

CHILD HEALTH SCREENING AND SURVEILLANCE: A CRITICAL REVIEW OF THE EVIDENCE

Report prepared by Centre for Community Child Health, Royal Children's Hospital Melbourne
for the National Health and Medical Research Council.

Endorsed 15 March 2002



TABLE OF CONTENTS

ACKNOWLEDGMENTS	3
1. EXECUTIVE SUMMARY	4
1.1. TERMS OF REFERENCE	4
1.2. METHODOLOGY	4
1.3. FINDINGS: OVERVIEW	5
1.4. FINDINGS: TOPIC SUMMARIES	6
1.5. APPROACHES TO EARLY DETECTION AND PREVENTION OF PROBLEMS	17
2. BACKGROUND	18
REFERENCES	18
3. DEFINITIONS AND CONCEPTS	19
3.1. SCREENING	19
3.1.1. DEFINITION OF SCREENING	19
3.1.2. TERMS USED IN APPRAISAL OF A SCREENING TEST	20
3.1.3. TERMS USED IN APPRAISAL OF A SCREENING PROGRAM	20
3.1.4. TERMS USED IN APPRAISAL OF A SCREENING TEST WITHIN A SCREENING PROGRAM	20
3.2. PREVENTION	21
3.3. SURVEILLANCE	21
3.3.1. DEFINITION OF SURVEILLANCE	21
3.3.2. APPRAISAL OF SURVEILLANCE	22
3.4. HEALTH PROMOTION	22
REFERENCES	22
4. PROGRAM ISSUES	23
4.1. HARM / BENEFIT	23
REFERENCES	25
5. METHODOLOGY	26
5.1. CONSULTATION STRATEGY	26
5.1.1. EXPERTS IN CHILD HEALTH	26
5.1.2. EXPERTS IN LITERATURE SEARCHING	26
5.1.3. EXPERTS IN TOPIC AREAS	26
5.2. LITERATURE RETRIEVAL METHODS	26
5.2.2. REVIEW OF EXISTING DATABASES OF REVIEWS	26
5.2.3. SEARCH OF THE PEER-REVIEWED LITERATURE 1993-2000	27
5.2.4. LIMITED REVIEW OF PRE-1993 PEER-REVIEWED LITERATURE	27
5.2.5. REVIEW OF ELECTRONIC PUBLICATIONS AND WEBSITES	27
5.2.6. REVIEW OF OTHER PUBLICATIONS	28
5.2.7. REVIEW OF UNPUBLISHED RESEARCH	28
5.3. EVALUATION	28
5.4. ASSESSING THE EVIDENCE	30
5.4.1. ASSESSING THE QUALITY OF SYSTEMATIC REVIEWS.	30
5.4.2. ASSESSING THE QUALITY OF RANDOMISED CONTROLLED TRIALS.	32
5.4.3. ASSESSING THE QUALITY OF OTHER STUDIES.	33
5.5. EXCLUSION CRITERIA	33
REFERENCES	33
6. DISEASES / CONDITIONS	35
6.1. TOPIC LIST	35
6.2. SELECTION CRITERIA	35
6.3. DISCUSSION OF EACH DISEASE / CONDITION	35
6.3.1. ATLANTO-AXIAL INSTABILITY (in Down Syndrome)	36
6.3.2. CONGENITAL ADRENAL HYPERPLASIA	41
6.3.3. CONGENITAL HEART DISEASE	49
6.3.4. CONGENITAL HYPOTHYROIDISM	55
6.3.5. CYSTIC FIBROSIS	62

6.3.6. DEVELOPMENTAL DYSPLASIA OF THE HIP	69
6.3.7a. CONDUCTIVE HEARING IMPAIRMENT	78
6.3.7b. PERMANENT CHILDHOOD HEARING IMPAIRMENT	88
6.3.8. HYPERTENSION	106
6.3.9. IRON DEFICIENCY	110
6.3.10. LEAD TOXICITY	120
6.3.11. PHENYLKETONURIA (PKU)	129
6.3.12. SCOLIOSIS	137
6.3.13. CRYPTORCHIDISM (UNDESCENDED TESTIS)	145
6.3.14. URINARY ABNORMALITIES	151
6.3.15. VISION DEFECTS	157
6.3.16. DENTAL HEALTH	172
6.3.17. DEVELOPMENT	179
6.3.18. LANGUAGE	189
6.3.19. HEIGHT	198
6.3.20. WEIGHT	206
7. LIMITATIONS OF THIS REVIEW	215
8. SUMMARY OF FURTHER RESEARCH RECOMMENDATIONS	216
9. THE LANGUAGE OF EARLY DETECTION: TIME FOR REVIEW?	220
9.1. CURRENT TERMINOLOGY	220
9.1.1. SCREENING	220
9.1.2. SURVEILLANCE	220
9.1.3. ACTIVITIES DESIGNED TO PREVENT PROBLEMS AND PROMOTE HEALTH	221
9.2. MOVE AWAY FROM THE CONCEPT OF PASS/FAIL	221
9.3. CLARIFICATION OF TERMINOLOGY	222
9.3.1. SCREENING	222
9.3.2. SURVEILLANCE	222
9.3.3. ACTIVITIES DESIGNED TO PREVENT PROBLEMS AND IMPROVE OR PROMOTE HEALTH	222
9.4. PRINCIPLES AND PROCESSES FOR SERVICE DELIVERY	223
REFERENCES	224
10. GLOSSARY	225
11. APPENDICES	226
A. REFERENCE GROUP	226
B. SEARCH STRATEGIES	227
C. WEBSITES	228

ACKNOWLEDGMENTS

In addition to those people acknowledged in Chapter 6 for their input into specific topic documents, the project team would like to thank the following people for their assistance with the development and preparation of this report:

Dr Julian Higgins for conceptual input into the development of appraisal strategies.

Mr Steve McDonald for advice and assistance with the development of literature search strategies.

Ms Judy Stoelwinder for assistance with the development of literature search strategies.

Ms Melissa Graham for photocopying and referencing.

Authors

This review was conducted by a project team from the Centre for Community Child Health that comprised:

Professor Frank Oberklaid

Dr Melissa Wake

Dr Claire Harris

Ms Kylie Hesketh

Dr Martin Wright

NHMRC Child Health Screening and Surveillance Steering Committee - Terms of Reference

Review and update the 1993 NHMRC document "*Review of Child Health Surveillance and Screening*" following the procedures recommended by the NHMRC 1998 document "*A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines*". Specifically:

- identify the scope of the guidelines;
- review and evaluate:
 - the 1993 NHMRC document "*Review of Child Health Surveillance and Screening*";
 - post 1993 literature and research related to child health screening; and
 - the extent, quality, level, relevance and strength of scientific evidence for the effectiveness and appropriateness of relevant interventions.
- undertake consultation within NHMRC requirements, including:
 - consulting groups developing similar guidelines for screening tests;
 - consulting key stakeholders; and
 - consulting the public.
- develop child health screening guidelines that:
 - lead to an improvement in the health outcomes for children and their families;
 - address the needs of disadvantaged Australians;
 - are based on the best available evidence;
 - identify areas for future research;
 - identify economic implications;
 - include a strategy for best dissemination and implementation; and
 - specify evaluation and updating requirements.
- prepare companion documents based on the guidelines:
 - an easy reference guide for health and allied health professionals; and
 - a consumer guide that provides accessible and understandable information in a culturally appropriate fashion to parents/carers and other consumers to aid informed decision making.
- present child health screening guidelines to the Health Advisory Committee by 30 June 2000.

Members

Professor Lesley Barclay (Chair)	Director, Centre for Family Health and Midwifery, UTS
Professor David Henderson-Smart	Director, Centre for Perinatal Health Services Research, Sydney University
Professor Graham Vimpani	Director, Child Adolescent and Family Health Service, Wallsend Campus
Ms Nan Davies	Director of Nursing, Planning and Development Child and Youth Health

Secretariat

Mr Phil Callan	Health Advisory Section, NHMRC
----------------	--------------------------------

1. EXECUTIVE SUMMARY

The early detection of health and other problems in children is a worthy goal. At first glance the benefits appear to be self-evident. It is not until one begins to systematically review the evidence for screening and surveillance, topic by topic, that the complexity of this endeavour becomes apparent.

There is surprisingly little evidence for the effectiveness of screening programs in many domains (that is not to say that the corollary is true – for some, there is inadequate evidence for lack of effectiveness either). There are scant data about cost effectiveness. There are major issues of program quality, monitoring of compliance with referrals for assessment, and whether facilities exist in many communities for assessment and follow up. There are concerns that much attention is paid to the test or procedure itself and little to the main elements of a community-wide program. In some cases, there is little evidence that therapy alters outcomes. These issues are outlined in relevant chapters on topics. Finally there are problems with terminology, and the way in which terms such as “screening” and “surveillance” are used.

This report should not be construed as diminishing the importance of attempts at early detection. Rather it provides evidence that perhaps we need to rethink how best we do this; a discussion of these issues is to be found in Chapter 9.

A distillation of the evidence, recommendations and research agenda for each topic is summarised in this section; details are to be found in the body of work for each topic. The topics are listed alphabetically for convenience.

1.1 TERMS OF REFERENCE

The NHMRC last undertook to review child health screening and surveillance activities in 1993. Since the last review there has been an emerging acknowledgment of the importance of evidence based medicine. In 1999 the NHMRC instigated an update of the 1993 review employing evidence based techniques. Initially this work was to be carried out as a single task of critically reviewing the evidence and developing clinical guidelines. Prior to commencement of the review, NHMRC decided to divide the task into a two stage process: Stage 1, consisting of a critical review of the evidence for screening and surveillance activities around fourteen child health topics, followed by Stage 2, consisting of the development of evidence based clinical guidelines for practitioners informed by Stage 1. This report constitutes Stage 1 of the process. Following consultation with the NHMRC-appointed Steering Committee and the contractors, the list of fourteen child health topics to be included in the review was revised to twenty-one. However one of the reviews (post-natal depression) was ultimately excluded from the final document, as a more comprehensive review of this subject had recently been released by NHMRC.

The guidelines set out by NHMRC to conduct the review included:

- identification of relevant information through searches of peer-reviewed literature published since the release of the last review (post-1993), consultation with groups undertaking similar work both within Australia and internationally, and consideration of submissions received by NHMRC during two public consultations.
- critically evaluate the scientific literature and other information for quality and strength of evidence using the ratings outlines in the NHMRC publication “A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines”. This requirement was modified to allow rating of non-clinical trial evidence which is not comprehensively covered by the NHMRC publication.

It was decided by the NHMRC-appointed Steering Committee that the review should be limited to the consideration of children from birth to eighteen years of age, and thus would not cover prenatal screening activities.

The findings within this report represent those of the authors of this report, based on their critical review of the literature up to August 2000 and informed by comment from the project’s Reference Group and a limited number of experts in each field. This report does not necessarily represent the views of the NHMRC. Because clinical guidelines were not developed, this review has not been subject to the wide external consultation usually associated with guideline development.

1.2 METHODOLOGY

An expert Reference Group was established (Section 5.1). This group provided advice and feedback to the project, from its inception to its completion. In addition, individual members were asked to advise on

specific topics. For each topic, experts in that topic area were also consulted to ensure that information in the report was factually correct, and that key works in that area had not been missed.

In consultation with experts in searching the electronic literature, search strategies were developed to locate studies published in 1993-2000 (ie since the previous NHMRC review) according to quality of evidence ratings, as defined in the NHMRC publication “Guidelines for the Development and Implementation of Clinical Practice Guidelines”. Search strategies are included in Appendix B. A limited review of the pre-1993 peer-reviewed literature was undertaken. Other data sources were also searched (Section 5.2).

Final selection of topics to be reviewed, search strategies, specified frameworks to assess the quality of different aspects of the literature (Sections 5.3 and 5.4), terminology (Chapter 3), and the proposed structure for the topic reviews were agreed between the Steering Group and the project team prior to commencement of individual topic reviews.

The draft of the final review was scrutinised by both the Steering and the Reference Groups and modifications made in light of their comments. This also led to the addition of Chapter 9, “The language of early detection: time for review?”.

1.3 FINDINGS: OVERVIEW

Relatively few topics could be recommended for formal screening programs.

For many conditions we could not recommend formal screening programs, for a variety of reasons including:

- multidimensional conditions on a continuum of normality-abnormality do not lend themselves to pass/fail criteria (see Chapter 9)
- available screening tests not considered sufficiently acceptable to the target population, based on reported uptake rates of either the screening test or definitive referral
- sensitivity could not readily be balanced against specificity – ie very large numbers of false positives a by-product of capturing all or most of those with the target condition
- the target condition itself was too variable over time to justify screening at a single time point, but evidence to support periodic screening (surveillance) not available.
- management for those detected by screening not shown to significantly alter outcomes
- not yet an agreed therapy.

Some conditions seemed important and management likely to alter outcomes, but there was inadequate research evidence to support this. For these conditions, further research seems vitally important.

For some conditions ‘systems issues’ seemed especially important. A good example is clinical examination in the newborn period. Early discharge (<72 hours post partum) is now commonplace in Australia, thus there is potential for screening to be missed or incomplete if the infant is discharged from hospital within a few hours of birth. It is important that there is good communication between hospitals and the community, that clinicians in both settings are knowledgeable and competent in the management of neonates, and that a line of responsibility is clear. The findings of examinations undertaken in the maternity unit and any follow up required should be clearly communicated to health professionals in the community. Systems are required to ensure that:

- a point of accountability exists for each child’s neonatal examination
- a complete neonatal examination has been undertaken
- the findings of the examination and any subsequent action, referrals, etc have been documented
- correspondence between health professionals in maternity settings and those in the community (particularly GPs and MCHNs) is adequate and appropriate.

Similar systems issues need to be considered for other conditions with respect to early identification.

For some conditions, the balance of harm vs benefit may change over time as new therapies or detection techniques become available, and thus suitability for screening will need to be reviewed. For example, prenatal screening may lower the yield of newborn screening for some conditions, while effective therapy for overweight might make screening or surveillance appropriate in the future.

1.4 FINDINGS: TOPIC SUMMARIES

1. ATLANTO AXIAL INSTABILITY

Insufficient evidence to make a recommendation for or against screening.

Comments:

Little is known about:

- the natural history of atlanto axial instability
- prediction of who will suffer adverse outcomes
- rate of progression, if any

Therefore, it is impossible to weigh the harms vs the possible benefits of screening or to make sensible recommendations about timing and frequency of such screening.

Recommendations

Implementation of a formal screening program for atlanto-axial instability in children with Down Syndrome is not recommended.

2. CONGENITAL ADRENAL HYPERPLASIA

Insufficient evidence to make a recommendation for or against screening

Comments:

There are many unanswered questions about the tests and protocols used, and there is wide variation in results from various programs.

Screening has the potential to detect life-threatening conditions prior to the onset of symptoms. However, death from undiagnosed congenital adrenal hyperplasia appears to be exceedingly rare in developed countries. Many children are detected clinically at birth because of virilisation, and many others are detected within weeks of birth at the onset of symptoms and treated appropriately.

Neonatal screening also has the potential to detect milder forms of disease that do not present until late childhood. Over-treatment is a potential adverse outcome for those with mild congenital adrenal hyperplasia detected by screening.

Some of the morbidity suffered prior to diagnosis through clinical vigilance can be avoided through screening. However, this has not been quantified or weighed against potential harms and costs. Therefore, it is impossible to make recommendations about screening at this time.

Recommendations

Implementation of a universal screening program for congenital adrenal hyperplasia is not recommended at this time.

3. CARDIAC DISEASE

Insufficient evidence to make a recommendation for or against screening

Comments:

The neonatal and 6 week postnatal examinations offer routine access to newborns to detect congenital abnormalities. Although there are limitations to clinical examination as a screening tool in this setting, it may be the only opportunity to detect CHD in the pre-symptomatic phase.

However, it appears that cardiac examination frequently falls short of *program* requirements for an effective detection system.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

Recommendations

Although there is little firm evidence to support the value of screening, we recommend continuation of newborn and 6 week cardiac examination, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical and nursing staff
- clear pathways and/or guidelines for referral
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records
- strategies to improve attendance at the 6 week postnatal check (particularly for disadvantaged groups)
- opportunistic examination of children not examined in the newborn period

4. CONGENITAL HYPOTHYROIDISM

Good evidence to make a recommendation for screening.

Comment: although no randomised controlled trials or comparative trials were identified, the quality of population-based cohort studies and the clear and sizeable benefits of screening approach this standard.

Recommendations

That screening programs continue.

That the impact of adding multiple new conditions to newborn screening programs is monitored, to ensure that the quality of existing congenital hypothyroidism programs is not threatened

5. CYSTIC FIBROSIS

Fair evidence to recommend screening

Comments:

The general expert consensus is that earlier diagnosis through screening is beneficial. However, this benefit is likely to be modest.

Short term advantages of reduced morbidity and nutritional benefits have been demonstrated.

With improved medical care, individuals now typically live into early middle age, so that cystic fibrosis is no longer usually a lethal condition in childhood. Systematic reviews report that there is as yet no conclusive evidence that neonatal screening alters long term prognosis in cystic fibrosis. As therapy through childhood, adolescence and early adulthood continues to improve, the relative benefits of screening may lessen.

Neonatal screening is currently available in most Australian states. There are short term advantages to early detection, the test is relatively low cost and has low false-positive rates, and screening is part of the current neonatal heel prick screening program.

The brief of this review was to focus on neonatal screening; pre-conceptual and antenatal screening programs also require debate.

Recommendations

Continuation of current neonatal screening programs for cystic fibrosis is recommended.

6. DEVELOPMENTAL DYSPLASIA OF THE HIP

Insufficient evidence to make a recommendation for or against screening by clinical examination

Fair evidence to recommend against screening by ultrasound

Comments:

The neonatal and 6 week postnatal examinations offer routine access to newborns to detect congenital abnormalities. Although there are limitations to clinical examination as a screening tool in this setting, it may be the only opportunity to detect developmental dysplasia of the hip in the pre-symptomatic phase.

However, it appears that hip examination frequently falls short of *program* requirements for an effective detection system. Although Australia has fewer cases missed by screening and a lower incidence of surgery than the UK where a similar screening regimen is in place, there is still considerable room for improvement. Given the uncertainty of the natural history, the knowledge that some cases will resolve spontaneously and the hypothesis put forward by some authors that repeated examinations may increase instability, further research is required in this area.

Consideration of strategies to improve the sensitivity and specificity of clinical examination is also required.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

The evidence suggests that ultrasound is not useful as a universal screening tool, and may result in unnecessary treatment and increased costs. However ultrasound screening for children at high risk of developmental dysplasia of the hip may have some merit.

Recommendations

Universal ultrasound screening for developmental dysplasia of the hip is not recommended.

Although there is little firm evidence to support the value of clinical screening for developmental dysplasia of the hip, we recommend continuation of newborn and 6 week examination using the Ortolani and Barlow maneuvers, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical and nursing staff
- clear pathways and/or guidelines for referral
- standardised follow up procedures for infants with abnormal findings on clinical examination
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records
- strategies to improve attendance at the 6 week postnatal check (particularly for disadvantaged groups)
- opportunistic examination of children not examined in the newborn period

Measures to increase the effectiveness of a screening program for developmental dysplasia of the hip should be implemented and the incidence of surgery and late presentations monitored. If these rates are not reduced by improvements in the screening program, the benefits of clinical screening need to be examined further.

7a. CONDUCTIVE HEARING LOSS

Good evidence to recommend against screening

Comments:

Strong evidence indicates that long term developmental effects of otitis media with effusion detected through screening are minimal. Otitis media with effusion is a fluctuating condition, so that one off screening makes no sense. If screening were to occur it should be repeated many times for each child. However, treatment of children detected through repeated screening has been shown to provide no long term developmental or academic benefit.

However, no research program has yet adequately researched this issue for the small number of children with significant persistent conductive hearing impairment, as opposed to those with persistent otitis media with effusion regardless of hearing status.

Recommendations

Universal screening programs to detect otitis media with effusion are not recommended.

7b. PERMANENT CHILDHOOD HEARING IMPAIRMENT

Fair evidence to recommend universal neonatal hearing screening

Insufficient evidence to make a recommendation for or against genetic screening

Insufficient evidence to make a recommendation for or against school entry screening

Good evidence to recommend against distraction testing

Comments:

Hearing screening before discharge for all NICU neonates and preferably all neonates admitted to special care nurseries for more than 48 hours is now accepted best practice.

There is excellent evidence that it is feasible to test the hearing of *all* newborns within maternity hospitals, and that this results in the detection of many children with significant hearing impairment much earlier than would otherwise be the case. There is fair evidence to support benefits of such screening in terms of language outcomes, but this has not been well quantified, and evidence to support benefits on academic outcomes and quality of life is awaited. Testing appears to be well accepted by parents but this requires further study.

However, there is as yet little evidence of *program* sensitivity when applied at population level. We await information on population benefits in terms of median age of diagnosis and proportions of children diagnosed late. We also await evidence of incremental benefits in language, academic and quality of life outcomes at a population level. US programs have reported relatively high levels of loss to follow up.

Population-based programs are likely to stand or fall on the logistics and quality of the testing system, and on the logistics and quality of the follow up system for screen positives. Given the expense of universal neonatal hearing screening, such issues need to be addressed *before* decisions are made to undertake universal screen for whole populations. In particular, Australia is large and has many small communities geographically removed from support services. These will be a particular issue for program delivery and may necessitate adapted or alternative systems if adequate outcomes are to be obtained.

Genetic tests are highly sensitive and specific for approximately 15-20% of deaf children with single gene mutations, most of whom do not have identified risk factors for deafness. This proportion is likely to double over the next five years. Logistic issues have already been extraordinarily well dealt with for universal neonatal heelprick testing. Should universal hearing screening prove very difficult to fully implement in Australia, an alternative could be to test the hearing of all NICU and special care babies who are at risk for deafness, backed up with genetic screening of all other babies through the existing heelprick system. If successful, this could offer major logistic and cost advantages. However, this is as yet only at the research stage.

Recommendations

1. Hearing screening before discharge for all NICU neonates and preferably all neonates admitted to special care nurseries for more than 48 hours is now accepted best practice, and should become a high priority at State level.
2. The evidence supports some form of neonatal screening for hearing impairment. However, to be successful implementation packages require careful trialing for the situation in which they will be used. Pilot projects should be mounted to trial the feasibility of different models of infant screening, paying careful attention to processes, acceptability and costs, in both large and small communities. These should include settings other than major metropolitan hospitals (already under study in West Australia.) Models to consider could include at-risk screening with and without genetic screening, hospital based UNHS, and community based UNHS.
3. An adequately funded economic analysis should be undertaken urgently to examine costs of different models (including genetic screening) of infant hearing screening for Australian settings. This should take into account likely achievable coverage and program sensitivity for each model at a population level.
4. A national forum should be convened to discuss possible infant screening models in terms of logistics, coverage, acceptability, program maintenance and quality, minimum standards, data collection/reporting and outcome analysis at a population level for each state and for Australia as a whole. This should take results of the economic analysis into account. Models to consider could include at-risk screening with and without genetic screening, hospital based UNHS, and community based UNHS.
5. Final decisions about implementation of universal neonatal screening programs in Australia should be guided by the above information when available.
6. If the above processes lead to an agreement about proceeding with a specific model(s), planning and implementation would ideally be overseen by a national body set up for this purpose. This could draw on the experience of BreastScreen, the most recent new screening program implemented in Australia. Formal channels of communication with the UK, US, and Canada should also be established.
7. Careful population-based documentation of a wide range of outcomes for children with permanent childhood hearing impairment should commence now, against which to assess the benefits of new programs if/when introduced.

8. HYPERTENSION

Fair evidence to recommend against screening

Comments:

Childhood hypertension and measurement of blood pressure in children do not fit criteria for screening.

We were unable to find any evidence that measurement of blood pressure as a screening procedure during childhood improves health outcomes, or that it is a valid or reliable screening test.

Recommendations

Implementation of screening programs to detect hypertension in well children is not recommended.

9. IRON DEFICIENCY

Fair evidence to recommend against screening

Comments:

Iron deficiency does not meet criteria for screening in infants and children since the most appropriate age for screening is unclear, and the benefits are unclear and may be outweighed by the harms. In addition, coverage rates in existing programs and research projects have been poor.

However, there is fair evidence to recommend that health professionals maintain a high level of vigilance for iron deficiency and iron deficiency anaemia in children (particularly among children from low socioeconomic areas), and that iron status be assessed based on risk factors.

Iron deficiency may be more appropriately seen as a public health, rather than an individual, issue.

Recommendations

It is recommended that primary health care professionals provide parents with age-appropriate dietary information congruent with the Australian Dietary Guidelines to assist in the prevention of diet-related iron deficiency in infants and children.

Health care professionals working in areas of lower socioeconomic status, where the prevalence of iron deficiency is likely to be higher than average, should be alert to the possibility of iron deficiency in their child patients and assess iron status based on risk factors (uncorrected iron deficiency in the mother during pregnancy, prematurity, age <24 months, introduction of cow's milk as the main milk source before 12 months of age, consumption of >600ml of cow's milk per day).

Children living in areas prone to lead toxicity and (as per current NHMRC guidelines) children with blood lead levels >0.72µmol/L should be tested for iron deficiency.

10. LEAD TOXICITY

Fair evidence to recommend against screening

Comments:

Lead toxicity does not meet criteria for screening in infants and children because it is poorly accepted by professionals and parents and lead exposure is low (and falling) for the great majority of Australian children. There is no information about the effectiveness of screening programs.

Lead toxicity may be more appropriately seen as a public health, rather than an individual, issue.

Recommendations

Implementation of screening programs to detect lead toxicity in well children is not recommended.

We recommend five yearly surveys of representative samples of children, particularly those aged 1-4 years, in high and low lead exposure areas to document the prevalence of lead toxicity. We note that a similar recommendation was made in the 1993 NHMRC Review of Child Health Surveillance and Screening.

Health care professionals working in areas where the prevalence of lead toxicity is likely to be higher than average should be alert to the possibility of lead toxicity in their child patients.

Public health preventive and management strategies are likely to be most effective in further reducing childhood lead levels.

11. PHENYLKETONURIA

Fair evidence to recommend screening

Comments:

Although never subjected to formal controlled trial, the benefits of screening for phenylketonuria are not in doubt.

However, although treatment of PKU detected by screening prevents intellectual disability, some lasting cognitive effects are common. In addition, long term disability that may be experienced by the offspring of women with PKU (due to high levels of phenylalanine during pregnancy) has not yet been well quantified at population level.

Recommendations

Continuation of current universal newborn screening programs for phenylketonuria is recommended.

Screening programs should incorporate population based monitoring and management programs for women of childbearing age, aiming to systematically prevent avoidable intellectual disability in their infants. Such monitoring should meet defined, stringent quality and reporting standards, and include reporting of long term outcomes for these children.

12. SCOLIOSIS

Good evidence to recommend against screening

Comments:

Although there are good, relatively inexpensive tests available to screen for idiopathic scoliosis, there is no evidence that screening programs are beneficial. Most cases of idiopathic scoliosis are detectable without screening, many do not require treatment and the treatments available do not appear to be of benefit to those with mild to moderate spinal curvature.

Recommendations

Implementation of new scoliosis screening programs is not recommended.

In communities where screening programs are already in place, continuation of these programs should be reassessed.

13. UNDESCENDED TESTES

Insufficient evidence to make a recommendation for or against screening

Comments:

No evidence regarding the effectiveness of screening programs or screening tests for undescended testes was found.

The current practice in Australia is for all children to have a neonatal examination conducted by a health professional that would routinely include examination of the genitalia. There are several factors that would encourage continuation of this practice:

- Clinical examination is the only test available for detecting undescended testes
- Consensus is that the optimum time for this examination is at birth
- There is evidence of damage to the undescended testis after the first year of life
- There is evidence that treatment before the age of 2 years improves adult fertility

However, it appears that screening for undescended testes frequently falls short of *program* requirements for an effective detection system. UK data suggests poor rates of detection from neonatal screening and that, of those detected, many are not referred for surgical opinion and treatment. In addition, lack of clear diagnostic criteria, lack of appropriate training for personnel involved, inadequate guidelines and pathways for referral have been identified as problems related to this process.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

No evidence is available about the effectiveness of screening for undescended testes at later ages, nor is there evidence regarding the benefits of treatment beyond the age of two.

Recommendations

Although there is little firm evidence to support the value of screening, we recommend continuation of specific examination of the genitalia at the newborn and 6 week checks, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical and nursing staff
- clear pathways and/or guidelines for referral
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records

- strategies to improve attendance at the 6 week postnatal check (particularly for disadvantaged groups)
- opportunistic examination of children not examined in the newborn period
- program evaluation and quality improvement

14. URINALYSIS

Fair evidence to recommend against screening

Comments:

There is no evidence that screening urinalysis prevents renal or other disease. There is information to suggest that treating asymptomatic bacteruria is harmful. Screening urinalysis is costly to the community, may result in physical and psychological costs to the patients and their families, and has high rates of misclassification.

Recommendations

Screening programs of urinalysis in well children are not recommended.

15. VISION

Insufficient evidence to make a recommendation for or against neonatal screening

Fair evidence to recommend against screening for risk factors for amblyopia

Insufficient evidence to make a recommendation for or against preschool visual acuity screening

Fair evidence to recommend against colour vision screening

Comments:

There is some evidence about the impacts of different disorders of vision, and that some interventions are effective. There is some evidence as to how vision screening tests function in selected samples, but less as to how they function in community samples. There is a dearth of high quality evidence about the effectiveness of vision screening programs coupled with intervention services.

A number of peak bodies currently support vision screening in children. Recommendations generally combine current understanding about the natural history and effect of different vision disorders, “common sense”, and lower levels of evidence.

There is a lack of good evidence that clinical examination of the neonatal eye is effective. The seriousness of the conditions being screened for, the agreement on interventions, and the improved effectiveness of sufficiently early interventions support the currently recommended practice of eye examination in the newborn. It may not be ethical to carry out an RCT in such a situation. However, examination requires considerable skill. There is little evidence as to how well this examination is currently carried out at program level and therefore its effectiveness.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

Recommendations

1. Although there is little firm evidence to support the value of screening, we recommend continuation of specific examination of the eye at the newborn check, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical staff
- clear pathways and/or guidelines for referral
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records

Because neonatal eye examination is difficult, minimum program standards of coverage and accuracy should be set and evaluated.

The examination should include external examination of the eye with particular attention paid to corneal size and clarity. The fundus red reflex should be examined in dim illumination with a direct ophthalmoscope to exclude media opacities. Examination should be carried out by well trained and skilled personnel as part of the neonatal examination.

Examination of high risk neonates by ophthalmologists should continue.

2. Screening for risk factors for amblyopia should be reviewed as photoscreening improves. Without further evidence, such programs are not recommended.

3. Programs of preschool or school entry visual acuity screening should not be instituted. What to do when such programs are already in existence is more problematic. Further research into visual acuity screening is required (see Chapter 8: Summary of Further Research Recommendations). On the basis of current evidence, we suggest that in such cases there be community discussion and education. The aim should be to replace visual acuity screening programs with greater support for visual acuity evaluation as part of a comprehensive assessment when children are identified to have a learning or behavioural problem.

4. Without further evidence, screening for abnormalities of colour vision is not recommended.

16. DENTAL HEALTH

Insufficient evidence to make a recommendation for or against screening

Comments:

There is no clear evidence or consensus opinion on whether caries in the primary dentition should receive treatment.

Recommendations

Screening programs for dental caries in the deciduous teeth are not recommended.

We recommend regular surveys to document the prevalence and severity of caries in preschool and school aged child populations.

Public health efforts should focus on preventive dental health in preschool and school aged children. This should be given higher priority than screening efforts.

17. DEVELOPMENT

Insufficient evidence to make a recommendation for or against developmental screening

Comments:

Currently developmental delay does not appear to meet criteria for screening programs. However, there is evidence that the identification of developmental delay/disability (or significant risk factors) and subsequent intervention can improve developmental and other social outcomes. There is some evidence that the earlier the intervention, the better the outcome. In addition, if it can be argued that in the case of child development there should be less stringent expectations in terms of test sensitivity and specificity (given the intrinsic difficulties in measurement of development, the fact that even false positives from developmental screening may benefit from some form of intervention, and the high prevalence in the community), there are some developmental screening tests that are acceptable. There still will be, however, a large number of children who will not be identified by developmental screening tests.

There is no high quality evidence putting together all the links in the chain and reporting on the effectiveness of developmental screening programs on child developmental outcomes.

Recommendations

The identification of children who would benefit from early intervention should not be based solely on the use of developmental screening tests, or limited to inquiry at one point of time.

It is not recommended that screening programs for developmental delay be implemented at this stage.

Until there is evidence that alternatives to developmental screening programs function better, existing programs should be reviewed to ensure that adequate tools and processes are used

Community programs of developmental screening should use tools that have been demonstrated to have adequate psychometric properties (sensitivity and specificity greater than 70%).

Community programs of developmental screening should allow for intervention or assistance of some degree for children who fail screening, but who pass diagnostic tests.

Individualised checklists of milestones should not be used as developmental screening tests.

18. LANGUAGE

Insufficient evidence to make a recommendation for or against screening

Comments:

Difficulties with accurate definition of language delay and its natural history currently argue against the introduction of formal screening programs. In addition, it may be difficult to implement screening programs in the community to a sufficiently high standard to make it effective.

On the other hand, treatment appears to be beneficial for children with established language delay, and one (flawed) RCT has suggested a distinct benefit for very young children detected through screening.

Recommendations

Implementation of formal screening programs for language delay is not recommended at this stage.

Further research is urgently needed to better quantify early predictors of later language delay, and to confirm promising results from early, community-based secondary prevention programs.

19. HEIGHT

Insufficient evidence to make a recommendation for or against screening

Comments:

We could identify little evidence demonstrating the effectiveness of this activity. Height screening invariably leads to large numbers of referrals of “short-normal” children; the cost, acceptability and uptake of referrals has not been quantified; and fundamental requisites of the screening test (sensitivity, specificity, positive and negative predictive value, and NNT) are yet to be quantified for different centile cutpoints.

Nonetheless, screening for short stature due to growth hormone deficiency is felt by many to be of benefit. With many areas now without routine height screening programs, a formal randomised controlled trial incorporating economic analysis could be mounted to answer what is seen by many to be an important question.

Regular population-based surveys of child height are important at a population level to understand secular trends and to better understand long term nutritional status of the child population and of disadvantaged subgroups.

Recommendations

New growth screening programs outside the research context are not recommended.

Should retrospective review suggest that age of diagnosis of growth hormone deficiency and Turner syndrome are rising following cessation of height monitoring, further study of the usefulness of height measurement as a screening tool should be undertaken.

Regular, systematic population-based surveys of height are recommended to monitor secular trends in height for the whole population and for particular subgroups.

20. WEIGHT

Insufficient evidence to make a recommendation for or against screening for failure to thrive

Comment: while there is some evidence that systematic detection and management of failure to thrive in the first year of life can improve outcomes, this has yet to be operationalised into a practicable screening system.

Fair evidence to recommend against screening for obesity

Comment: childhood overweight and obesity are emerging as a public health problem of a massive dimension, with incalculable future costs. If effective and accessible therapy becomes available, this recommendation should be urgently reviewed.

Recommendations

Routine weight monitoring at birth, at 6-8 weeks, and at 8-12 months is recommended as part of routine clinical care. This does not however constitute a screening program.

1.5 APPROACHES TO EARLY DETECTION AND PREVENTION OF PROBLEMS

Like others, we recommend a three tier approach to the early detection and health promotion (Chapter 9).

For defined conditions, there is strong evidence for the benefits of screening. For a few conditions, the absence of such evidence should not be construed as being equivalent to there being evidence of the lack of benefit to such an approach. However, given the complex and multi-dimensional nature of child development and behaviour, there is unlikely ever to be a test that fulfils the scientific criteria for screening and the term “developmental screening” should not be used for these conditions.

Health surveillance as presently conceived is considered to be a vague and sometimes confusing term which means different things to different people. We would suggest that surveillance be more narrowly defined as activities designed to detect specific, definable problems in children at an early stage – essentially, sequential or periodic screening. Few of these activities can currently be recommended on the existing evidence; this does not mean that detection of these problems is not important.

We believe that the most pressing need is for further research and critical thinking around prevention of problems and promotion of health. These activities may be universal or targeted, based on clearly-defined indicators of risk. Such activities should be comprehensive and planned at a population level, while responsive to local needs. The evidence base for these activities should in general be as rigorous as that we have applied to screening in this critical review, and their practice subject to similar quality requirements.

Efforts at early detection and health promotion should be linked closely to the service delivery system for children and families and should take place in all settings and by all professionals who come into contact with children and their families; they should not be regarded as solely the responsibility of health professionals. An active research program into efficacy and effectiveness of various approaches to detection, intervention and health promotion should be strongly encouraged, and evaluation be considered an integral part of program development.

2. BACKGROUND

Risk factors in childhood are related to many adverse outcomes in later life. This is as true for conditions that are predominantly psychosocial in nature as for those that are predominantly organic. If we could detect these risk factors early and then treat or manage them effectively, we could lift a huge individual and societal burden of lifelong impairment - especially when it is clear that late treatment is ineffective and often very expensive.

Unfortunately, there is still a huge divide between the logic of these statements and the successful implementation of effective screening and surveillance programs. In the last 75 years, we have weighed and measured children, watched them draw circles and crosses and hop and skip, and physically examined them over and over again - and only recently have we realised that this may be futile. Only a small proportion of childhood screening and surveillance activities has been demonstrated to be effective. The remainder have either not been evaluated, or, worse, have been demonstrated to be ineffective.

The requirements of screening and surveillance programs are diverse and complex. Detection of a condition or risk factor requires firstly that we are able to define it clearly, and secondly that we have an accurate test that can be implemented within an acceptable and cost-effective program. Effectiveness, acceptability and affordability of subsequent management of the condition in question must also be considered. Other questions also complicate screening and surveillance in the late 20th century and into the beginning of the 21st century, with the rise in chronic conditions which have psychosocial and environmental roots. Whom should we screen, and whom should we treat, when there are no clear cutoff points and perhaps everyone could benefit from available interventions? Even if we have an effective management, what do we do when people are unable to modify their behaviour to implement it as we would wish? As genetic screening becomes more and more feasible, what are its ethical and cost implications, particularly when testing is available antenatally? How do we deal with populations where there may be differential risk? And if "all screening programmes do harm; some can do good as well" (Muir Gray 1997), how do we balance the harm and the good in judging the overall worth of a program?

Attempts to develop effective screening programs span several decades. While for some conditions (eg phenylketonuria) there has existed for many years a simple, reliable and cost effective test which meets all the WHO guidelines for a screening test and has been shown to be both effective and efficient, for other areas (eg children's developmental problems) the search for an appropriate methodology has been more problematic. There exist a large number of developmental screening tests, for example, with little consensus about the best way to proceed in order to detect problems at the earliest possible stage. Many screening tests have had psychometric shortcomings, and have been plagued by poor rates of sensitivity and specificity. Others have been too time consuming so that they are impractical in a primary care setting; have needed considerable expertise in administration, creating problems with training and reliable administration; or have worked reasonably well in a controlled setting where they are administered by well trained professionals, but have not readily translated into population-wide screening.

Recent years have seen a shift in emphasis from screening (which implies professionals administering tests to children) to surveillance (which actively elicits parental concerns and makes parents and families a focus of efforts of early detection). More recently there has been a further shift towards attempting to develop systems that promote the health and well-being of all children, recognising that a number of adverse circumstances, especially environmental, may have a significant impact on outcomes.

This review and update of child health screening guidelines by NHMRC is, therefore, timely, and of significant importance in facilitating a shift in the delivery of health services in Australia away from treating established conditions towards prevention, early detection and early intervention.

REFERENCES

Muir Gray, J.A. 1997, "Evidence-based healthcare: How to make health policy and management decisions", Churchill Livingstone, New York.

3. DEFINITIONS and CONCEPTS

Screening, surveillance and other related terms are used widely in the literature, but there is notable inconsistency in both formal definitions and common usage. Some of these differences are due to variations in the provision of health services internationally and can be seen in relation to the nature of programs within the country of origin. Others are historical, associated with changes in practice driven by issues such as accountability, harm versus benefit and the need for a sound evidence base in decision making.

Numerous definitions have been published about screening and surveillance activities in general, and others relating to child health in particular. The various terminologies used have both strengths and weaknesses and, in the course of this research, we critiqued previous works and developed a series of definitions appropriate to this review. The following working definitions have been drawn from a wide range of sources and focus on screening and surveillance in a child health setting.

These working definitions were agreed upon in conjunction with the NHMRC-appointed Steering Committee and the Reference Group for this project prior to commencement of individual topic reviews. However, the review raised a number of issues as to the appropriateness and clarity of some of these definitions. Therefore, on completion of the review we critique the commonly accepted definitions presented here. We suggest new definitions of both screening and surveillance, together with activities designed to improve child health outcomes; these are discussed in Chapter 9.

3.1. SCREENING

3.1.1. DEFINITION OF SCREENING

The definition of screening most commonly cited in the literature comes from the US Commission on Chronic Illness (1957).

‘Screening is the presumptive identification of unrecognised disease or defect by the application of tests, examinations or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment’ (US Commission on Chronic Illness 1957).

While obviously withstanding the test of time, this definition has some limitations. In particular, it fails to acknowledge the implications of an overall screening program, that administration of the program and the follow up of those who undergo a screening test is as important as the test itself. In addition, the modern concept of screening for ‘risk factors’ is not allowed for. We feel that this definition is inadequate for the purposes of this review and propose that it is necessary to define and evaluate both screening tests and screening programs, and that the two are interdependent.

Screening test

‘Any measurement aimed at identifying individuals who could potentially benefit from intervention. This includes symptoms, signs, lab tests, or risk scores for the detection of existing or future disease’ (Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests 1996).

We plan to base this report on this definition of screening, although we will interpret “disease” to include disease, condition, or specified adverse health outcome.

Screening program

‘In a screening program, a test, or a series of tests, is performed on a population that has neither the signs nor symptoms of the disease being sought but whose members have some characteristic that identifies them as being at risk from that disease, the outcome of which can be improved by early detection and treatment. Screening actually consists of all the steps in a program from the identification of the population at risk to the diagnosis of the disease or its precursor in certain individuals to the treatment of those individuals’ (Muir Gray 1997).

Types of screening

Mass or population screening involves the screening of a whole population.

Targeted or pro-active screening involves identifying members of the population at risk.

Opportunistic screening involves taking the opportunity to administer a screening test when the contact with the individual or group is not primarily for screening purposes: eg during a medical consultation when the individual was attending for an unrelated health problem.

Aims of screening tests

By definition, the purpose of a screening test is not to provide a diagnosis. However, in practice, the boundaries between screening and diagnosis are not always clear. Some screening tests identify individuals who are at increased risk of developing a disease or condition at some time in the future, others will identify those who are at increased risk of existing disease and a third group where the method of screening is actually a diagnostic test.

3.1.2. TERMS USED IN APPRAISAL OF A SCREENING TEST

Sensitivity

The sensitivity of a test is the proportion of people with the condition who are correctly identified by a positive test result. It is a measure of the true positive rate.

Specificity

The specificity of a test is the proportion of people who do not have the condition who are correctly identified by a negative test result. It is a measure of the true negative rate.

False positive

An individual incorrectly identified by a positive test result who does not actually have the condition.

False negative

An individual incorrectly identified by a negative test result who does actually have the condition.

3.1.3. TERMS USED IN APPRAISAL OF A SCREENING PROGRAM

Efficacy

Efficacy is the outcome of a specific intervention under ideal conditions.

Effectiveness

Effectiveness is the outcome of a specific intervention under normal, everyday conditions. Muir Gray (1997) puts this in the context of a screening program:

'The effectiveness of any screening program is determined by:

- The sensitivity of the series of tests applied to the population
- The effectiveness of the therapy offered to those individuals discovered to have the condition

Thus screening effectiveness = test accuracy + therapeutic effectiveness'

Efficiency

Efficiency is a measure of the economy with which an intervention of known efficacy and effectiveness is carried out. This is frequently used synonymously with cost-effectiveness, which in most cases is appropriate. As well as the actual costs of a screening program, a measure of efficiency or cost-effectiveness must consider the costs of any potential harms of the intervention versus the benefits, and the opportunity costs of other interventions that are foregone in favour of the program in question.

Yield

The yield of a screening program is the number of new cases detected for every 100 cases screened. The incremental yield is the number of new cases that would not have been detected without the screening program.

3.1.4. TERMS USED IN APPRAISAL OF A SCREENING TEST WITHIN A SCREENING PROGRAM

Positive Predictive Value (PPV)

The positive predictive value of a test is the proportion of people with a positive test result who actually have the disease. It is the probability that a person with a positive test is a true positive.

The positive predictive value is related to the prevalence of the condition being screened for. If the condition has a high prevalence in the population, a test with good sensitivity and specificity will have a high PPV. If the condition has a low prevalence, a test will have a low PPV.

Negative Predictive Value (NPV)

The negative predictive value of a test is the proportion of people with a negative test result who are actually free of the disease. It is the probability that a person with a negative test is a true negative.

Number needed to test

The number of individuals in the target population that must be tested in order to find one case of the condition being screened for.

3.2. PREVENTION

Primary prevention

Interventions that prevent the occurrence of disease, eg immunisation, effective health education programs, screening programs aimed at detecting risk of later disease.

Secondary prevention

Interventions that prevent the development of disease by early detection, eg screening programs aimed at detecting early existing disease such as phenylketonuria or hypothyroidism.

Tertiary prevention

Interventions that impede the progress of established disease or disability through effective intervention, eg screening for sensorineural deafness, language disorders or developmental delay.

3.3. SURVEILLANCE

3.3.1. DEFINITION OF SURVEILLANCE

Again, there is lack of consistency in the definition and use of the term surveillance. Previous authors have recognised this, and some have even recommended that because of the nature of child health surveillance it warrants a unique definition (Stone 1990; Butler 1989; Hall 1996). A modification of the Hall definition was used in the previous NHMRC review (1993).

'Surveillance aims to optimise the health of children through the ongoing overview of the physical, social and emotional health and development of all children. It includes:

- the measurement and recording of physical growth
- monitoring of developmental progress
- the administration of screening tests
- offering and arranging intervention when necessary
- prevention of disease by immunisation and other means
- providing information and support to parents
- health education

Whilst other screening tests may be undertaken in childhood (eg by service clubs or in shopping centres) they are not considered a part of the surveillance if they are conducted in isolation from organised primary health care.

It should be noted that child health surveillance:

- is initiated by health professionals but involves partnership with parents
- involves whole populations of children
- comprises primary, secondary and tertiary prevention activities

This is a statement of the aims of child health surveillance and what a surveillance program should include. We found this descriptive format to be conceptually unhelpful, because of the confusion between process issues and health issues, and because some health issues (eg growth, development) are included while others that might meet criteria for surveillance (eg vision) are not. We have therefore elected to use the following framework by Stone (1990) as our working definition.

Child Health Surveillance

‘Child health surveillance is the systematic and ongoing collection, analysis, and interpretation of indices of child health, growth, and development in order to identify, investigate and, where appropriate, correct deviations from predetermined norms’ (Stone 1990).

Surveillance is essentially about gathering and utilising data. In formulating specific definitions, it is also useful to consider surveillance at two levels – the individual and population levels.

Individual Surveillance

Individual surveillance focuses on a particular child, and will include gathering data from screening tests, physical examinations, discussions with parents and other caregivers, etc. This is also sometimes referred to as clinical surveillance.

Population Surveillance

Population surveillance focuses on groups or entire populations, and enables observation of changes and trends at a public health level. This is also sometimes referred to as monitoring.

3.3.2. APPRAISAL OF SURVEILLANCE

Following our appraisal of the available evidence of the effectiveness of screening tests and programs, we will consider how these programs might fit into a longitudinal surveillance policy. It is our intention to also evaluate, where available, any evidence for surveillance programs that are not based on screening. At the end of the review, we will attempt to pull all these threads together and make recommendations for overall screening and surveillance programs.

3.4. HEALTH PROMOTION

Definitions of Health Promotion cover wide ranging, multifaceted interventions and approaches to population health and wellbeing. An example is the commonly cited WHO Ottawa Charter for Health Promotion (1986).

‘The process of enabling people to increase control over and improve their health. It involves the population as a whole in the context of their everyday lives, rather than focusing on people at risk for specific diseases, and is directed toward action on the determinants or causes of health’.

Similarly, the previous NHMRC Review of Child Health Screening and Surveillance (NHMRC, 1993) considered Health Promotion to be: ‘Any combination of health education and related organisational, economic and environmental supports for behaviour of individuals, groups or communities conducive to health.’

Successful screening and surveillance activities should contribute to better health at a population level, and from that point of view, they fall within the realms of health promotion. However, the overall concept of health promotion and specific health promotion strategies are beyond the scope of this review, falling more naturally into the NHMRC “Phase 2” project.

REFERENCES

- Butler, J. 1989, "Child health surveillance in primary care: a critical review", Department of Health, London.
- Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests: Recommended Methods. Available at <http://som.flinders.edu.au/fusa/cochrane> last updated 6 June 1996.
- Hall, D.M.B. (Ed). 1996, "Health for all children (3rd ed.)", Oxford University Press, New York.
- Muir Gray, J.A. 1997, "Evidence-based healthcare: How to make health policy and management decisions", Churchill Livingstone, New York.
- National Health and Medical Research Council. 1993, "Review of child health surveillance and screening", Australian Government Publishing Services, Canberra.
- Stone, D.H. 1990, "Terminology in community child health – an urgent need for consensus", *Arch Dis Child*, vol.65, pp. 817-821.
- Commission on Chronic Illness. 1957 *Prevention of chronic illness*, Harvard University Press for The Commonwealth Fund, Cambridge (not sourced for this review).
- World Health Organisation. 1986 "Ottawa Charter for Health Promotion", *First International Conference on Health Promotion*, Ottawa, Canada.

4. PROGRAM ISSUES

For all screening programs, consideration will be given to the following:

4.1. HARM / BENEFIT

Screening programs, like any other intervention, have the potential for both benefit and harm. Screening aims to identify a previously unrecognised disease or condition, or those people who are at high risk of developing it. Because it involves testing large numbers of people to identify the small number who have or are likely to have the disease, screening may only benefit a small proportion of the population screened. It is essential therefore that the potential benefit is weighed against the possibility of harm (Peckham & Dezateuz 1998).

The benefits of early detection of disease appear self evident. In some instances it is possible to prevent a disease or condition completely (primary prevention) – for example detecting and treating squint in young children so as to prevent the development of amblyopia. It is also possible to prevent the clinical consequences of an established condition (secondary prevention), for example in conditions such as PKU and others screened for in the immediate postnatal period. Finally there is increasingly good evidence of the effectiveness of early intervention for conditions such as developmental delay and other developmental and behavioural problems (tertiary prevention), so a strong case can be made for the early identification of these problems.

Hall summarises the possible benefits of early detection under the following headings: parents value early diagnosis; improved outcome; improved quality of life for child and family; access to educational and social services (Hall 1996). However, none of these automatically follows screening – see chapter on Hearing impairment (Hind & Davis 2000).

One might argue that to screen for a condition which does not fulfil Wilson and Jungner criteria could be considered 'harmful'. Or, put another way, screening for a condition should only be considered to be of benefit if it needs these criteria (Hotzman 1997; Gray & Austoker 1998).

A program of population screening, like any prevention program, must have proven benefits that are large enough to outweigh any adverse effects. There are often major problems in determining the benefits of a screening or intervention program (Marshall 1996b). Furthermore, even where there is evidence for population based screening effectiveness, those participating in such programs have the right to decide whether the benefit to harm ratio is acceptable to them as individuals (Marshall 1996b).

If the quality of a screening program is low, then the adverse effects are increased and the benefits are reduced (Figure 1) while the reverse is true if the program is of high quality (Figure 2). Any benefits of a screening program may in fact be outweighed by the potential harms if an adequate level of quality is not achieved (Muir Gray 1997).

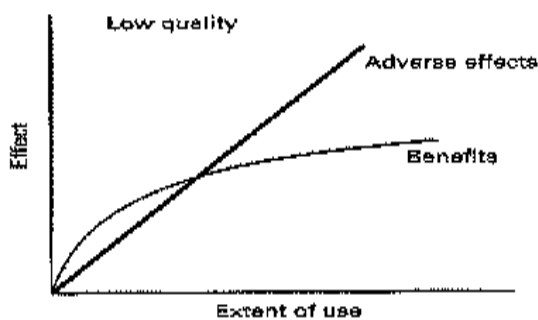


Figure 1*

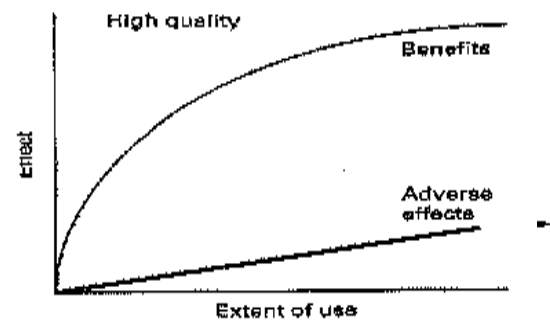


Figure 2*

* Figures 1 & 2 are from Muir Gray (1997), page 49.

Marshall (1996a) suggests that four questions need to be answered before embarking on any preventive intervention; these are also relevant for screening tests.

1. Is there any proven benefit from the intervention?
2. If there is, how great is it?
3. Are there any adverse effects of the intervention?
4. If there are, what are they, how serious are they and how often do they occur?

He also suggests there are four standard methods of reporting benefits (Marshall 1996a).

1. Relative reduction of morbidity or mortality rates (as a result of the screening test). This is usually reported as the relative reduction rate. For example, a screening test which leads to a specific treatment or intervention should lower the morbidity or mortality of the childhood population by a certain proportion compared to the situation where the screening test is not available. These rates are not always simple to calculate.
2. Absolute reduction of morbidity or mortality rate. This usually provides a better sense of the true benefit of a screening test, and for several conditions in child health (eg PKU) is easier to determine.
3. Number needed to treat. This refers to the number of individuals who would need to be detected by screening in order to avoid one adverse outcome due to screening. An example of this in childhood is screening to prevent amblyopia – how many children would need to be patched to avoid one case of irreversible amblyopia?
4. Total cohort mortality rate. This is much less relevant to child than to adult screening; this rate is often reported for screening for cancers in adults.

The other two (“number needed to treat”, and “total cohort mortality rate”) are less relevant to child health screening and surveillance.

There may be physical, psychological, social or ethical harms associated with screening tests (Marshall 1996c).

Physical harm can be the discomfort or pain associated with the test, for example, the pain of a heel prick to take blood for neonatal screening tests. It can be due to the investigations performed as part of the assessment process following an abnormal screening test (Austoker & Ong 1998) or from the treatment of detected abnormalities or diseases., or even to the treatment of a condition that is not present due to a false positive test result.

Psychological harm refers to the anticipated discomfort or perception of discomfort or harm, or the anxiety over the results of a screening test, or the false reassurance of a false negative screening test (Hall, Bobrow & Marteau 2000) or the anxiety of a false positive screening test until the diagnosis is proven to be unfounded (Sorenson, Levy, Mangione & Sepe 1984).

Social harm refers to the inconvenience that may be entailed by participating in a screening program, including time, travel, costs of loss of productive work time for parents, and so on.

Ethical harm is where informed consent is not obtained for participation in the screening program. This is of major concern in adult screening programs (O'Connor, Rostom, Fiset, Tetroe, Entwistle, Llewellyn-Thomas, Holmes-Rovner, Barry, Jones 1999; Austoker 1999) and for example, in antenatal screening. One could also argue about the ethics of screening programs that do not meet the criteria.

Perrin (1998) has also argued that it is unethical to screen for problems, referring especially to developmental, behavioural and psychosocial problems, if there are not the resources available to undertake diagnostic assessments and especially to provide treatment.

All of these considerations need to be addressed in deciding whether to introduce a new screening program. First the condition to be screened for should fulfil the Wilson and Jungner criteria (discussed in detail elsewhere in this report). In paediatrics, the properties of the screening test (especially false negative and false positive rates) would seem to be especially pertinent. A false negative result on antenatal screening for Down's Syndrome was found to have an adverse effect on parental adjustment two to six years after the birth of an affected child (Hall, Bobrow, Marteau 2000). In a survey of parents of infants with false positive neonatal screening tests for cystic fibrosis, feelings of “anxiety described as shock, denial, depression and anger” were described on learning of a positive screening test. The authors concluded that “psychosocial risks are associated with cystic fibrosis screening” (Tluczek, Mischler, Farrell, Fost, Peterson, Carey, Bruns & McCarthy 1992). Simply waiting for the results of retesting or definitive diagnostic testing also causes anxiety and ‘psychological harm’ (Sorenson et al. 1984). Bergman and Stamm, in a seminal study of children who had been investigated for a heart murmur subsequently shown to be innocent, found that even

years afterwards parents still worried about their child and tended to treat them differently (Bergman & Stamm 1967).

Screening is therefore not as straightforward as might seem from first glance. A consideration of potential harms and benefits is important in determining whether a screening test should be introduced. After the introduction of a screening program, harms and benefits need to be considered at an appropriate point in time and linked to decisions about program continuation. Finally, for some conditions harms and benefits (and therefore continuation or commencement of programs) may need to be reviewed as new therapy and/or detection methodologies alter the existing harm-benefit ratio.

REFERENCES

- Austoker, J., & Ong, G. 1998, "Do women who undergo further investigation for breast screening suffer adverse psychological consequences? A multi-centre follow up study comparing different breast screening result groups five months after their last breast screening appointment", *J Public Health Med*, vol. 20, pp. 396-403.
- Austoker, J. 1999, "Gaining informed consent for screening (Editorial)", *BMJ*, vol. 319, pp. 722-723.
- Bergman, A. B., & Stramm, S. J. 1967, "The morbidity of cardiac non-disease in school children", *N Engl J Med*, vol. 267, pp. 1008-1013.
- Gray, J.A., & Austoker, J. 1998, "Quality assurance in screening programs", *British Medical Bulletin*, vol. 54, pp. 983-992.
- Hall, D.M.B. (Ed). 1996, "Health for all children (3rd edition)", Oxford Medical Publications, New York.
- Hall, S., Bobrow, M., & Marteau, T.M. 2000, "Psychological consequences for parents of false negative results on prenatal screening for Down's syndrome; retrospective interview study", *BMJ*, vol. 320 pp. 407-412.
- Hind, S., & Davis, A. 2000, "Outcomes for children with permanent hearing impairment", in Seewald RC (ed) *A Sound Foundation Through Early Amplification; Proceedings of an International Conference* pp. 199-212 Phonak AG.
- Hotzman, N.A. 1997, "Genetic screening and public health (editorial)", *American Journal Public Health*, vol. 87, pp. 1275-1277.
- Marshall, K.G. 1996a, "Prevention. How much harm? How much benefit? 1. Influence of reporting methods on perception of benefits", *CMAJ*, vol. 154, pp. 1493-1499.
- Marshall, K.G. 1996b, "Prevention. How much harm? How much benefit? 2. Ten potential pitfalls in determining the clinical significance of benefits", *CMAJ*, vol. 154, pp. 1837-1843.
- Marshall, K.G. 1996c, "Prevention. How much harm? How much benefit? 3. Physical, psychological and social harm", *CMAJ*, vol. 155, pp. 169-176.
- Muir Gray, J.A. 1997, "Evidence-based healthcare: How to make health policy and management decisions", Churchill Livingstone, New York.
- O'Connor, A.M., Rostom, A., Fiset, V., Tetroe, J., Entwistle, V., Llewellyn-Thomas, H., Holmes-Rovner, M., Barry, M., & Jones, J. 1999, "Decision aids for patients facing health treatment or screening decisions: systematic review", *BMJ*, vol. 310, pp. 731-734.
- Peckham, C.S., & Dezateuz, C. 1998, "Issues underlying the evaluation of screening programs", *British Medical Bulletin*, vol. 54, pp. 767-778.
- Perin, E.C. 1998, "Ethical questions about screening", *Journal of Developmental and Behavioural Pediatrics*, vol. 19, pp 350-352.
- Sorenson, J.R., Levy, H.L., Mangione, T.W., & Sepe, S.J. 1984, "Parental response to repeat testing of infants with 'false-positive' results in a newborn screening program", *Pediatrics*, vol. 73, pp. 183-187.
- Tluczek, A., Mischler, E.H., Farrell, P.M., Fost, N., Peterson, N.M., Carey, P., Bruns, W.T., & McCarthy, C. 1992, "Parents' knowledge of neonatal screening and response to false positive cystic fibrosis testing", *Journal Dev Behav Pediatrics*, vol. 13, pp. 181-186.

5. METHODOLOGY

5.1. CONSULTATION STRATEGY

5.1.1. EXPERTS IN CHILD HEALTH

The project team is aware of the importance of consultation with other groups doing similar work, both in Australia and overseas, and with other key stakeholders. Several people who are international leaders in the field of child health screening and surveillance agreed to be members of the reference team for this project. Contact was made with experts around the world to seek advice the topics being considered and access to the latest information available. The following people were contacted in the preliminary stages of the review:

USA	Richard Wasserman MD (AAP & APA)
Canada	Emmett Francoeur MD (CPS) Charles Larson (Child Public Health Unit) Prof Ron Barr (Head, Child Development Unit, McGill University)
UK	Prof David Hall (President Elect, Royal College of Paediatrics and Child Health, UK) Dr Sarah Stewart-Brown (Director, Health Services Research Unit, Oxford University)
Netherlands	Prof Micha de Winter (Chair, Working party to review child health screening guidelines)
Switzerland	Jean Vuille
Australia	Prof David Henderson-Smart (Australian and New Zealand neonatal network)

We continued to liaise with these child health professionals, and also consulted with the other key individuals who agreed to be part of our consultation network (Prof Paul Dworkin, Prof Frances Page Glascoe, Dr. Aidan MacFarlane, Prof Neil Halfon, and so on). In addition, many of the key opinion leaders in Australia were members of the reference group (Dr. Neil Wigg, Dr. Garth Alperstein, A/Prof Gay Edgecombe, etc.). We also consulted with other key personnel, including Prof Graham Vimpani, Prof Allan Carmichael, and other members of the National Council of Community Child Health.

5.1.2. EXPERTS IN LITERATURE SEARCHING

We established links with several specialists in this field who offered their expertise to assist us in our search of the literature. We sought expert advice regarding general and on-line search strategies from Sue Shaw (Centre for Clinical Effectiveness, Monash University), Steve McDonald (Institute of Public Health, Monash University (Australasian Cochrane Collaboration)), and Judy Stoelwinder (Network Library Services Manager, Women's and Children's Health Care Network). All provided advice on appropriate MeSH headings for our on-line searches. They generously agreed to give us ongoing support and advice.

5.1.3. EXPERTS IN TOPIC AREAS

In addition to the experts in child health and the project reference team, we consulted local Australian experts in each of the topic areas being covered. Experts in each of the topic areas (minimum of two experts per topic) were identified and contacted. Each generously agreed to provide feedback on the draft topic documents in their area of expertise, to check the factual information in the documents and to direct us to any key work on the topic (published or unpublished) that we had not identified.

The purpose of consulting these experts was to ensure the background information provided on each topic (eg. clinical picture etc) was factually correct, not to provide expert opinion on the recommendations or conclusions reached.

5.2. LITERATURE RETRIEVAL METHODS

5.2.2. REVIEW OF EXISTING DATABASES OF REVIEWS

Our initial approach was to explore existing databases of reviews, particularly the Cochrane Database of Systematic Reviews (including registered titles, protocols, and completed reviews), the York Database of Abstracts of Reviews of Effectiveness (DARE), and the Cochrane Controlled Trials Register (CTR). We also conducted a search for reviews on Medline, HealthStar, CINAHL and ClinPSYC, using the appropriate MeSH headings and text words (listed under each topic in Section 7), and a modified version of the

systematic review search filter developed by [NHS Centre for Reviews and Dissemination](#), University of York. Where relevant reviews were at protocol stage, we contacted the authors for information.

5.2.3. SEARCH OF THE PEER-REVIEWED LITERATURE 1993-2000

Following our search for reviews, we conducted systematic searches of the published and unpublished literature by each topic using the appropriate MeSH headings and text words (listed under each topic in Section 7). In the first instance we searched major on-line databases, ie Medline, HealthStar (which is more geared towards the health policy, health management and health technology literatures than Medline), CINAHL and ClinPSYC.

Although we initially considered searching Embase (which has a strong coverage of European material) it was later decided following consultation with Judy Stoelwinder (Network Library Services Manager, Women's and Children's Health Care Network) that this was probably unnecessary. The Embase database has a strong pharmacology focus which was not particularly relevant to screening and surveillance. In addition, it was thought unlikely that Embase searches would result in detection of key articles that had not been picked up by:

- (1) our comprehensive search of four other databases as well as Cochrane, DARE and CCTR,
- (2) scanning of citations in articles we had already sourced,
- (3) our international reference group and local experts who were asked to inform us of any key articles we had not included.

We designed strategies to locate studies according to quality of evidence ratings, as defined in the NHMRC publication "Guidelines for the Development and Implementation of Clinical Practice Guidelines" (NHMRC 1999). All strategies described were designed and thoroughly tested by researchers and librarians skilled in search methods and evidence based appraisal. We also explored on-line search strategies developed by centres involved in evidence based medicine (including NHS Centre for Reviews and Dissemination & UK Institute for Health Sciences), which are designed to retrieve methodologically sound studies. Filters were added to allow targeting of specific methodologies and of fields of particular importance such as 'diagnosis' (including sensitivity and specificity) and 'screening' when searching for articles about a particular condition. Key words used for searches within individual domains were inserted as appropriate. A full list of key words (mapped to the appropriate MeSH headings) used for individual topics are detailed under each topic discussion in Section 7. We drew on the expertise of search specialists to assist us in the preparation of our final search strategies.

Initially we searched for evidence in the form of systematic reviews and/or meta-analyses. Following this, we searched for RCTs and controlled clinical trials (CCTs). However, because few child health screening and surveillance activities have been subjected to such rigorous testing, we proceeded to broader searches (including literature reviews and cohort studies). Strategies were developed into a tiered approach with the most precise (specific) search terms listed first to allow efficient sifting of detected evidence. For topics where good quality systematic review(s) were identified, further searches were limited to evidence not included in the systematic review (ie published after the systematic review).

Search Strategies are included in Appendix B.

5.2.4. LIMITED REVIEW OF PRE-1993 PEER-REVIEWED LITERATURE

The 1993 "Review of Child Health Surveillance and Screening" (NHMRC 1993) probably did not retrieve all relevant pre-1993 literature, especially papers published in 1992-93, those in the foreign language literature, and those cited on non-Medline sources. It also seems unlikely that the pre-1993 literature was appraised for quality and strength of evidence to the standard that is now required before a review can be accepted as "evidence-based", as these methods of appraisal have only been adopted on a wide scale in the last five years. We therefore re-searched the pre-1993 scientific literature where it seemed likely to contain important information and included pre-1993 evidence where appropriate.

5.2.5. REVIEW OF ELECTRONIC PUBLICATIONS AND WEBSITES

A comprehensive list of paediatric and health websites was examined for publications and information related to childhood screening and surveillance. From the comprehensive list, we identified websites suitable for more in depth searching (ie containing a search facility and containing child health information). These websites were then explored for information not available on the major databases (eg Cochrane, Medline).

The comprehensive list of websites examined are included in Appendix C.

5.2.6. REVIEW OF OTHER PUBLICATIONS

Some high-quality publications from other organisations were of direct relevance to this review. For example, the NHS Centre for Reviews and Dissemination has completed a number of systematic reviews (eg pre-school hearing, speech, language and vision screening and Child health surveillance: an evaluation of visual screening) and others are in progress.

Publications reviewed included:

- NHS Centre for Reviews and Dissemination (CRD) reports
- Effectiveness Matters
- Effective Health Care Bulletins
- Journals of secondary publication, including ACP Journal Club, Evidence-Based Medicine, and Evidence-based Health Policy and Management Journal
- NHS R&D Health Technology Assessment Programme reports
- Registers of published studies, including the Register of Cost-Effectiveness Studies
- Reports of the Centre for Evidence-Based Child Health, Institute of Child Health, London

5.2.7. REVIEW OF UNPUBLISHED RESEARCH

As far as possible, we identified relevant research in progress using registers including the UK-based NHS National Research Register and the US Database of current health services research (HSRProj) as well as informal networks within and outside the Reference Group members and local experts.

5.3. EVALUATION

Where appropriate, the randomised controlled trial (RCT) is considered the 'gold standard' in assessing the value of interventions. However, only a small proportion of information about screening and surveillance programs was available from RCTs.

The validity of the information was assessed by considering the level of evidence (a measure of bias in study design) and the quality of evidence (a measure of bias in the conduct of the study) supporting that activity. The strength, magnitude and relevance of the clinical effect was also considered. Definitions of these terms are summarised in Table 1.

In this review, we attempted to assess the evidence within the format outlined in the NHMRC "Guidelines for the Development and Implementation of Clinical Practice Guidelines" (Table 2) (NHMRC, 1999) where clinical intervention studies were available. However, few intervention studies were identified in this review. Similarly for many areas of screening no good quality studies of program effectiveness are available. However, there may be good evidence about how well a specific diagnostic test performs in detecting a condition which may support or not support the introduction of a screening program. Where no evidence of program effectiveness was available, we assessed the quality of evidence of available diagnostic tests for each condition.

Table 1. Assessing the evidence (NHMRC 2000)

Strength of evidence	
Level	The study design used, as an indicator of the degree to which bias has been eliminated by design (see Table 2).
Quality	The methods used by investigators to minimise bias within a study design.
Statistical precision	The <i>p</i> -value or, alternatively, the precision of the estimate of the effect (as indicated by the confidence interval). It reflects the degree of certainty about the existence of a true effect.
Size of effect	The distance of the study estimate from the ‘null’ value and the inclusion of only clinically important effects in the confidence interval.
Relevance of Evidence	The usefulness of the evidence in clinical practice, particularly the appropriateness of the outcome measures used.

Table 2. Levels of evidence (NHMRC 1999)

Level of Evidence	Intervention
I	Evidence obtained from a systematic review of all randomised controlled trials
II	Evidence obtained from at least one properly-designed randomised controlled trial
III-1	Evidence obtained from well-designed pseudo-randomised controlled trials (alternative allocation or some other method)
III-2	Evidence obtained from comparative studies with concurrent controls and allocation not randomised (cohort studies), case-control studies, or interrupted time series with a control group
III-3	Evidence obtained from comparative studies with historical control, two or more single-arm studies, or interrupted time series without a parallel control group
IV	Evidence obtained from case series, either post-test or pre-test and post-test
V	<i>The current tables exclude expert opinion and consensus from an expert committee as they do not arise from scientific investigation</i>

Prior to the era of evidence based practice, criteria were developed by various authors for appraisal of screening programs. While many of the principles of earlier guidelines are still relevant and appropriate, the fundamental criteria for evaluating current best practice is evidence of effectiveness. In a sense, quantifying properties such as sensitivity and specificity is irrelevant if there is evidence that a program is or is not effective in improving health outcomes, though such information may help in understanding why/why not or how to potentially improve a program (Muir Gray 1997).

For all conditions we assessed the suitability for screening using criteria for the appraisal of screening tests developed by Cochrane and Holland (1971) and principles to consider in the appraisal of screening programs

developed by Wilson and Jungner (1968). Various groups have subsequently modified these criteria, but these remain the most widely known and presented.

Characteristics of a Screening Test (Cochrane & Holland 1971)

- Simple, quick and easy to interpret: capable of being performed by paramedics or other personnel
- Acceptable to the public, since participation in screening programs is voluntary
- Accurate: ie gives a true measure of the attribute under investigation
- Repeatable: this involves the components of observer variability, both within and between tests, subject variability, and test variability
- Sensitive: this is the ability of a test to give a positive finding when the individual screened has the disease or abnormality under investigation
- Specific: this is the ability of the test to give a negative finding when the individual does not have the disease or abnormality under investigation

Criteria for a Screening Program (Wilson & Jungner 1968)

- The condition sought should be an important health problem
- There should be an accepted treatment for patients with recognised disease
- Facilities for diagnosis and treatment should be available
- There should be a recognisable latent or early symptomatic stage
- There should be a suitable test or examination
- The test should be acceptable to the population
- The natural history of the condition, including development from latent to declared disease, should be adequately understood
- There should be an agreed policy on whom to treat as patients
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
- Case-finding should be a continuing process and not a 'once and for all' project

5.4. ASSESSING THE EVIDENCE

Our emphasis was on locating the best literature, not the most literature. Therefore, we initially considered evidence from systematic reviews and randomised controlled trials (RCT). Where recent high quality systematic reviews were identified or where guidelines based on good evidence were available we did not repeat this process. However, findings of these papers were reviewed in light of local conditions and an appraisal made of the quality of the review/guideline using standard review criteria (as detailed for example in the JAMA guides "How to use an overview" and "How to use clinical practice guidelines").

We evaluated the evidence using the criteria outlined above in 'Evaluation' (Section 6.3). Where there was no information as to the effectiveness of a screening program but the condition is deemed important, we sought evidence of the properties of individual screening tests.

As well as assessing the evidence of effectiveness, we addressed other issues that must be considered in the evaluation of a screening program. We also reviewed 'numbers needed to test' in order to find one case; information about any potential adverse effects of screening; the training and resource implications; costs; etc for each condition.

5.4.1. ASSESSING THE QUALITY OF SYSTEMATIC REVIEWS.

Wherever possible we considered evidence from published systematic reviews. However, it cannot be assumed that the reviews have been conducted and reported without flaws and the possibility of bias and other limitations in the review process must be considered. In order to do this in a systematic way, we developed a system for evaluating systematic reviews and for reporting on their quality.

The reviews were graded according to whether or not they satisfy three major requirements. These are:

1. A clearly defined research question and specified criteria for inclusion of studies in the review
2. Evidence of a comprehensive search of the literature

3. The presence of specified critical appraisal criteria to assess the validity of studies reviewed, and whether the findings of the critical appraisal have been taken into consideration in the conclusions and recommendations of the review

Each of these requirements was divided into three categories. Those reviews represented by the first category were considered to be of high quality in terms of this particular requirement, those graded to the second category were considered to have some limitations, and those rated in the third were considered to be of unsatisfactory quality.

For simplicity and clarity throughout the document we use a systematic description of whether or not each review satisfied these requirements. For each category, the quality of the review is described as:

- A high quality
- B some limitations
- C unsatisfactory quality

1. *Clearly defined question and inclusion criteria*

It is very important that a systematic review has been conducted using a clearly defined research question. Equally important is the use of explicit inclusion and exclusion criteria for the studies under review.

- A (high quality)** A clear aim or research question **AND** a list of clear inclusion criteria was reported
- B (some limitations)** A clear aim or research question **OR** a list of clear inclusion criteria was reported
- C (unsatisfactory quality)** No aims or clear research questions were reported. It was unclear what had been done.

2. *Comprehensive search*

Search strategies for systematic reviews should be comprehensive and clearly stated. There should be evidence that the authors have sought information extensively beyond the electronic databases.

- A (high quality)** Electronic databases and one or more of the following were used: the 'grey' literature, internet sources, conference proceedings, hand searches, reference lists, technical reports, experts, theses, etc **AND** the search was not restricted to the English language literature only
- B (some limitations)** More than 1 electronic database was used **OR** the search was restricted to the English language literature only
- C (unsatisfactory quality)** Only 1 major electronic database was used (eg Medline)

3. *Critical appraisal of the validity of studies reviewed*

- A (high quality)** Explicit criteria for critical appraisal were stated and the findings of that appraisal were considered in the conclusions and recommendations
- B (some limitations)** Critical appraisal was undertaken but explicit criteria were not stated
- C (unsatisfactory quality)** No critical appraisal of the studies reviewed was reported

4. *Consistency of results*

Systematic reviews may be of high quality, but if only a few studies were included the value of the results obtained will be questionable. The heterogeneity of the results is also relevant. Where there is agreement between studies on the size and direction of effect, this homogeneity will suggest confidence in the results. Where results are more heterogeneous, the effect is less clear. We supplement an assessment of the quality of each review with the number of studies included and the degree of confidence in the results.

- A (high quality)** Homogeneity of results (ie all studies indicate positive/negative effects of similar magnitudes)
- B (some limitations)** Heterogeneity of size of effect but trend obvious (ie all studies indicate positive/negative effects, but of differing magnitudes)
- C (unsatisfactory quality)** Heterogeneity of direction of effect (ie some studies indicate positive effect, others negative)

As the number of studies included in a review impacts on the value of the results, the number of studies reviewed will be added to the assessment of the quality of the systematic review.

Examples

The discussion of the review findings is accompanied by four ratings of the review on the above measures of quality and the degree of homogeneity of the results. In addition a number indicating the number of studies considered in the review is included in the rating.

For example:

Smith A, Jones A and Black A. Issues in screening. *Screening Journal* 1:1-10, 1999

Clearly defined question and inclusion criteria

C (unsatisfactory quality) This review did not have a clearly stated research question or criteria for inclusion of studies.

Comprehensive search

B (some limitations) There are some limitations in the search methods used. More than one electronic database has been accessed but no attempt was made to search for information outside this framework. Alternatively, the limitation may be that the authors explicitly stated that the search was restricted to the English language literature only, potentially overlooking significant research.

Critical appraisal of the validity of studies reviewed

A (high quality) The studies included in the review were appraised using explicitly stated criteria and the findings of the appraisal were considered in the conclusions and recommendations.

Confidence in results

B (some limitations) All the studies reviewed had findings that indicated the same direction of effect (eg all positive or all negative) but the size of the effects differed. The interpretation of this evidence needs to be considered given the limitation that only 3 studies were included, and the size of the effect may have differed widely between them.

Number of studies in the review

3 (Only three studies were included in this review)

5.4.2. ASSESSING THE QUALITY OF RANDOMISED CONTROLLED TRIALS.

Where systematic reviews were not available, we next considered evidence from randomised controlled trials. However again, it cannot be assumed that the trials have been conducted and reported without flaws and the possibility of bias and other limitations in the trial process must be considered. In order to do this in a systematic way, we developed a system for evaluating randomised controlled trials and for reporting on their quality, based on a scale developed by Jadad and colleagues (Jadad, Moore, Carroll, Jenkinson, Reynolds, Gavaghan & McQuay 1996).

The trials were graded according to whether or not they satisfy three major requirements. These are:

1. The use of appropriate randomisation techniques and concealment of allocation
2. Blinding of outcomes (where possible/appropriate)
3. A clear description of withdrawals and dropouts ie all subjects entering the trial are accounted for at the completion of the trial.

Similar to the assessment of reviews, the first two requirements were divided into three categories. Those trials represented by the first category were considered to be of high quality in terms of this particular requirement, those graded to the second category were considered to have some limitations, and those rated in the third were considered to be of unsatisfactory quality. The third requirement was divided into two categories. Those trials represented by the first category were considered to be of high quality, having presented a documented account of all subjects at the completion of the trial, and those rated in the second category were considered to have an unsatisfactory account of subjects at the completion of the trial.

As with the assessment of reviews, for simplicity and clarity throughout the document we use a systematic description of whether or not each review satisfied these requirements. For each category, the quality of the review is described as:

A high quality

- B some limitations
- C unsatisfactory quality

1. *Randomisation and concealment of allocation*

It is important that randomised controlled trials (RCTs) use an appropriate method of randomisation and conceal the allocation of subjects to groups.

- A (high quality)** the method of randomisation is appropriate and allocation was concealed eg table of random numbers, computer generated randomisation, sealed opaque envelopes
- B (some limitations)** the trial is stated as randomised but the method of randomisation and/or allocation concealment is not clearly described
- C (unsatisfactory quality)** method of randomisation is not appropriate (quasi-randomised) eg subjects were randomised by their date of birth, alternating consecutive subjects to alternate groups OR subjects were randomised with inadequate concealment eg read off an open table of random numbers, on-the-spot coin flip

2. *Blinding of outcomes*

Although this criterion is more appropriate to RCTs of treatments, where possible there should be blinding of outcomes.

- A (high quality)** double blind and description of blinding clear
- B (some limitations)** stated as double blind or use of placebo or dummy mentioned, but blinding not explicitly described
- C (unsatisfactory quality)** not double blind or inappropriate blinding

3. *Withdrawals and dropouts*

All subjects entering the trial should be accounted for at its completion, therefore it is important that any withdrawals or dropouts are described.

- A (high quality)** all subjects entering the trial were accounted for at its completion
- B (some limitations)** some description of withdrawals and dropouts
- C (unsatisfactory quality)** no description of withdrawals and dropouts

As the number of subjects included in a trial impacts on the value of the results, the number of subjects randomised will be added to the assessment of the quality of the RCT.

5.4.3. ASSESSING THE QUALITY OF OTHER STUDIES.

For topics where systematic reviews or RCTs were unavailable, studies leading to lower levels of evidence were reviewed. These reports were critically appraised using NHMRC evaluation guidelines and principles outlined in the JAMA 'User's Guides to the Medical Literature'. The criteria in Table 1 Assessing the evidence (NHMRC 2000) were utilised in overall evaluation of these materials.

5.5. EXCLUSION CRITERIA

Studies that do not satisfy our criteria for evaluation will be excluded. In situations where the study may have particular historical importance or other significance to a screening program, it will be cited and the reasons for exclusion will be recorded.

REFERENCES

Cochrane, A. L., & Holland, W. W. 1971, "Validation of screening procedures", *British Medical Bulletin*, vol.27, pp.3-8.

Jadad, A.R., Moore, A., Carroll, D., Jenkinson, C., Reynolds, J.M., Gavaghan, D.J., & McQuay, H.J. 1996, "Assessing the quality of reports of randomised clinical trials: is blinding necessary", *Controlled Clinical Trials*, vol. 17, pp.1-12.

Muir Gray, J.A. 1997, "Evidence-based healthcare: How to make health policy and management decisions", Churchill Livingstone, New York.

National Health and Medical Research Council. 2000, "How to use the evidence: assessment and application of scientific evidence", Commonwealth of Australia, Canberra.

National Health and Medical Research Council. 1999, "Guidelines for the Development and Implementation of Clinical Practice Guidelines", Commonwealth of Australia, Canberra.

National Health and Medical Research Council. 1993, "Review of child health surveillance and screening", Australian Government Publishing Services, Canberra.

Wilson, J. M. G., & Jungner G. 1968 "Principals and practice of screening for disease", World Health Organisation, Geneva.

6. DISEASES / CONDITIONS

6.1. TOPIC LIST

Specific disorders

1. Atlanto-axial instability
2. Congenital adrenal hyperplasia
3. Congenital heart disease
4. Congenital hypothyroidism
5. Cystic fibrosis
6. Developmental dysplasia of the hip
7. Hearing – *a.* conductive hearing impairment
 - *b.* permanent childhood hearing impairment
8. Hypertension
9. Iron deficiency
10. Lead toxicity
11. Phenylketonuria
12. Scoliosis
13. Undescended testes
14. Urinary abnormalities
15. Vision

Normal progress

16. Dental health
17. Development
18. Language
19. Height
20. Weight

6.2. SELECTION CRITERIA

All topics specified by NHMRC on the tender document were included.

6.3. DISCUSSION OF EACH DISEASE / CONDITION

Each topic is discussed in detail and a summary of the findings for all topics is included in tabular form.

6.3.1. ATLANTO-AXIAL INSTABILITY (in Down Syndrome)

6.3.1.1. BACKGROUND

Clinical picture

Atlanto-axial instability is also known as atlanto-axial subluxation or C1-C2 subluxation. It is an instability, partial dislocation or subluxation of the first and second cervical vertebrae (C1 and C2, or “atlas” and “axis”) caused by slackness of the transverse ligament or abnormalities such as hypoplasia, malformation, or complete absence of the odontoid (Committee on Sports Medicine 1984). Atlanto-axial instability can lead to full dislocation of the atlanto-axial joint (Hungerford, Akkaraju, Rawe, & Young 1981).

Many cases of atlanto-axial instability are asymptomatic (Pueschel, Herndon, Gelch, Senft, Scola, & Goldberg 1984). However, some individuals present with symptoms including neck pain or stiffness, head tilt, problems with bowel and/or bladder function, torticollis, difficulty walking, muscle weakness, gait disturbance, increasing clumsiness and lack of coordination, spasticity, hyperreflexia, clonus, toe-extensor reflex and other neurological signs (Committee on Sports Medicine 1984; Pueschel et al 1984). It appears that the presence and intensity of symptoms may be related to the joint interval; individuals with a wide joint interval appear to experience more intense symptoms whilst those with a narrow joint interval are more likely to be asymptomatic (Pueschel et al. 1984). Generally, nearly all individuals with a joint interval less than 7mm are asymptomatic and of seven patients with an interval ≥ 7 mm, all exhibited neurologic symptoms (Pueschel et al. 1984). However, the relationship between joint interval and symptoms has not been reliably quantified.

Atlanto-axial instability is most common in individuals with Down Syndrome but has also been reported in individuals with rheumatoid arthritis, dwarfism and those with abnormalities of the odontoid process of C2 (Committee on Sports Medicine 1984). Individuals with Down Syndrome commonly suffer from generalised poor muscle tone and joint laxity which is thought to partially explain the high prevalence of atlanto-axial instability in this population.

As the majority of cases of atlanto-axial instability are seen in children with Down Syndrome, and all of the research identified was related to children with Down Syndrome, this review will relate only to children with Down Syndrome.

Natural History

The natural history of this disorder is unclear (Pueschel, Scola, & Pezzullo 1992). Onset appears to occur most often in childhood (mean age of onset of 40 studied cases was 10.5 years (SD 5.6 years)), but cases have been documented where diagnosis based on the onset of symptoms has occurred at age 45 years (Pueschel et al. 1984). The intensity of symptoms appears to be related to the size of the joint interval, and it appears that in some cases the joint interval can become wider over time (Pueschel et al. 1984). However a direct relationship between the time since onset and severity (size of joint interval and intensity of symptoms) has not been established.

Prevalence

It is estimated that between 10% and 40% of individuals with Down Syndrome experience atlanto-axial instability, although incidence estimates differ depending on age and the criterion used to diagnose the condition (Newton 1998). Atlanto-axial instability occurs more commonly in females than males, with an estimated ratio of 2.3:1, and onset is more common in childhood, although cases have been detected in middle aged adults (Pueschel et al. 1984).

Genetics relevant to screening issues

Down Syndrome results from trisomy of chromosome 21. However no direct genetic contribution has been identified for atlanto-axial instability.

Diagnosis

Diagnosis involves radiological detection of larger than normal space between the odontoid (bone that passes from C2 into C1) and the anterior arch of the C1 vertebra (Committee on Sports Medicine 1984). The criterion for diagnosing atlanto-axial instability is based on the shortest distance between the inferior aspect of the anterior atlas arch and the anterior aspect of the odontoid process (referred to as the atlanto-dens interval) in flexion, extension or neutral views of the spine. The shortest atlanto-dens interval required for diagnosis varies. Generally children should have an atlanto-dens interval of less than 3.5mm in flexion (Wheless 1999). In children with Down Syndrome the atlanto-dens interval is variously considered normal up to 4.4mm (Newton 1998) or 5mm (Wheless 1999). However, as stated above, symptoms are usually not present in cases where the atlanto-dens interval is less than 7mm (Pueschel et al. 1984).

Treatment/Management

Treatment of individuals with Down Syndrome and atlanto-axial instability is based on the size of the atlanto-dens interval and the intensity of symptoms. Children with an atlanto-dens interval considered abnormal but less than 7mm are usually asymptomatic and are often managed by avoiding certain activities which involve high impact flexion loading on the cervical spine, rather than active treatment (Wheless 1999). However, there is no evidence that avoidance of certain activities is of greater benefit than no treatment (Cremers, Bol, de Roos & van Gijn 1993). In addition avoidance can lead to missed social opportunities and feelings of social isolation for the child.

Where the atlanto-dens interval is between 7mm and 9mm treatment normally consists of monitoring, cervical orthosis and avoidance of contact sports and other high risk activities. Children with severe instability where the atlanto-dens interval is 10mm or more usually receive surgical treatment (Wheless 1999). In one study examining 404 patients with Down Syndrome, 59 (14.6%) were diagnosed with atlanto-axial instability, but only six (1.5%) were symptomatic and treated surgically (Pueschel & Scola 1987).

The main aim of surgical treatment is to reduce as much as possible the atlanto-axial subluxation and stabilise the upper section of the cervical spine to prevent additional spinal cord injury. Surgery appears to be effective for patients with recent onset of neurological symptoms. However for those whose symptoms have persisted over a number of years prior to surgery or who display marked neuronal damage, surgery appears to have little or no effect. As the natural history of the disorder is unclear, there is contention over treatment of asymptomatic atlanto-axial instability patients. Some doctors advocate that all patients should receive surgery irrespective of symptoms or the size of the joint interval and some believe that surgery is unwarranted for asymptomatic patients (Pueschel et al. 1984).

Prevention

None known.

How might screening reduce the burden of suffering?

As the natural history of atlanto-axial instability is unclear, the impact of screening on the burden of suffering is not apparent. Some cases of atlanto-axial instability experience progressive neurologic deterioration (Giblin & Micheli 1979) which may be avoided if detected and treated early. Once symptoms of atlanto-axial instability are obvious, irreversible spinal cord damage can result, but is a rare complication. It is possible that early detection of atlanto-axial instability, allowing monitoring of asymptomatic patients for the appearance of symptoms, may be beneficial to prevent such complications. Dislocation, although also very rare, can result in death (Hungerford et al. 1981) which again may be avoidable if the condition is identified and treated.

Due to a small number of individual cases of dislocation while playing sport, participation in sporting activities is thought to be dangerous for individuals with atlanto-axial instability, potentially causing complete dislocation and neurological damage. This has led to restrictions for some activities. For example, children with Down Syndrome are now required to be screened for atlanto-axial instability prior to competing in the Special Olympics (Committee on Sports Medicine and Fitness 1995) and are often advised to avoid particular sports and activities thought to increase risk of dislocation for children with atlanto-axial instability. However, a randomised controlled trial of Down Syndrome children with an atlanto-axial distance 4-6.5cm, randomised to a normal sport and activities group or a group advised to avoid "risky" sports and movements, found no differences in neurological signs or changes in atlanto-axial distance between the groups after one year (Cremers et al. 1993).

6.3.1.2 EVIDENCE

Tests

All tests require some form of clinical interpretation, and as such are open to variation in accuracy dependent upon the person interpreting the screen result. There is no 'gold standard' test for the detection of atlanto-axial instability.

Radiography

Most screening is done using radiographs of the upper spine during flexion (chin towards the chest) and extension (head back). Sometimes radiographs are also taken when the child's neck is in the neutral position. As part of a prospective study of 135 children aged 6-14 years with Down Syndrome, 19 children received two consecutive radiographs to screen for atlanto-axial instability. Five children (26%) were diagnosed with atlanto-axial instability based on the first radiograph and four children (21%) based on the second radiograph. However, only three of these children were diagnosed on both occasions (Selby, Newton, Gupta & Hunt 1991).

Other

Polytomographic myelography and computed tomography can also be used to diagnose atlanto-axial instability. However, we could identify no information reporting on the effectiveness of these methods for this condition.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971		Criteria for a Screening Program Wilson & Jungner 1968	
Simple, quick and easy to interpret	X	Important health problem	✓
Acceptable to the public	✓	Accepted treatment	X
Accurate	?	Facilities for diagnosis and treatment	✓
Repeatable	X	Latent or early symptomatic stage	?
Sensitive	X	Suitable test or examination	?
Specific	X	Test acceptable to the population	✓
		Natural history adequately understood	X
		Agreed policy on whom to treat	X
		The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

We found no reports of formal screening programs for atlanto-axial instability.

Cost effectiveness

No information on the cost-effectiveness of screening for atlanto-axial instability was identified.

Quality of evidence

No evidence was identified for atlanto-axial instability screening programs, and only one study of screening tests was identified. The information identified was of a low level of evidence. There have been no prospective follow up studies which assess the value or outcome of screening programs. This is in part due to the relative rarity of the condition (particularly identified in children), the large number of asymptomatic cases, the diversity in type and severity of symptoms reported for those who are symptomatic, and the rarity of adverse outcomes.

Generalisability of evidence

The issues discussed here relate only to screening children with Down Syndrome. There have been no suggestions to screen other populations.

6.3.1.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening.

Comments:

Little is known about:

- the natural history of atlanto axial instability
- prediction of who will suffer adverse outcomes
- rate of progression, if any

Therefore, it is impossible to weigh the harms vs the possible benefits of screening or to make sensible recommendations about timing and frequency of such screening.

6.3.1.4. RECOMMENDATIONS

Implementation of a formal screening program for atlanto-axial instability in children with Down Syndrome is not recommended.

6.3.1.5. FURTHER RESEARCH

- Natural history of atlanto-axial instability in Down Syndrome
- Reduction in burden of disease through preventive management
- Accuracy of radiograph interpretation
- Harms (particularly quality of life) related to banning from particular activities

6.3.1.6. SEARCH TERMS

Atlanto-axial joint

Joint instability

Occipital bone

Atlanto-axial instability

6.3.1.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Dinah Reddihough – Director, Child Development and Rehabilitation, Royal Children’s Hospital, Melbourne

Doug Bryan – Divisional Director, Community Orientated Paediatric and Adolescent Services, Royal Children’s Hospital, Melbourne

Ian Torode – Orthopaedic surgeon, Department of Orthopaedics, Royal Children’s Hospital, Melbourne

David Hall – Reference Group

Katrina Williams – Reference Group

6.3.1.8 REFERENCES

Committee on Sports Medicine 1984, "Atlantoaxial instability in Down Syndrome", *Pediatrics*, vol. 74,no. 1, pp. 152-154.

Committee on Sports Medicine and Fitness 1995, "Atlantoaxial instability in Down Syndrome: Subject Review (RE9528)", *Pediatrics*, vol. 96,no. 1, pp. 151-154.

Creemers, M. J., Bol, E., de Roos, F., & van Gijn, J. 1993, "Risk of sports activities in children with Down's syndrome and atlantoaxial instability", *Lancet*, vol. 342,no. 8870, pp. 511-514.

Giblin, P. E. & Micheli, L. J. 1979, "The management of Atlanto-axial Subluxation with Neurologic Involvement in Down's Syndrome: A Report of Two Cases and review of the Literature", *Clinical Orthopaedics and Related Research*, vol. 140, pp. 66-71.

Hungerford, G. D., Akkaraju, V., Rawe, S. E., & Young, G. F. 1981, "Atlanto-occipital and atlanto-axial dislocations with spinal cord compression in Down's Syndrome: a case report and review of the literature", *British Journal of Radiology*, vol. 54, pp. 758-761.

Newton, R. W. Atlanto-axial instability in Down Syndrome. DownsNet Online Library . 1998. Last updated 6-1-2000. http://www.downsnet.org/library/news_update/atlanto-axial/

Pueschel, S. M., Herndon, J. H., Gelch, M. M., Senf, K. E., Scola, F. H., & Goldberg, M. J. 1984, "Symptomatic Atlantoaxial subluxation in persons with Down syndrome", *Journal of Pediatric Orthopedics*, vol. 4,no. 6, pp. 682-688.

Pueschel, S. M. & Scola, F. H. 1987, "Atlantoaxial instability in individuals with Down syndrome: epidemiologic, radiographic, and clinical studies", *Pediatrics*, vol. 80,no. 4, pp. 555-560.

Pueschel, S. M., Scola, F. H., & Pezzullo, J. C. 1992, "A longitudinal study of atlanto-dens relationships in asymptomatic individuals with Down syndrome", *Pediatrics*, vol. 89,no. 6 Pt 2, pp. 1194-1198.

Selby, K. A., Newton, R. W., Gupta, S., & Hunt, L. 1991, "Clinical predictors and radiological reliability in atlantoaxial subluxation in Down's syndrome", *Arch Dis Child*, vol. 66,no. 7, pp. 876-878.

Wheless, C. R. Wheless' Textbook of Orthopaedics Online. Gurvinder S. Deol, Jan Van Der Bauwhede, and Eugene Sherry. Wheless' Textbook of Orthopaedics Online. 1999. <http://www.medmedia.com/Welcome.html>

6.3.2. CONGENITAL ADRENAL HYPERPLASIA

6.3.2.1. BACKGROUND

Clinical picture

Congenital Adrenal Hyperplasia (CAH) is a group of autosomal recessive disorders leading to abnormal adrenal function. It is also known as adrenogenital syndrome, and the terms macrogenitosomia praecox in males and pseudohermaphroditism have also been used synonymously with CAH. The type and severity of the disorder depends on the genetic mutation, the enzyme affected and the degree of enzyme insufficiency. Defects in mineralocorticoid synthesis results in electrolyte imbalances, hypo or hypertension, and syncope. Disorders of electrolyte metabolism may be life threatening in the neonate. The 'salt wasting' forms of CAH result in hyponatraemia, hypovolaemia, hyperkalaemia and hypotension. Abnormalities of glucocorticoid production result in an excess of androgenic steroids causing short stature, early puberty, acne, and virilisation and infertility in females. Milder, non-classic forms of CAH present later in life.

Deficiency of 21-hydroxylase (21OHD)

This is the most common form of CAH, accounting for 90% of cases and results in a reduced ability to synthesize cortisol and aldosterone. The classic form results in a salt-wasting crisis and female pseudohermaphroditism in the infant and postnatal virilisation as the child grows. The non-classic form presents later in life with precocious or disordered puberty, menstrual irregularity, hirsutism, acne and infertility.

Deficiency of 11 β -hydroxylase

This is very similar to the deficiency of 21-hydroxylase with 2 exceptions. Salt wasting is not a feature in the infant, and hypertension is usually present.

Deficiency of 17 α -hydroxylase

This deficiency results in male pseudohermaphroditism, hypertension and lack of sexual maturity.

Deficiency of 3 β -hydroxysteroid dehydrogenase.

The classic form results in a salt-wasting crisis, male and female pseudohermaphroditism in the infant and precocious or disordered puberty. The non-classic form presents later in life with precocious or disordered puberty, menstrual irregularity, hirsutism, acne and infertility.

Lipoid congenital adrenal hyperplasia

This is the most severe form of CAH and affected individuals cannot synthesis steroid hormones. This results in all infants being phenotypic females (male pseudohermaphroditism) with a severe, potentially life threatening salt-wasting crisis.

Prevalence

Classical CAH is rare and reports of incidence vary. Several studies cite an incidence rate of 1 in 14-17,000 newborns (Balsamo, Cacciari, Piazzini, Cassio, Pirazzoli, & Zappulla 1996; Brosnan, Brosnan, Kemp, Domek, Jelley, Blackett, & Riley 1999; Deaton, Glorioso, & McLean 1999; Fitness, Dixit, Webster, Torresani, Pergolizzi, Speiser, & Day 1999) while others suggest a lower rate of 1 in 20-23,000 (Cutfield & Webster 1995; Kwon & Farrell 2000). Milder forms are more common, but the true incidence is not known and is suggested to be between 1 in 100 and 1 in 1,000 (Deaton, Glorioso, & McLean 1999; Fitness et al. 1999; Thilen, Nordenstrom, Hangenfeldt, van Dobein, Guthenberg, & Larsson 1998). Deficiency of 21-hydroxylase is the most common, affecting 90% of patients (Deaton, Glorioso, & McLean 1999) (Al Saedi, Dean, Dent, Stockl, & Cronin 1996). Of these, 75% have the salt wasting variant (SW) and the remainder have the simple virilising form (SV) (Brosnan et al. 1999).

All references to CAH cite the issue of the potentially life-threatening salt-wasting crisis, however none mention the mortality rate and in the literature sourced for this review there was only one report of deaths from SW. In the period 1969-86 prior to the introduction of screening in Sweden, 2 boys died due to an adrenal crisis in the neonatal period.

Genetics relevant to screening issues

CAH due to 21OHD is an autosomal recessive disease affecting the gene CYP21. These abnormalities involve misalignments, deletion and other sequence aberrations, and many infants have combinations of mutations. Variations in CYP21 mutations have been linked to the severity of clinical disease (Nordenstrom, Thilen, Hagenfeldt, Larsson, & Weddell 1999). The 'Null' mutations are associated with the most severe forms of CAH, with salt loss and female virilisation. The '12 splice' mutation is less severe, and although associated with severe female virilisation, not all children have salt loss. The 'Ile¹⁷²Asn' mutation is linked

with the simple virilising phenotype and only 10% of children have salt wasting. The 'Val²⁸¹Leu' abnormality is associated with mild, late onset symptoms. Genotyping of CYP21 mutations could be used for prediction of clinical outcome in CAH patients (Nordenstrom et al. 1999). There are other variants, but they are rare.

Diagnosis

Biochemical tests

Increased levels of hormones are detected in the urine and blood of affected individuals. The particular hormones found to be in excess are related to the relevant enzyme deficiency. In the most common form of CAH, 21OHD, there is a marked increase in the serum concentration of 17 α -hydroxyprogesterone (17 OHP). This can be measured by enzyme immunoassay, radioimmunoassay (RIA) or fluoro-immunometric assay (FIA).

There is some debate in the literature about the appropriate cut-off values for what constitutes a positive test; what is normal; and what is the watershed between the two that requires repeat testing or further investigation. Preterm and low birth weight babies have markedly higher levels of 17OHP than term babies and cut off levels need to be adjusted for this group.

Genetic tests

CYP21 genotyping can be carried out on venous blood samples and is suitable for detecting gene mutations in the neonatal heel prick sample. This detects 95% of alleles that carry any of the common mutations.

Treatment/Management

Treatment involves correction of the hormone imbalances and surgery to ambiguous genitalia when appropriate. Hydrocortisone is generally used as maintenance therapy in glucocorticoid replacement as it can be given in pulses to mimic normal cortisol secretion and it has a lower potential for growth suppression in children (Deaton, Glorioso, & McLean 1999). The longer acting prednisolone or dexamethasone can also be used. Complications of treatment include Cushingoid features and growth retardation when glucocorticoid replacement exceeds the levels required for adrenal suppression.

Addition of aldosterone analogues may improve adrenal suppression; androgen inhibitors may allow smaller doses of hydrocortisone to be given; and aromatase inhibitors may limit ensuing short stature by preventing conversion of androgens to oestrogen (Deaton, Glorioso, & McLean 1999).

Prevention

Pre-implantation diagnosis is available for families who have a child with CAH and wish to utilise IVF technology for subsequent pregnancies.

Prenatal therapy for CAH can reduce the incidence and severity of virilisation in female fetuses. In families at risk (ie where one or both parents have some form of CAH or a child with CAH) treatment with dexamethasone should be commenced as soon as the pregnancy is confirmed. Chorionic villus sampling or amniocentesis can be used to diagnose whether the fetus is affected and the sex of the child. In boys, the treatment can be discontinued until after the infant is born. In female fetuses identified with CAH, treatment is required throughout the pregnancy to avoid virilisation.

How might screening reduce the burden of suffering?

Some authors have compared CAH with PKU – they have similar prevalence and onset in neonatal period. However there is one particularly relevant difference – PKU is rarely identified without screening before the onset of irreversible disability, whereas CAH is often detected clinically and is rarely associated with developmental disability. The value of screening for CAH is to prevent deaths from salt wasting prior to diagnosis, prevent sex misassignment of female infants with virilised external genitalia and prevent precocious puberty and premature epiphyseal closure in children with milder forms of CAH.

Even in areas with adequate neonatal and endocrinological services, clinical diagnosis of CAH may be delayed or misinterpreted and salt wasting crises could result in death. CAH screening could be an effective tool for identifying male infants without a family history of CAH and preventing salt loss.

6.3.2.2 EVIDENCE

Tests

Clinical examination

Examination of the genitalia is part of the routine examination of infants at birth and at the 6 week post-natal check. While the diagnosis of CAH cannot be made on physical examination alone, in the absence of serological screening tests, detection of ambiguous or abnormal genitalia may be the earliest indicator of the disease. In situations where no biochemical screening tests are used, clinical examination is the only potential indicator of CAH in the well infant. Even in those sites where biochemical screening is undertaken, there will still be a delay until results are available, and the possibility of false negative reports, so clinical vigilance is still important.

Immunoassay for 17 α -hydroxyprogesterone

In the most common form of CAH, 21 OHD, there is a marked increase in the serum concentration of 17 α -hydroxyprogesterone (17 OHP). This can be measured by enzyme immunoassay or radioimmunoassay (RIA) and can be applied as a screening tool in infancy using the routine neonatal heel prick sample. More recently, a fluoroimmunoassay (FIA) has been developed specifically for direct measurement from blood spotted on filter paper.

Genetic tests

Genotyping for mutations of the CYP 21 gene can be carried out on the neonatal heel prick sample. Experimental studies assessing the value of testing for CYP 21 mutations found that genotyping was in agreement with the confirmed diagnosis. There were no CYP 21 mutations found in the false positive group, and false negatives on biochemical testing were associated with genotypes of milder forms of CAH. An association was found between very high levels of 17 OHP and genotypes related to severe disease (Fitness et al. 1999; Nordenstrom et al. 1999).

Programs

Several different protocols have been established in screening programs around the world. Some have involved single tests using either RIA or FIA, others have added follow up with either a second biochemical test or genetic testing. Different cut off values that vary with gestational age and birth weight have been used.

In a recent study of national data from newborn screening programs throughout the USA in 1993-4, the total number of infants screened, positive test results and confirmed disease cases were reviewed (Kwon & Farrell PM 2000). The 1993 and 1994 reports were chosen because of completeness of the data and uniformity of presentation. However, for analysis of CAH results, data from 2 states were not included in the calculations due to inconsistency or incompleteness. The findings were compared with those of the previous 2 years and no statistical variance was found. From these data, sensitivity was calculated to be 100%, specificity 99% and positive predictive value 0.54%.

For PPVs approximating 0.5% there will be about 200 false positives for every true positive case detected. The financial burden incurred and the personal cost to families due to the magnitude of the false positive rate are significant adverse outcomes of screening for CAH.

No mention is made in the Kwon and Farrell study above of different protocols used across the country and the data from the various states is used collectively. Many other studies, with a variety of protocols and cut off points, are reported in the literature. These offer Level III-IV evidence and are summarised in Table 1 below. Where sensitivity and specificity are reported, they are slightly less than those from the national US data. An interesting finding is that 3 of these studies report very low false positive rates of 0.05 – 0.65% in contrast to the PPV above (Brosnan et al. 1999; Therrell, Jr., Berenbaum, Manter-Kapanke, Simmank, Korman, Prentice, Gonzalez, & Gunn 1998; Thilen et al. 1998). There is not enough detail in the reports of these studies or the US national study to comment on the features of a protocol that may result in lower false positive rates.

There are many issues relating to biochemical or genetic screening programs for CAH that need to be resolved. These include:

Determination of the best screening regime – 17 OHP single test; 17 OHP 2-test screen; 17 OHP test followed by genotyping; or single CYP 21 mutation analysis

If 17 OHP screen is to be used –

which test should be applied – RIA or FIA

what is the appropriate cutoff for positive results in term babies and in preterm or low birth weight infants.

How to interpret an abnormal 17 OHP when the child has no other signs or symptoms and what action to take

How to interpret a neonatal 17 OHP screening value in terms of disease severity

How to manage infants confirmed to have CAH, but without symptoms, to avoid over-treatment of mildly affected cases.

Hormonal screening programs specifically test for classical 21 OHD, children with non-classical, milder forms are not covered.

The possibility of false negative cases suggests that clinical observation should never be abandoned or vigilance relaxed due to the presence of a screening program

To achieve maximal benefit from screening, prompt notification of positive results and rapid follow up procedures are required so that treatment can be commenced to avoid neonatal salt wasting crises. Diseases included in the current neonatal screening programs in Australia do not require the same urgency of reporting, hence new systems and protocols will need to be instigated and adhered to.

The advantages of neonatal screening are avoidance of salt wasting crises, earlier correct gender assignment and prevention of precocious puberty and premature epiphysial closure in children with milder forms of CAH. However, while earlier diagnosis allows for treatment to be commenced prior to irreversible complication, in some infants with mild forms of CAH, over treatment has been recognised as an adverse outcome of screening (Brosnan et al. 1999;Nordenstrom et al. 1999).

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✗ *
Repeatable	?
Sensitive	✓
Specific	✓

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓ @
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓ #
Suitable test or examination	?
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?

* PPV 0.5%

@ Early treatment for mild cases not well established, over-treatment reported as possible adverse effect

Very small window prior to symptomatic stage for SW

Cost effectiveness

Three of the studies in Table 1 below offer cost effectiveness information.

- Brosnan et al found that the incremental cost for the screening detection of 8 newborns with classic CAH who were not recognised clinically was \$US 147,093 per case.
- Thilen et al assessed the cost of their program as \$US 2.70 per infant screened.
- Cutfield and Webster estimated the cost of each case of classic CAH detected to be \$8527, and \$12,181 for each case identified by screening alone.

Table 1 Neonatal screening programs for CAH

EVALUATION OF SCREENING PROGRAMS								
Study	Test	Screening protocol	Cut off for recall/retest (nmol/L)	Age at diagnosis	Additional outcomes	Effectiveness		Level of Evidence
Balsalmo et al Period A (1980-83) Prospective cohort	RIA	Day 5 (median) Range: 2-9 term, 4-30 prem/LBW	27 (all infants)	20 days (median) Range: 12-150	3 out of 5 cases detected were unsuspected	Sens 83% Spec 99.8% NPV 99.9%		CEBM 4 NHMRC III-3
Period B (1983-91) Case survey	Clinical diagnosis			SW 13 days (median) Range: 4-19 days SV 30 months (median) Range: 7-44 months	2 cases were diagnosed after salt wasting episode			
Period C (1991-95) Prospective cohort	FIA	Day 4 (median) Range: 2-9 term, 4-45 prem/LBW	Routine: 30-60 term, 36-75 prem/LBW Urgent: >60 term, >75 prem/LBW	11 days (median) Range: 5-35	5 out of 8 cases detected were unsuspected	Sens 90% Spec 92% NPV 99.9%		
Cutfield and Webster (1984-1993) Prospective cohort	RIA	Day 3-5 4.3 +/- 1(SD)	23 nmol/L > 1500 gms 32 nmol/L < 1500 gms	Initial identification Screened: 10 days (\pm 8) Clinical eval: 2 days (\pm 2)	6 of 20 cases were detected clinically prior to screening results	Sens 100% Spec 99.78% PPV 1.9% 80% of false positives in infants <2700 gm		
Brosnan et al (1989-94) Retrospective cohort: comparison between screened and unscreened populations	RIA	1. First days of life 2. Repeated at 1-2 weeks of age	Unspecified	Significant difference for male infants only (p=0.01) Median days (range) Screened: 12 (1-40) Unscreened: 26 (0-50)	100% SW were detected clinically (57%) or on first screen (43%). 52% SV and 83% NC were initially detected on second screen	False positive rate: 0.65% on first screen 0.40% for 2 screens		CEBM 4 NHMRC III-3
Thilen et al (1989-94) Prospective cohort compared with retrospective study	RIA 89-91 FIA 91-94	Day 3-5	75 nmol/L: >37 wks 400 nmol/L: 33-36 wks 600 nmol/L: <33 wks	Boys Screened: 9 days Unscreened: 21 days	35 (53%) were suspected or diagnosed before recall 31 (47%) diagnosed by screening	False pos <0.05% (60% of these were preterm) False negative 10%		CEBM 4 NHMRC III-3
Therrell et al (1989-95) Case series Period 1	RIA	1. First days of life 2. Repeated at 1-2 weeks of age	3500 ng/dL normal BW 6000 ng/dL low BW	SW 11 days (median) Range: 0-40 SV 34 days (median) Range: 0-111 NC 60 days (median) Range: 22-144	100% SW were identified on first screen. 61% SV and 87% NC were initially detected on second screen	False positive rate on first or second test = 0.4%	Recall rate Period 1: 1%	CEBM 4 NHMRC IV
Period 2	RIA	1. First days of life 2. Repeated at 1-2 weeks of age	Routine: 4000-10,000 ng/dL, 6500-20,000 low BW Urgent: > 10,000 ng/dL, > 20,000 low BW				Recall rate Period 2: 0.5%	
Al Saedi et al (Part 1) Retrospective cohort	FIA	Day 3	Retesting: 40-70 Urgent endocrinological consultation: >70			Specificity by BW >2500gm: 99.9% <2500gm: 77.2% <1500gm: 36.8%	Recall by BW >2500gm: 0.1% <2500gm: 23% <1500gm: 63%	CEBM 4 NHMRC III-3

SW – Salt wasting, SV – Simple virilising, NC – Non classic, PCA – Post conceptual age

EXPERIMENTAL / RESEARCH STUDIES				
Study	Test	Research protocol	Cut off for positive test (nmol/L)	Findings relevant to screening programs
Al Saedi et al (Part 2) Prospective cohort of preterm infants <31 weeks post conceptual age (PCA)	RIA and FIA on all samples	Weekly samples taken from day 5 until either 37 weeks PCA or discharge	N/A	FIA: 64% of first samples >40nmol/L, 44% >70nmol/L % of high levels decreased with time, no sample taken after the 5 th week was >70nmol/L No infant had evidence of adrenal insufficiency FIA is consistently higher than RIA, particularly at lower PCA Mean of all samples by FIA 38.96 +/- 37.3, by RIA 11.4 +/- 11.1 (p<0.0001)
Nordenstrom et al (1986-97) Prospective cohort – positive findings on 17OHP screening undergo genotyping	RIA 1986-90 FIA 1991-97 Positive samples tested for CYP21 mutations	Day 3-5 RIA/FIA Timing of genotyping unspecified	1986-88 - 200 (RIA) 1988-91 - 150 (RIA) 1991-97 - 75 term (FIA) - 200 prem (FIA)	10% of false positives had inconclusive 17OHP values on re-testing No CYP21 mutations were found in the false positive group False negatives had genotypes associated with the milder forms of CAH There was an association between very high levels of 17 OHP (>500 nmol/L) and genotypes associated with severe disease, but no association at lower 17 OHP levels
Fitness et al Genotyping of neonatal screening samples	Genotyping	Genotyping of random sample of neonatal screening samples – blinded to 17 OHP results	N/A	Genotyping was in agreement with ultimate clinical diagnosis DNA analysis identified males with unusually low initial 17 OHP levels

Generalisability of evidence

There is no reason to suspect that Australia will be different to the other populations already utilising a screening program.

6.3.2.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening

Comments:

There are many unanswered questions about the tests and protocols used, and there is wide variation in results from various programs.

Screening has the potential to detect life-threatening conditions prior to the onset of symptoms. However, death from undiagnosed congenital adrenal hyperplasia appears to be exceedingly rare in developed countries. Many children are detected clinically at birth because of virilisation, and many others are detected within weeks of birth at the onset of symptoms and treated appropriately.

Neonatal screening also has the potential to detect milder forms of disease that do not present until late childhood. Over-treatment is a potential adverse outcome for those with mild congenital adrenal hyperplasia detected by screening.

Some of the morbidity suffered prior to diagnosis through clinical vigilance can be avoided through screening. However, this has not been quantified or weighed against potential harms and costs. Therefore, it is impossible to make recommendations about screening at this time.

6.3.2.4. RECOMMENDATIONS

Implementation of a universal screening program for congenital adrenal hyperplasia is not recommended at this time.

6.3.2.5. FURTHER RESEARCH

- Optimal screening regime
- Interpretation and cut off points for abnormal results
- Overall population benefits of screening could be assessed by multi-centre randomised controlled trial, following the Wisconsin model for cystic fibrosis.

6.3.2.6. SEARCH TERMS

MeSH headings: Genitalia; Urogenital abnormalities; Adrenal hyperplasia, congenital

Keywords: Ambiguous genitalia; Congenital adrenal hyperplasia; Genitalia; Urogenital abnormalities

6.3.2.7 CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment

Reference Group: Caroline Briggs, Neil Wigg

Clinical experts: Fergus Cameron, Dept of Endocrinology and Diabetes, Royal Children's Hospital, Melbourne
Steve Kahler, Clinical Director, Murdoch Children's Research Centre, Melbourne
Wayne Cutfield, Director of Endocrinology, Starship Children's Hospital

6.3.2.8 REFERENCES

- Al Saedi, S., Dean, H., Dent, W., Stockl, E., & Cronin, C. 1996, "Screening for congenital adrenal hyperplasia: the delfia screening test overestimates serum 17 hydroxyprogesterone in preterm infants", *Pediatrics*, vol. 97,no. 1, pp. 100-102.
- Balsamo, A., Cacciari, E., Piazzzi, S., Cassio, A., Pirazzoli, P., & Zappulla, F. 1996, "Congenital adrenal hyperplasia: neonatal mass screening compared with clinical diagnosis only in the Emilia-Romagna region of Italy, 1980-1995", *Pediatrics*, vol. 98,no. 3 (Part 1), pp. 362-367.
- Brosnan, P. G., Brosnan, C. A., Kemp, S. F., Domek, D. B., Jelley, D. H., Blackett, P. R., & Riley, W. J. 1999, "Effect of newborn screening for congenital adrenal hyperplasia", *Archives of Pediatrics & Adolescent Medicine*, vol. 153, pp. 1272-1278.
- Cutfield, W. S. & Webster, D. 1995, "Newborn screening for congenital adrenal hyperplasia in New Zealand", *Journal of Pediatrics*, vol. 126,no. 1, pp. 118-121.
- Deaton, M.A., Glorioso, J.E., & McLean, D.B. 1999, "Congenital Adrenal Hyperplasia: not really a zebra", *American Family Physician*, vol. 59,no. 5, pp. 1190-1196.
- Fitness, J., Dixit, N., Webster, D., Torresani, T., Pergolizzi, R., Speiser, P. W., & Day, D. J. 1999, "Genotyping of CYP21, linked chromosome 6p markers, and a sex-specific gene in neonatal screening for congenital adrenal hyperplasia", *Journal of Clinical Endocrinology & Metabolism*, vol. 84,no. 3, pp. 960-966.
- Kwon, C. & Farrell, P.M. 2000, "The Magnitude and challenge of False-Positive Newborn Screening Test Results", *Archives of Pediatrics & Adolescent Medicine*, vol. 154, pp. 714-718.
- Nordenstrom, A., Thilen, A., Hagenfeldt, L., Larsson, A., & Weddell, A. 1999, "Genotyping is a valuable diagnostic complement to neonatal screening for congenital adrenal hyperplasia due to steroid 21-hydroxylase", *Journal of Clinical Endocrinology & Metabolism*, vol. 84,no. 5, pp. 1505-1509.
- Therrell, B. L., Jr., Berenbaum, S. A., Manter-Kapanke, V., Simmank, J., Korman, K., Prentice, L., Gonzalez, J., & Gunn, S. 1998, "Results of screening 1.9 million Texas newborns for 21-hydroxylase-deficient congenital adrenal hyperplasia", *Pediatrics*, vol. 101,no. 4 Pt 1, pp. 583-590.
- Thilen, A., Nordenstrom, A., Hagenfeldt, L., van Dobein, U., Guthenberg, C., & Larsson, A. 1998, "Benefits of neonatal screening for congenital adrenal hyperplasia (21-hydroxylase deficiency) in Sweden", *Pediatrics*, vol. 101, pp. 1-5.

6.3.3. CONGENITAL HEART DISEASE

6.3.3.1 BACKGROUND

Clinical picture

Congenital Heart Disease (CHD) is a collective term for the various defects in the heart and major vessels that occur as a result of abnormal fetal development. The most common conditions are ventricular septal defect (VSD), pulmonary valve stenosis (PS), coarctation of the aorta (COA), tetralogy of fallot (TOF), transposition of the great arteries (TGA), patent ductus arteriosus (PDA), pulmonary atresia (PA), atrial septal defect (ASD), aortic valve stenosis (AS), hypoplastic left heart (HLH), total anomalous pulmonary venous connection (TAPVC), common arterial trunk (CAT), and interruption of aortic arch (IAA). These may be classified according to whether there is a shunt between the arterial and venous circulations or by the presence or absence of cyanosis.

The aetiology of these conditions is generally unknown. Genetic defects and environmental and teratogenic influences have been implicated. Teratogens and adverse conditions during pregnancy that increase the risk of CHD include maternal factors such as diabetes, PKU, systemic lupus erythematosus, rubella and certain drugs (Lithium, Ethanol, Warfarin, Thalidomide, Antimetabolites and Anticonvulsants).

Natural history

Formation of the heart is completed by the end of the first trimester and defects in fetal anatomy are frequently compensated for by the parallel nature of the circulation in utero. The haemodynamic changes that occur after birth, closure of the ductus arteriosus and foramen ovale, remove that compensation and the abnormalities become apparent. An exception to this is tricuspid valve regurgitation, which results in heart failure in utero and often hydrops fetalis.

Following the significant circulatory changes that occur at birth, the majority of cardiac anomalies are evident clinically, usually by the presence of cyanosis, auscultation of a murmur or abnormal pulses. However the circulation continues to develop throughout the neonatal period and later haemodynamic effects may make underlying abnormalities more apparent. For example, as pulmonary vascular resistance falls over the first few weeks of life, left to right shunting increases and a VSD may become symptomatic.

More significant lesions will be evident at birth and require immediate, often life-saving, treatment. However, the absence of abnormal clinical findings, either at birth or in the first few months, do not exclude underlying disease. It is not unusual for CHD to present with signs or symptoms later in infancy (Wren, Richmond, & Donaldson 1999) (Abu-Harb, Hey, & Wren 1994). There is wide variation in severity of congenital heart defects and only 25-30% of the children with CHD will be symptomatic in the first year of life (Bernstein 2000). The severity of some lesions can change with growth. VSDs may become smaller and even close, but Aortic Stenosis can become more significant as the child grows.

If left untreated, many congenital heart lesions, including some that are asymptomatic during childhood, will lead to complications in later childhood or adult life. These vary with the underlying pathology but can include: heart failure, pulmonary hypertension, failure to thrive, cardiac hypertrophy or dilatation, atrial dysrhythmias, heart block, valvular insufficiency or stenosis, infective endocarditis, systemic embolisation, hypertension and cerebrovascular disease. Reduced life span or sudden death are also possible sequelae.

Prevalence

The rate of congenital malformations of the heart and circulatory system in Australia is 44.4 per 10,000 live births (AIHW 1998). Other studies report higher rates, between 50 and 90 children per 10,000 live births in the general population (Bernstein 2000) (Ainsworth, Wyllie, & Wren 1999) (Wren, Richmond, & Donaldson 1999) (Gregory, Emslie, Wyllie, & Wren 1999). When one first degree relative has CHD the risk is increased, the degree of risk depending on the particular lesion. When two first degree relatives are affected the risk may be as high as 20-30%. There are gender differences in the occurrence of specific cardiac lesions - TGA and left sided obstructive lesions are slightly more common in boys and ASD, VSD, PDA and PS are more common in girls (Bernstein 2000). As noted above, children with particular chromosomal abnormalities or specific syndromes have a much higher incidence of cardiac anomalies.

Genetics relevant to screening issues

Familial CHD has been recognised, and although no gene defect has been found, inheritance is thought to be autosomal dominant in some families (Johnson, Payne, Grant, & Strauss 1995). This group includes TAPVC,

abnormalities of the tricuspid valve and A-V canal. Single gene defects have been found for cardiac anomalies associated with Marfan's and Noonan's syndromes. CHD is also associated with chromosomal abnormalities - 90% of patients with Trisomy 18, 50% with Trisomy 21 and 40% of those with Turner's syndrome have a cardiac defect (Bernstein 2000). The conotruncal lesions (TOF, PA, IAA, Double outlet right ventricle and Truncus Arteriosus) are also linked to specific chromosomal abnormalities and are grouped together under the mnemonic CATCH 22 – Cardiac, Abnormal facies, Thymic hypoplasia, Cleft palate, Hypoglycaemia and deletions on chromosome 22q11 (Johnson et al. 1995). These abnormalities are found in the DiGeorge and Velo-Cardio-Facial Syndromes. Another microdeletion on chromosome 7 is responsible for William's syndrome, while a single gene defect (affecting the Elastin gene) in the same area on chromosome 7 produces a familial syndrome of supra-aortic stenosis and peripheral pulmonary stenoses without the other features of William's syndrome.

Diagnosis

Clinical examination

Clinical examination of the cardiovascular system includes inspection for cyanosis, examination of the praecordium, auscultation of the heart, palpation of pulses and measurement of blood pressure. If abnormalities are detected, a diagnosis of the underlying condition is sometimes made clinically at this point, however more usual practice is referral for further investigation, in particular echocardiography.

Chest X-ray and Electrocardiogram

Chest radiography and the performance of an ECG are simple and readily available investigations which often provide useful information about the severity of cardiac abnormalities.

Echocardiography

Echocardiography is used to determine cardiac structure, measure pressures, gradients and directions of flow and contractile functioning. Two-dimensional echocardiography provides real time images of cardiac structures and has largely replaced invasive diagnostic procedures in congenital heart disease.

Cardiac catheterisation

Cardiac catheterisation can play a significant role in the diagnosis of congenital heart lesions when echocardiography is incomplete or if there is a difference between clinical and echo findings. This invasive procedure carries potential complications and thus should be reserved for cases where non-invasive diagnostic tools are unsuitable or inadequate. Its primary function is for pre or post-surgical interventions, biopsy, electrophysiological studies or transcatheter procedures.

Magnetic Resonance Imaging

MRI is useful in diagnosis of particular lesions in areas less well visualised by echocardiography eg pulmonary vasculature.

Treatment/Management

Medical management may be required initially in symptomatic patients to control heart failure or prevent closure of the ductus arteriosus. Subsequent action may be definitive surgical procedures or palliative surgery (eg a shunt operation) to buy time for a more favourable outcome.

The treatment of choice, the timing of the procedure and the associated risks and complications are obviously specific to the abnormality. Surgery or catheter directed procedures are performed routinely in paediatric tertiary centres, the success rates also dependent on the underlying pathology and the expertise of the surgical team. Some procedures carry little risk and a 100% recovery can be expected (eg VSD); however when the cardiac lesion is extensive the results are less successful (eg Hypoplastic Left Heart).

Some conditions may not require active treatment and can be observed for spontaneous resolution eg small VSD, PDA in premature infant. The use of antibiotic prophylaxis for procedures such as dental extractions, in which a bacteraemia is likely to be caused by the surgical manipulation, is required in those conditions susceptible to infective endocarditis.

Prevention

Few congenital heart defects can be prevented, however avoidance of known teratogens should be considered where appropriate, and adequate immunity to Rubella ensured.

How might screening reduce the burden of suffering?

It is important to detect CHD early as development of symptoms and deterioration may be sudden and some treatable defects may cause death prior to diagnosis. Early recognition and treatment can avoid irreversible complications and disability.

3.3.3.2 EVIDENCE

Tests

Clinical examination is the only screening test used for CHD.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✗
Acceptable to the public	✓
Accurate	✗
Repeatable	✗
Sensitive	✗
Specific	✗

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	?
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

Assessment of the cardiovascular system is part of the routine neonatal medical examination, at the 6 week post-natal check and, in some programs, at pre-school or school entry.

A prospective cohort study evaluating the neonatal screening program of a large UK hospital considered all births over a 2 year period (Ainsworth, Wyllie, & Wren 1999). All babies (n=7204) were screened by clinical examination within 48 hours of delivery, echocardiography was performed on all those found to have murmurs, and the regional paediatric cardiology database was used to determine which children in the cohort were diagnosed with CHD in the first year of life. The whole birth cohort was screened and followed up, however only those with murmurs were examined with echocardiography. The authors report a sensitivity of 44% (95% CI 31-58) (25 of the 57 babies with CHD were detected by neonatal examination) and a positive predictive value of 54% (95% CI 39-69%) (25 of the 46 babies with murmurs were found to have CHD). Specificity is reported as 99.7%, but this is based on an audit of the paediatric cardiology register and not on evaluation of all the infants in the cohort with echocardiography. No comment is made about the completeness of this register, or local population mobility, however it is unlikely that a large number of children with CHD were missed by this method. In this study, 15% of infants with CHD presented before the routine neonatal check, 37% were diagnosed by screening and a further 48% were diagnosed after the neonatal examination.

In their retrospective review of screening performed at the 6-8 week postnatal check, Gregory et al set out to measure the efficacy of this program in the UK (Gregory et al. 1999). The standard practice was for a GP or community paediatrician to examine all infants at 6-8 weeks, and all practitioners in the study area were advised to refer every baby found to have a murmur. The authors have acknowledged the significant limitations of the study, but several points can still be made. Only 83% of eligible babies (n=7132) received the routine examination. Of the 47 infants found to have a murmur at this time, only 25 were referred for follow up, suggesting a problem with the screening protocol. However, of the 25 referred, 11 were diagnosed with CHD, but of the 22 not referred, none had a diagnosis of CHD and remained well at the 12 month review time. The reasons why those who were subsequently found to have pathology were in the group that was referred, and those with

benign murmurs were not, are not known, but would be relevant in further evaluation of the screening process. In this study, 51% of infants with CHD were diagnosed before 6 weeks, 31% at the 6-8 week examination and 17% between 6 weeks and 12 months.

Wren and his colleagues conducted a retrospective review of the screening examinations of a cohort of infants (n=1590) with CHD in a large regional area of the UK over a period of 8 years. With the exception of some difficulty obtaining medical records of the 6 week examination, the methodology appears sound. They had similar findings to the two previous papers. Of those undiagnosed at the time of the routine neonatal examination, 45% were detected, but only 16% had a definitive diagnosis made before discharge. The authors note that of the babies discharged home undiagnosed, 35% were diagnosed by 6 weeks and 57% by 3 months. If every baby who was found to have signs or symptoms of CHD had been examined by echocardiograph within 4 weeks of the abnormality being noted, 58% could have been diagnosed by 6 weeks and 76% by 3 months.

These three papers suggest that both the neonatal and postnatal examinations are useful in detecting CHD. It is not clear whether the low sensitivity of the clinical examination is due to the fact that murmurs are not present at the time of examination or whether they were missed in the screening examination. Poor rates of referrals for specialist opinion are also an issue for a screening program. Further investigation of the examination and follow up procedures could lead to improved screening results.

However, only half of the infants found to have murmurs actually had underlying heart disease. This means that many babies undergo further investigation and are given the label of 'heart condition' unnecessarily. The Ainsworth study attempted to minimise this stigma by ensuring that the echo was performed within 24 hours of the screening examination to reassure parents quickly. It has been demonstrated that some parents whose children have received a false positive diagnosis in various screening programs continue to treat their child differently and remain cautious about normal activities even after being reassured that there is no underlying problem (Tluczek, Mischler, Farrell, Fost, Peterson, Carey, Bruns & McCarthy 1992).

Cost effectiveness

No information regarding the cost-effectiveness of screening for CHD was found in the course of this review.

Quality of evidence

S. B. Ainsworth, J. P. Wyllie, and C. Wren. Prevalence and clinical significance of cardiac murmurs in neonates. *Archives of Diseases in Childhood Fetal Neonatal Edition* 80 (1): F43-F45, 1999.

Level: CEBM 4, NHMRC IV

Quality: This was a prospective cohort study where the whole birth cohort was screened. There are 2 major limitations:

only those with positive findings on clinical examination, rather than the whole cohort, went on to have echocardiography.

the Regional Paediatric Cardiology Register was used to determine which children in the cohort were diagnosed with CHD in the first year of life. No comment is made about the completeness or accuracy of this database or local population mobility.

Relevance of evidence: This information is directly relevant to the evaluation of screening for CHD.

J. Gregory, A. Emslie, J. P. Wyllie, and C. Wren. Examination for cardiac malformations at six weeks of age. *Archives of Diseases in Childhood Fetal Neonatal Edition* 80:F46-F48, 1999.

Level: CEBM 4, NHMRC IV

Quality: The limitations of this study are:

Results of the clinical examination were identified by retrospective examination of medical records

The Regional Paediatric Cardiology Register was used to determine which children in the cohort were diagnosed with CHD in the first year of life. No comment is made about the completeness or accuracy of this database or local population mobility.

Children were defined as not having CHD if they were not on the register

Relevance of evidence: This information is directly relevant to the evaluation of screening for CHD.

C. Wren, S. Richmond, and L. Donaldson. Presentation of congenital heart disease in infancy: implications for routine examination. *Archives of Diseases in Childhood Fetal Neonatal Edition* 80:F49-F53, 1999.

Level: CEBM 4, NHMRC IV

Quality: The limitations of this study are:

Results of the clinical examination were identified by retrospective examination of medical records, which, in many cases, could not be traced.

Relevance of evidence: This information is directly relevant to the evaluation of screening for CHD.

Generalisability of evidence

These studies are from the UK where clinical practice is very similar to Australia. While we have no information about the rates of detection or referral patterns in Australia, there are no obvious reasons to expect that they are different to those in the UK. However, GPs in the UK are assigned a cohort of patients that they have the opportunity to follow over time. In Australia this is not the case, with families attending many different GPs, or only attending emergency departments for acute care. A greater number of abnormalities, particularly those that become clinically evident later, could potentially be missed in this situation.

6.3.3.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening

Comments:

The neonatal and 6 week postnatal examinations offer routine access to newborns to detect congenital abnormalities. Although there are limitations to clinical examination as a screening tool in this setting, it may be the only opportunity to detect CHD in the pre-symptomatic phase.

However, it appears that cardiac examination frequently falls short of program requirements for an effective detection system.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

6.3.3.4. RECOMMENDATIONS

Although there is little firm evidence to support the value of screening, we recommend continuation of newborn and 6 week cardiac examination, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical and nursing staff
- clear pathways and/or guidelines for referral
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records
- strategies to improve attendance at the 6 week postnatal check (particularly for disadvantaged groups)
- opportunistic examination of children not examined in the newborn period

6.3.3.5. FURTHER RESEARCH

Evaluation and continuing quality improvement of programs

Methods to improve the sensitivity and specificity of the clinical examination in detecting important cardiac disease

6.3.3.6. SEARCH TERMS

MeSH headings: Heart defects, congenital; Down syndrome

Keywords: Congenital heart defect; Hypertrophic obstructive cardiomyopathy; Down syndrome

6.3.3.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Reference group – Garth Alperstein, David Hall

Clinical Experts – Dr Jim Wilkinson, Director, Dept of Cardiology, Royal Children’s Hospital, Melbourne

6.3.3.8. REFERENCES

Abu-Harb, M., Hey, E., & Wren, C. 1994, "Death in infancy from unrecognised congenital heart disease.", *Archives of Disease in Childhood*, vol. 71, p. F3-F7.

AIHW 1998, *Australia's health 1998: the sixth biennial health report of the Australian Institute of Health and Welfare* Canberra.

Ainsworth, S. B., Wyllie, J. P., & Wren, C. 1999, "Prevalence and clinical significance of cardiac murmurs in neonates", *Archives of Diseases in Childhood Fetal Neonatal Edition*, vol. 80, no. 1, p. F43-F45.

Bernstein, D. 2000, "Part XIX, The Cardiovascular System," in *Nelson Textbook of Pediatrics*, 16th edn, Behrman RE, Kliegman RM, & Jenson HB, eds., WB Saunders Company, Philadelphia, pp. 1362-1363.

Gregory, J., Emslie, A., Wyllie, J. P., & Wren, C. 1999, "Examination for cardiac malformations at six weeks of age", *Archives of Diseases in Childhood Fetal Neonatal Edition*, vol. 80, p. F46-F48.

Johnson, M. C., Payne, R. M., Grant, J. W., & Strauss, A. W. 1995, "The genetic basis of paediatric heart disease.", *Annals of Medicine*, vol. 27, no. 3, pp. 289-300.

Gluczek, A., Mischler, E.H., Farrell, P.M., Fost, N., Peterson, N.M., Carey, P., Bruns, W.T., & McCarthy, C. 1992, "Parents' knowledge of neonatal screening and response to false positive cystic fibrosis testing", *Journal of Developmental & Behavioral Pediatrics*, vol. 13, pp. 181-186.

Wren, C., Richmond, S., & Donaldson, L. 1999, "Presentation of congenital heart disease in infancy: implications for routine examination", *Archives of Diseases in Childhood Fetal Neonatal Edition*, vol. 80, p. F49-F53.

6.3.4. CONGENITAL HYPOTHYROIDISM

6.3.4.1 BACKGROUND

Clinical picture

Congenital hypothyroidism is defined as defective function of the thyroid gland from birth (Pollitt, Green, McCabe, Booth, Cooper, Leonard, Nicholl, Nicholson, Tunaley, & Viridi 1997). Its effects are due to lack of normally-functioning thyroid hormone, either through insufficient production of normal thyroid hormone or through production of abnormal thyroid hormone.

Congenital hypothyroidism may be primary or secondary. *Primary hypothyroidism* is usually due to the thyroid gland being absent (thyroid agenesis, athyreosis) or failing to migrate downwards to the normal position during embryogenesis, resulting in a rudimentary gland unable to produce sufficient thyroid hormone (ectopic thyroid). Rarely, primary hypothyroidism is due to dysmorphogenesis or to suppression of the fetal thyroid by maternal medication. *Secondary hypothyroidism* results from abnormally low production of thyroid stimulating hormone (TSH, or thyrotropin) by the pituitary gland, without which the (normal) thyroid gland fails to produce thyroid hormone. In practice, screening for congenital hypothyroidism is directed towards detection of thyroid agenesis and ectopia. Only these conditions will be discussed further.

Hypothyroidism at any age leads to a clinical syndrome of “coarse” facies, lethargy and “sluggish” behaviour, cool skin, constipation, and other features. If it occurs prenatally or in early childhood, it also leads to mild-moderate intellectual disability, growth and bone age retardation, enlarged protruding tongue, micrognathia, immature facies, umbilical hernia, persistent jaundice, hoarse cry and feeding problems. Neurological abnormalities such as spasticity and poor coordination can occur. The Kocher-Debré-Semelaigne syndrome is hypertrophy of muscles with slowness of contraction and movement. Difficulties with mathematics over and above that expected for IQ is not uncommon (Brook 1995).

Natural History

Signs and symptoms are related to severity and duration of lack of thyroid hormone.

Babies born with thyroid agenesis typically have some signs and symptoms present at birth, since although some thyroid hormone is supplied transplacentally by the mother prenatally, the fetus must supplement this in the latter half of pregnancy to meet its needs. Babies with functioning thyroid remnants produce some thyroid hormone and serum levels may even be in the normal range at birth. Thus they may display few or no symptoms in the neonatal period, but develop worsening signs as growth outstrips supply, the ectopic remnant fails, and the duration of hormone insufficiency lengthens. Clinical diagnosis may be delayed weeks, months or years.

Without replacement thyroid hormone therapy, full-blown clinical hypothyroidism (“cretinism”) usually develops and is lifelong. However, replacement therapy has been available for many years, so that even when intellectual disability was not prevented, many of the day to day effects of hypothyroidism (eg poor growth and constipation) were reversed once the condition was diagnosed.

Prior to introduction of screening, outcomes were variable, due to the varying severity of the conditions and also because some children were detected and treated relatively early even without screening. Some achieved IQs within the normal range (Weber, Siragusa, Rondanini, Cerai, Mora, Colombini, Medaglini, Lia, Locatelli, Comi, & Chiumello 1995), while others developed irreversible intellectual disability. Brook reports that prior to screening approximately 10% of infants were diagnosed within the first month of life, 35% within three months, 70% within the first year and 100% by 3-4 years. In a series of 651 infants with congenital hypothyroidism in the pre-screening era, the mean IQ was 76% with only a small number of children in the 110-120 range. 78% of children diagnosed before 3 months of age had IQ >85, compared to just 19% of those diagnosed aged 3-6 months and 0% diagnosed >6 months (Brook 1995).

If at any age children receive insufficient replacement or compliance by the child or parent is poor, signs and symptoms of hypothyroidism will reappear. Treatment must be lifelong. Endocrinologic opinion is that after two years of age *permanent* intellectual deficit does not appear to occur, but *concurrent* cognitive functioning falls at any age when serum levels are low and rises when replacement is optimised (New England Congenital Hypothyroidism Collaborative 1994).

Prevalence

The incidence of congenital primary hypothyroidism detected by screening is consistently approximately 1 in 3-4,000 live births across screening programs in many countries (Toublanc 1992). The incidence detected

clinically prior to commencement of screening programs was approximately 1 in 5-10,000 (Brook 1995). This apparent increase with screening is probably due to a combination of increased sensitivity (ie some cases, often the milder ones, were not previously detected early in life) and transient neonatal hypothyroidism (false positives). In a Swedish series of neonatal blood spots analysed retrospectively when the children were 5 years old, 31 screened positive to high neonatal TSH of which 15 (48%) had been clinically diagnosed. Seven (23%) were biochemically hypothyroid but had not yet been diagnosed by 5 years even though cognitive deficits were present (false negatives), and 9 (29%) were euthyroid (presumed transient hypothyroidism). In some cases raised neonatal TSH has been reported to be associated with postnatal exposure to iodine in antiseptics, eg for umbilical cleansing. However, many cases of apparent transient hypothyroidism may represent true borderline hypothyroidism and thus warrant careful follow up (Daliva, Linder, DiMartino-Nardi, & Saenger 2000).

The incidence is slightly higher in Hispanic and Asian populations and lower in black populations. Females are more often affected than males, and children with birthweights <2.0 kg and >4.5 kg are at higher risk (Waller, Anderson, & Lorey 2000).

Secondary hypothyroidism is extremely rare (1 in 135,000 to 1 in 25,000 (Pollitt et al. 1997)).

Genetics Relevant to Screening

Thyroid agenesis and ectopia are considered to be sporadic conditions with no increased risk in subsequent pregnancies.

Diagnosis

A presumptive diagnosis is made with the demonstration of high serum TSH in combination with low or low-normal serum thyroxine (T4 and/or T3). The infant is then recalled for confirmatory testing of serum levels, compared to normal concentration for age. At this time other tests may also occur if causes other than athyreosis or ectopia are suspected. Radionuclide scans are usual to demonstrate whether there is any functioning thyroid tissue, and an ultrasound scan will demonstrate presence or absence of the gland. This should not however delay commencement of replacement therapy (American Academy of Pediatrics 1993).

In the past diagnosis usually occurred after obvious clinical signs developed.

Treatment/Management

Thyroxine replacement aims to (1) maintain levels in the upper end of the normal serum range (about 65-155 nmol/l, depending on the laboratory) (2) avoid symptoms of hyperthyroidism and (3) to fully suppress TSH secretion (<0.5 mU/ml). Doses are titrated against serum thyroxine and TSH levels. Some authors accept even higher serum levels of T4 in the first four months of life with the goal of normalising TSH by eight weeks of life, arguing that theoretical side effects of excessive thyroxine such as craniosynostosis and irritability are rarely if ever seen (Garcia, Calzada-Leon, Perez, Martinez, Gonzalez, Perez, de le Lus Ruiz, & Altamirano 2000).

The best evidence about treatment levels come from prospective study of children identified; no randomised controlled trials comparing correlates and outcomes of differing doses has been reported. High doses in the neonatal period have generally been reported to result in better cognitive outcomes, but may be related to more behaviour problems in the primary school years (Rovet & Ehrlich 1995). In the primary school years, serum thyroxine levels above the upper end of the normal range may also be associated with poorer attention and memory tasks (Rovet & Alvarez 1996; Rovet & Ehrlich 1995).

Recently, it has been suggested that partial replacement of a small amount of thyroxine with tri-iodothyronine may lead to better neuropsychological function in adults (Bunevicius, Kazanavicius, Zalinkevicius, & Prange Jr 1999). No studies have yet been reported in children.

Starting replacement doses of 8-10 µg/kg/day recommended in the early days of screening programs have been superseded by starting doses of 10-15 µg/kg/day. Required doses usually fall to approximately 6-7 µg/kg/day by 2 years of age. Doses are titrated against serum levels of T4 and TSH. It is recommended that levels should be monitored 1-2 monthly in the first year of life, 2-3 monthly from 1-3 years of age, then 3-12 monthly till growth is complete (American Academy of Pediatrics 1993).

How might screening reduce the burden of suffering?

Early diagnosis and thyroxine replacement prevents, either completely or largely, the devastating lifelong effects of congenital hypothyroidism. Neonatal screening programs have now been adopted throughout the developed world.

6.3.4.2. EVIDENCE

Tests

Primary T4 with backup TSH measurements

Traditionally most North American programs measured thyroxine (T4) on the initial neonatal blood spot. If lower than an arbitrary cutpoint (usually between the 10th and 20th centile for that population), then the same initial blood spot is retested for TSH. This approach has the advantage of being able to detect primary as well as secondary hypothyroidism, but this is a rare condition and other signs are often present. By definition the initial false positive rate is high (ie 10-20%) in order not to miss children with ectopic thyroids, who often have low-normal T4 but high TSH. If only infants with raised TSH are recalled, the final recall rate is usually about 0.05% (ie two infants recalled for every true positive). If children with two low T4 levels are also recalled, the final recall rate is about 0.3% (ie 12 infants recalled for every true positive) (American Academy of Pediatrics 1993).

Primary TSH measurements

Australian and New Zealand programs, like most European programs, measure TSH levels as the initial screen. Children are recalled if levels are high, the recall rate typically being about 0.05% (ie two infants recalled for every screen positive). Children with primary hypothyroidism will not be diagnosed by this screen (American Academy of Pediatrics 1993).

One study compared primary T4 with primary TSH screening. This study estimated that one case out of 93,000 infants screened would have been missed by the primary T4 but detected by the primary TSH approach, while two cases would have been missed by the primary TSH but detected by the primary T4 approach (American Academy of Pediatrics 1993).

As discussed in the recent HTA systematic review, some problems arise because hypothyroidism is not an all-or-nothing condition. Thus cutpoints must be optimised to maximise sensitivity and specificity, with some false positives and false negatives inevitable. The apparent rise in prevalence of congenital hypothyroidism following introduction of screening probably resulted both from true positives that would previously have remained undiagnosed (often children with functioning thyroid remnants) and from transient hypothyroidism that resolves spontaneously (eg exposure to iodine, maternal autoantibodies). False negatives may occur in late-developing cases with normal TSH on the initial screen. This may be more common when neonatal screens are taken very early. For example, the Northwest Regional Screening Program in the US routinely takes specimens within 48 hours of birth and at 6 weeks, and reports an additional 5% of cases diagnosed by the second screen (LaFranchi, Hanna, Krainz, Skeels, Miyahira & Sesser 1995 as cited in Pollitt et al. 1997).

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✓
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	✓

Programs

National programs for screening for congenital hypothyroidism were widely introduced in the 1970s and early 1980s. In these programs, a single sample of capillary blood is collected and stored on filter papers, usually in conjunction with the “Guthrie” card which was initiated for phenylketonuria screening. Age at which blood is collected varies; in the UK it usually occurs at 6-14 days, in the US at 3-5 days, and in Australia at 2-3 days. Two basic systems exist (see “Tests” above). In the first, blood spots are tested for thyroxine level and those in the lowest 10-20% are retested for TSH level. In the second, blood spots are tested for TSH level, and all those higher than a set cutpoint are retested and then recalled if still high for further diagnostic testing.

The performance of reported screening programs is high. Late-diagnosed hypothyroidism in infancy has fallen by >95%. The HTA review notes that in France between 1978-87, 50 cases (one in 118,000 tests) were false negatives of screening; of these 23 appeared to be true false negatives (ie initial screen confirmed normal), 15 were due to failures in the collection or transmission of specimens, and 12 were due to errors in the laboratory (Lejer 1990 as cited in Pollitt et al. 1997). Stated concerns related mainly to organisational and logistic issues to improve audit and maximise the coverage and benefit of programs. It was suggested that new additions to screening programs, for example tandem mass spectrometry for which much earlier routine blood spots (eg in the first 48 hours of life) might be recommended, required careful monitoring for their impact on sensitivity and specificity of TSH screening.

No randomised controlled trials of outcomes of screening programs for congenital hypothyroidism were identified, and it is highly unlikely that anyone would now contend that one should be undertaken. Congenital hypothyroidism is now routinely detected in the neonatal period and treatment is begun in the first days or weeks of life, and late diagnosis is now rare. Even though subtle problems are common in children identified through screening, there appears to be universal agreement that screening programs result in major benefits and are both logistically and economically viable.

In a 1996 meta-analysis of outcomes of children in the screening era, Derksen-Lubsen compared a combined total of 675 children with congenital hypothyroidism with 570 normal controls. The overall IQ deficit was 6.3 points (95%CI 4.7-7.8). The most important predictor of outcome was the severity of the condition at birth (low thyroxine level, delayed bone age) (Derksen-Lubsen & Verkerk 1996). Even with adequate replacement, babies with very low plasma thyroxine levels (<40 nmol/l) at birth average 10 IQ points lower than those born with higher thyroxine levels. This indicates that while postnatal cognitive deficits can be fully prevented, prenatal cognitive insults cannot be fully reversed even with optimal treatment from soon after birth. Deficits tend to be in the performance and motor domains, with verbal IQ tending to be maintained. Up to 20% are reported to have persistent mild hearing loss, which has been reported to occur mainly in those in whom treatment is delayed beyond 14 days. Virtually all babies achieve normal growth and final height, though an inverse relationship

between bone age at birth and height has been reported to at least nine years of age (Aronson, Ehrlich, Beliley, & Rovet 1990).

Surprisingly, while many studies have reported outcomes for children with congenital hypothyroidism detected by screening compared to normal controls, few studies of outcomes before and after introduction of screening have been reported. Weber et al compared IQ of 15 children diagnosed after screening began (mean age of starting therapy 33 days, mean age at outcome 9.9yr, SD 0.81) with IQ of 11 children diagnosed before screening commenced (mean age of starting therapy 10.1 months, mean age at outcome 14.4yr, SD 2.2) (Weber et al. 1995). Median IQ in the screened group was 110 (range 84-124) and in the unscreened group was 77 (range 69-94) (Hulse 1983) (Klein 1979).

Cost effectiveness

One systematic review of the cost, yield and outcome of neonatal screening for congenital hypothyroidism was identified (Pollitt et al. 1997). Six studies assessing the costs of implementing a congenital hypothyroidism screening program and the likely benefits that would accrue were included in the review. All but one concluded that such a screening program is beneficial. Weaknesses of the studies included failure to conduct sensitivity analyses and failure to incorporate measures of health gain. All showed screening to be cost-effective, but across a very large range. Overall, the combined studies were assigned a Grade B strength of recommendation (fair evidence to support a congenital hypothyroidism screening program). Readers are referred to the systematic review pages 119-22 for full discussion of the papers and the analysis.

Quality of evidence

Overall the combined evidence is of high quality, being derived from large scale cohorts with careful ascertainment.

Generalisability of evidence

The evidence discussed here is generalisable to most populations.

6.3.4.3. CONCLUSIONS

Good evidence to make a recommendation for screening.

Comment

Although no randomised controlled trials or comparative trials were identified, the quality of population-based cohort studies and the clear and sizeable benefits of screening approach this standard.

6.3.4.4. RECOMMENDATIONS

That screening programs continue.

That the impact of adding multiple new conditions to newborn screening programs is monitored, to ensure that the quality of existing congenital hypothyroidism programs is not threatened.

6.3.4.5. FURTHER RESEARCH

Optimum screening cutpoints to maximise sensitivity and specificity of blood spots taken in the first 48 hours of life

Randomised controlled trials using a broad range of outcome measures of:

- medium vs high initial doses of thyroxine
- use of T3 as a partial substitute for thyroxine in children

Quality of life of children and adults with congenital hypothyroidism

6.3.4.6. SEARCH TERMS

Hypothyroidism

Congenital hypothyroidism

Neonatal hypothyroidism

6.3.4.7. CIRCULATION OF DRAFT FOR COMMENT

Dr Fergus Cameron – Department of Endocrinology & Diabetes, Royal Children's Hospital, Melbourne.

6.3.4.8. REFERENCES

American Academy of Pediatrics 1993, "Newborn screening for congenital hypothyroidism: recommended guidelines", *Pediatrics*, vol. 91, pp. 1203-1209.

Aronson, R., Ehrlich, R. K., Beliley, J. D., & Rovet, J. F. 1990, "Growth in children with congenital hypothyroidism detected by neonatal screening", *Journal of Pediatrics*, vol. 116, pp. 33-37.

Brook, C. G. D. (Ed) 1995, *Clinical Paediatric Endocrinology*, 3rd edn, Blackwell Science, Oxford.

- Bunevicius, R., Kazanavicius, G., Zalinkevicius, R., & Prange Jr, A. J. 1999, "Effects of thyroxine as compared with thyroxine plus triiodothyronine in patients with hypothyroidism", *New England Journal of Medicine*, vol. 340, pp. 424-429.
- Daliva, A. L., Linder, B., DiMartino-Nardi, J., & Saenger, P. 2000, "Three-year follow-up of borderline congenital hypothyroidism", *Journal of Pediatrics*, vol. 136, pp. 53-56.
- Derksen-Lubsen, G. & Verkerk, P. H. 1996, "Neuropsychologic development in early treated congenital hypothyroidism: analysis of literature data", *Pediatric Research*, vol. 39, pp. 561-566.
- Garcia, M., Calzada-Leon, R., Perez, J., Martinez, M. P., Gonzalez, R., Perez, M., de le Lus Ruiz, M., & Altamirano, N. 2000, "Longitudinal assessment of L-T4 therapy for congenital hypothyroidism: difference between athyreosis vs ectopia and delayed vs normal bone age", *Journal of Pediatric Endocrinology & Metabolism*, vol. 13, pp. 63-69.
- Hulse, A. 1983, "Congenital hypothyroidism and neurological development", *Journal of child psychology and psychiatry*, vol. 24, pp. 629-635.
- Klein, R. Z. 1979, "Neonatal screening for hypothyroidism", *Adv Pediatrics*, vol. 26, pp. 417-440.
- New England Congenital Hypothyroidism Collaborative 1994, "Correlation of cognitive test scores and adequacy of treatment in adolescents with congenital hypothyroidism", *Journal of Pediatrics*, vol. 124, pp. 383-387.
- Pollitt, R. J., Green, A., McCabe, C. J., Booth, A., Cooper, N. J., Leonard, J. V., Nicholl, J., Nicholson, P., Tunaley, J. R., & Viridi, N. K. 1997, "Neonatal screening for inborn errors of metabolism: cost, yield and outcome", *Health Technology Assessment*, vol. 1, no. 7, pp. 1-203.
- Rovet, J. & Alvarez, M. 1996, "Thyroid hormone and attention in school-age children with congenital hypothyroidism", *Journal of child psychology and psychiatry*, vol. 37, no. 5, pp. 579-585.
- Rovet, J. F. & Ehrlich, R. M. 1995, "Long-term effects of L-thyroxine therapy for congenital hypothyroidism", *Journal of Pediatrics*, vol. 126, pp. 380-386.
- Toublanc, J. E. 1992, "Comparison of epidemiological data on congenital hypothyroidism in Europe with those of other parts in the world", *Hormone Research*, vol. 38, pp. 230-235.
- Waller, D. K., Anderson, J. L., & Lorey, F. 2000, "Risk factors for congenital hypothyroidism: an investigation of infant's birth weight, ethnicity, and gender in California, 1990-1998", *Teratology*, vol. 62, pp. 36-41.
- Weber, G., Siragusa, V., Rondanini, G. F., Cerai, L. M. P., Mora, S., Colombini, J., Medagliani, S., Lia, C., Locatelli, T., Comi, G., & Chiumello, G. 1995, "Neurophysiologic studies and cognitive function in congenital hypothyroid children", *Pediatric Research*, vol. 37, pp. 736-740.

6.3.5 CYSTIC FIBROSIS

6.3.5.1 BACKGROUND

Clinical picture

Cystic Fibrosis (CF) is an autosomal recessive disorder involving abnormal chloride and sodium transport in epithelial cells. This results in altered secretions in affected organ systems such as the airways, pancreatic exocrine ducts, biliary ducts, GIT and vas deferens. The clinical picture is characterised by suppurative lung disease, pancreatic exocrine insufficiency, biliary cirrhosis and reduced fertility.

Natural history

The earliest clinical presentation of CF is meconium ileus in the neonatal period, occurring in 10-20% of patients. Other symptoms present later in infancy or childhood. Pulmonary complications usually dominate the course of the disease. Bronchiectasis and recurrent infection are common to most patients, and pneumothorax, haemoptysis and cor pulmonale are often present in those with advanced disease. Pancreatic insufficiency results in maldigestion of fat, protein and complex carbohydrates with consequent failure to thrive, poor growth and delayed puberty. Biliary cirrhosis, male infertility and reduced female fertility are additional complications.

Prevalence

The incidence of CF in Australia is 1 in 2,500 live births, with approximately 4% of the population carrying a gene mutation. Parents who are both carriers have a 1 in 4 chance of having an affected child with each pregnancy. However, more than 80% of affected individuals are born to families with no prior family history.

Genetics relevant to screening issues

Individuals with CF have mutations in the cystic fibrosis transmembrane regulator gene (CFTR). CFTR regulates chloride transport in epithelial cells. There are over 800 known mutations in the CFTR gene, the commonest being $\Delta F508$. Some of the other mutations result in milder forms of clinical disease.

Diagnosis

The Sweat Test is considered the gold standard for diagnosis of CF. The affected CFTR does not allow sodium and chloride to be absorbed as the sweat passes along the duct. Sweat is induced by pilocarpine iontophoresis and collected for analysis. Chloride values > 60 mmol/L are diagnostic of CF, while chloride levels < 40 mmol/L are considered normal. Values between 40 and 60 mmol/L are borderline and require further investigation.

Treatment/Management

Physiotherapy, antibiotics, nutritional supplements and pancreatic enzyme replacement are the mainstay of CF management. In some centres, antibiotics are being given prophylactically to prevent pulmonary infection. In addition, newer treatments such as anti-inflammatory agents and gene therapy are being trialed in clinical settings. Lung transplantation has been used successfully in some older patients with end stage lung disease.

Prevention

Pre-conceptual or pre-natal screening for CFTR mutations is possible. Screening can be used to detect carrier status at a general population level, in a family planning setting, or as a 'cascade' method within the families of affected individuals. Couples who both carry a gene mutation can consider avoidance of pregnancy, artificial insemination by donor, pre-implantation diagnosis or pre-natal screening. Pre-natal screening of fetal DNA may be advised when couples would consider termination if a positive diagnosis was made. The Health Technology Assessment (HTA) Review of Screening for Cystic Fibrosis (Murray, Cuckle, Taylor, Littlewood, & Hewison 1999) recommends that antenatal genetic screening be offered routinely, that pre-conceptual screening be made available to couples on request and that genetic screening be available for infertile men and sperm donors.

How might screening reduce the burden of suffering?

CF can be difficult to recognise, and thus diagnosis is often delayed. At the time of diagnosis, many patients are already malnourished or have established lung disease (Farrell, Kosorok, Laxova, Shen, Kosciak, Bruns, Splaingard, & Mischler 1997). Screening to detect CF at an earlier age than by clinical manifestations may postpone the development of these complications.

Neonatal screening will also identify the parental carrier state with a possible reduction in subsequent affected pregnancies.

6.3.5.2 EVIDENCE

Tests

Sweat Test

While considered the ‘gold standard’ for making the definitive diagnosis of CF, the sweat test is not considered useful as a screening test in the neonatal period. It is often impossible to collect sufficient sweat from the newborn for accurate analysis and the test itself is time-consuming and expensive.

BM meconium Test

The level of albumin in meconium is raised in children with CF and can be measured either by qualitative dipstick test or immunoassay. This has high false-positive and false-negative rates, and has generally been replaced by immunoreactive trypsinogen testing in centres performing neonatal screening.

Immunoreactive Trypsinogen (IRT)

The serum levels of IRT are raised in CF. Measurement can be made by radio immunoassay using dried blood spots from neonatal heel prick samples (day 3-5). One advantage of this method is that it can be combined with the neonatal testing for PKU and Congenital Hypothyroidism. False positives for this test are related to ethnicity, CF carrier status and perinatal health. Single stage IRT testing has been reported to have sensitivity of 85.7% and specificity of 99.8% (Gregg, Simantel, Farrell PM, Kosciak, Kosorok MR, Laxova A, Laessig, Hoffman, Hassemer, Mischler, & Splaingard 1997).

DNA analysis

Several assays are available commercially and many are able to detect several different genetic mutations. This test is usually combined with IRT testing to increase accuracy. Two stage IRT/DNA screening is reported to have sensitivity of 95.2% and specificity of 99.9% (Gregg et al. 1997).

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✓
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

A number of protocols have been employed using various combinations of meconium albumin, IRT and DNA testing. Most neonatal screening programs currently use IRT on the heel prick blood, however it is recognised as having a low discriminatory power in this age group (Murray et al. 1999). In order to increase the positive predictive value and reduce the number of false positives, a two-stage protocol has been widely adopted. Neonates who have a positive test on the initial sample undergo a second test. In an IRT+IRT protocol a second blood sample is taken at 4 weeks of age and retested for IRT. When an IRT+DNA protocol is used the genetic analysis can be performed on the original blood sample. Three stage protocols (IRT+Meconium+IRT and IRT+DNA+IRT) have been devised to reduce the false positive rate further.

In Australia, IRT+IRT protocols were introduced in the 1980s, but have been replaced by IRT+DNA since the early 1990s. NSW and Victoria have tested for $\Delta F508$ alone (accounting for over 70% of CFTR mutations in the region), and SA has performed analyses based on 5 genetic mutations (accounting for 80% of CFTR mutations). The IRT/DNA protocol in Australia is reported to have false positive rate of 0.05% and false negative rate of 5-6%; PPV of 5% after first test and 47% after second (Wilcken 1997).

Two randomised controlled trials of neonatal screening have been conducted and are discussed in a Cochrane systematic review (Merelle, Nagelkerke, Lees, & Dezateux 1999). The UK trial took place in 1985-1989 in Wales and the West Midlands, and involved approximately 474,142 infants, with a mean (SD) age of diagnosis of 9.1 (3.1) weeks in the screened group and 50.7 (60.5) weeks in the control group. The Wisconsin trial took place between 1985-1994, and involved 650,341 infants, initially screening with IRT only, and subsequently with an IRT/DNA two stage protocol. Mean (SD) age at diagnosis was 13 (36) weeks and 107 (118) weeks in the case and control groups respectively. For methodological reasons, only outcomes data from the Wisconsin trial was included in the Cochrane review.

In the most recent unpublished study from the Wisconsin group cited in the Cochrane review, mean z-scores for both weight and height were reported as significantly higher among screened compared with control patients up to the age of 11 years (Merelle et al. 1999). The proportion of patients with weight as well as height below the 5th percentile by the age of 11 years was significantly less among those screened compared with controls. The odds ratio for the risk of a weight below 5th percentile in the control group as compared with the screened group was 6.16 (95% CI 2.44, 15.57) and for height was 5.03 (95% CI 1.63, 15.57). Data about lung function, respiratory exacerbations, hospital admissions, hospital days, respiratory complications, diabetes mellitus, liver cirrhosis, and age at death have not yet been reported in comparable ways for the screened and unscreened groups. Longer term analysis is likely to provide more good-quality evidence.

Analysis of data from the Wisconsin trial subsequent to the two systematic reviews demonstrated that 50 CF patients, diagnosed early by neonatal screening and followed with comprehensive nutritional therapy, were able to achieve normal growth with overall height and weight at ~ 48th NCHS percentiles from birth to 12 years (Hui-Chuan L, Kosorok MR, Laxova A, Davis LA, FitzSimmon SC, & Farrell PM 2000). It should be noted that 6 children were excluded from this evaluation because their treatment did not commence in the first 12 weeks due to false negative screening tests, a false positive sweat test and late response to scheduled clinic visit. This illustrates the potential for a screening test to improve health outcomes, but that the potential is not always realised in the widespread implementation of a screening program.

An Australian study comparing a historical cohort of 57 children with CF born prior to the introduction of neonatal screening with 60 children born afterwards found significant differences in height, weight and lung function favouring the screened group (Waters, Wilcken, Irwig, Van Asperen, Mellis, Simpson, Brown, & Gaskin 1999). However management of CF is improving all the time and studies using historical comparisons have the limitation that patients in the screened cohort may also have had access to better treatment.

The Cochrane reviewers conclude that 'from the data available at this time, there is little evidence suggesting benefit from screening for CF in the neonatal period, although there is similarly little evidence of harm.'

A Health Technology Assessment (HTA) review of screening for CF has also been undertaken, however this is not a systematic review. While the search strategy is very comprehensive, there are no specified criteria for inclusion of studies or critical appraisal. All types of studies have been included, their various strengths and weaknesses discussed, and the resulting recommendations are based on an overview. The report reviewed the same two randomised trials as the Cochrane review, as well as five case-control studies, a study of sib-pairs and a trial of prophylactic versus symptomatic treatment of early disease. It reported that in their study of 20 screening programs over 5 million neonates have been tested with 'a low false-positive rate (0.5 per 1000), acceptable detection rate (90%), and favourable positive predictive value (33%).' Benefits were either predominantly short term or felt to be subject to strong statistical bias. The HTA reviewers conclude that 'the ability of screening to alter long-term prognosis has not been conclusively proven' and go on to say 'nevertheless there is some circumstantial evidence favouring a benefit.' Their recommendation is that, in view of the short term advantages (eg avoidance of prolonged hospitalisation), low false-positive rates and low cost of the test, that health authorities should consider introducing neonatal screening.

The benefits of screening for CF can be summarised as benefits to the health of the individual, improved quality of life for the individual and their family, and reduction in subsequent affected pregnancies by early recognition of parental carrier state. However, there are also possible adverse effects, particularly in those families who receive false positive results. These have been reported as concern, shock, disbelief, depression, anger and confusion. Even after being reassured with negative sweat test results, 85 of patients (n=104) reported they had

definitely changed their minds about future children and 22% were uncertain (Tluczek, Mischler, & Farrell 1992).

There are also some ethical issues associated with a neonatal screening program that should be considered. Screening to identify an individual affected with CF also identifies the carrier status of the parent. This could be considered unsolicited information and a breach of confidentiality.

There is an additional ethical consideration about whether neonatal screening should be in widespread use when pre-conceptual and antenatal screening is not. Is it justifiable to wait until an affected child is born before offering a diagnosis?

Cost effectiveness

Several studies have considered the actual costs of neonatal screening, however they yield little information about long term cost-benefits of a program. The Cochrane review reports the cost per case diagnosed from the Wisconsin study as \$US 7,613 for IRT and sweat tests, and \$US 7,403 for IRT/DNA and sweat tests. The cost per CF patient diagnosed on clinical grounds was estimated to be \$US 11,377.

An additional four studies were reported in the HTA review. The cost of a screening program was found to be fairly consistent across the studies, between \$6,000 - \$10,000 per case detected. (Murray et al. 1999) A CDC examination of the economic impact of CF screening reported that 'the Australian newborn screening protocol is the preferred prevention strategy because it detected the most CF affected newborns at the least cost'. (Qualls, Cono, Kelly, & Khoury 1997) An Australian study estimated the costs to be \$US 5.80 per newborn tested and \$US 18,710 per newborn detected as having CF (Wilcken 1997).

As there is uncertainty about the benefits of early detection to the patient, the family and society, it is difficult to quantify the cost-benefit of a neonatal screening program. The cost of treating a patient with CF can be very high, but also quite variable depending on severity, and a small change in long term outcomes could have a marked impact (either positive or negative) on the cost-benefit of the screening program.

Positive effects would be related to:

- Decreased costs related to delayed diagnosis ie multiple investigations and hospitalisations prior to the diagnosis being made. The Wisconsin study found that it was almost \$4,000 cheaper to diagnose CF by screening than by other methods.
- Decreased costs of medical care due to improved outcomes
 - short term improvements already identified (reduced hospitalisation time, fewer respiratory infections).
 - long term improvements still awaiting evaluation
- Lifetime benefits for patient and their family in improved quality of life and increased productivity. This could potentially be significant, but difficult to measure
- Reduced costs related to reproductive decisions made by parents who are aware of their carrier status early enough to prevent subsequent affected pregnancies. This could also have a flow-on effect through cascade testing of relatives.

Negative costs would be related to:

- Increased costs of prophylactic treatments
- Increased costs of medical care for extended life-span

Costs of a neonatal screening program are also dependent on whether pre-conceptual and antenatal screening is offered.

Quality of evidence

PAPER	STRENGTH OF EVIDENCE		Statistical precision	SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality			
M. E. Merelle, A. F. Nagelkerke, C. M. Lees, and C. Dezateux. Newborn Screening for cystic fibrosis. <i>The Cochrane Library (Oxford)</i> (Issue 3):1-16, 1999.	Level I	<p><i>Clearly defined question and inclusion criteria:</i></p> <p>A (high quality) A clear aim or research question AND a list of clear inclusion criteria was reported</p> <p><i>Comprehensive search</i></p> <p>A (high quality) Electronic databases and one or more of the following were used: the ‘grey’ literature, internet sources, conference proceedings, hand searches, reference lists, technical reports, experts, theses, etc AND the search was not restricted to the English language literature only</p> <p><i>3. Critical appraisal of the validity of studies reviewed</i></p> <p>A (high quality) Explicit criteria for critical appraisal were stated and the findings of that appraisal were considered in the conclusions and recommendations</p> <p><i>4. Consistency of results</i></p> <p>Methodological issues in one of the RCTs meant that comparison of results could not be carried out</p> <p><i>5. Number of studies included in review - Two</i></p>	<p>The Odds Ratio for the risk of a child with CF having a growth measurement below the 5th percentile in the control group compared with the screened group was: OR (95% CI)</p> <p>Weight 6.16 (2.44 – 15.57)</p> <p>Height 5.03 (1.63 – 15.63)</p>	<p>Although there is a fairly wide confidence interval, it appears that the risk of lower height and weight is clinically significant</p>	<p>These are appropriate outcome measures</p>

Generalisability of evidence

The HTA review compares regions with high mutation detection rates, regions with lower mutation rates, Asians, Ashkenazi Jews and Blacks - and in all 5, the positive predictive value is 1 in 4. Information from Australian studies is similar to that from other international findings.

6.3.5.3. CONCLUSIONS

Fair evidence to recommend screening

Comments:

The general expert consensus is that earlier diagnosis through screening is beneficial. However, this benefit is likely to be modest.

Short term advantages of reduced morbidity and nutritional benefits have been demonstrated.

With improved medical care, individuals now typically live into early middle age, so that cystic fibrosis is no longer usually a lethal condition in childhood. Systematic reviews report that there is as yet no conclusive evidence that neonatal screening alters long term prognosis in cystic fibrosis. As therapy through childhood, adolescence and early adulthood continues to improve, the relative benefits of screening may lessen.

Neonatal screening is currently available in most Australian states. There are short term advantages to early detection, the test is relatively low cost and has low false-positive rates, and screening is part of the current neonatal heel prick screening program.

The brief of this review was to focus on neonatal screening; pre-conceptual and antenatal screening programs also require debate.

6.3.5.4. RECOMMENDATIONS

Continuation of current neonatal screening programs for cystic fibrosis is recommended.

6.3.5.5. FURTHER RESEARCH

A recent Cochrane review states that long term follow up and analysis based on 'intention to treat' is required and note that this is under way in RCTs in the US and the UK. HTA reviewers call for re-analysis of existing research information.

Possible research areas:

- Evaluation of the consequences of detecting carriers as a result of neonatal screening
- RCTs to evaluate specific early treatments in individuals detected through screening
- Development and evaluation of audit procedure to ensure parents give informed consent for neonatal screening
- Continuation of research into gene replacement therapy
- Ongoing study of parent-reported outcomes, including impact of screening on reproductive decisions and quality of life

6.3.5.6. SEARCH TERMS

MeSH headings: Cystic fibrosis

Key words: cystic fibrosis

6.3.5.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Reference group: Bob Williamson, Garth Alperstein

Clinical experts: Phil Robinson, John Massie (Paediatric Respiratory Physicians, Dept of Thoracic Medicine, Royal Children's Hospital Melbourne)

6.3.5.8 REFERENCES

- Farrell, P. M., Kosorok, M. R., Laxova, A., Shen, G., Kosciak, R. E., Bruns, W. T., Splaingard, M., & Mischler, E. H. 1997, "Nutritional benefits of neonatal screening for cystic fibrosis. Wisconsin Cystic Fibrosis Neonatal Screening Study Group", *New England Journal of Medicine*, vol. 337, no. 14, pp. 963-969.
- Gregg, R.G., Simantel, A., Farrell, P.M., Kosciak, R.E., Kosorok, M.R., Laxova, A., Laessig, R. H., Hoffman, G., Hassemer, D., Mischler, E. H., & Splaingard, M. 1997, "Newborn Screening for Cystic Fibrosis in Wisconsin: Comparison of Biochemical and Molecular Methods", *Pediatrics*, vol. 99, no. 6, pp. 819-924.
- Hui-Chuan, L., Kosorok, M.R., Laxova, A., Davis, L.A., FitzSimmon, S.C., & Farrell, P.M. 2000, "Nutritional status of patients with cystic fibrosis with meconium ileus: a comparison with patients without meconium ileus and diagnosed early through neonatal screening", *Pediatrics*, vol. 105, no. 1, pp. 53-61.
- Merelle, M. E., Nagelkerke, A. F., Lees, C. M., & Dezateux, C. 1999, "Newborn screening for cystic fibrosis", *The Cochrane Library (Oxford)* no. Issue 3, pp. 1-16.
- Murray, J., Cuckle, H., Taylor, G., Littlewood, J., & Hewison, J. 1999, "Screening for Cystic fibrosis", *Health Technology Assessment*, vol. 3, no. 8, pp. 1-103.
- Qualls, N. L., Cono, J., Kelly, A. E., & Khoury, M. J. 1997, "The economic impact of population based newborn screening for cystic fibrosis", *MMWR - Morbidity & Mortality Weekly Report*, vol. 46, no. RR16, pp. 1-22.
- Tluczek, A., Mischler, E. H., & Farrell, P. M. 1992, "Parents' knowledge of neonatal screening and response to false-positive cystic fibrosis testing", *Journal of Developmental & Behavioral Pediatrics*, vol. 13, pp. 181-186.
- Waters, D. L., Wilcken, B., Irwig, L., Van Asperen, P., Mellis, C., Simpson, J. M., Brown, J., & Gaskin, K. J. 1999, "Clinical outcomes of newborn screening for cystic fibrosis", *Archives of Diseases in Childhood Fetal Neonatal Edition*, vol. 80, p. F1-F7.
- Wilcken, B. 1997, "Newborn Screening for Cystic Fibrosis in Australasia". *Newborn Screening for Cystic Fibrosis: A Paradigm for Public Health Genetics Policy Development Proceedings of a 1997 Workshop*.

6.3.6. DEVELOPMENTAL DYSPLASIA OF THE HIP

6.3.6.1. BACKGROUND

Clinical picture

Congenital dislocation of the hip (CDH) is the traditional term referring to a disorder of the joint in which the head of the femur is partly or completely displaced from the acetabulum. However hips are infrequently 'dislocated' at birth, but may be 'dislocatable'. It is more common for hips to dislocate after delivery, so it is inaccurate to refer to them as 'congenital'. 'Developmental Dysplasia of the Hip' (DDH) is now preferred over the traditional term as it also considers the relationship between dysplasia and stability. DDH covers the same range of hips problems as CDH (ie dislocation, subluxation and instability) but also includes poorly developed joints that may not actually dislocate and abnormalities determined after the newborn period.

For normal development of the hip to occur, there must be a balance between growth of the acetabular and tri-radiate cartilages and a well centred femoral head (Weinstein 1987). This balance may be affected by the intrauterine environment. Risk factors for DDH include family history, breech presentation, deformities of the foot, first born babies and oligohydramnios. Female infants are reported to have 4-6 times greater incidence.

Natural history

The majority of unstable hips detected at birth will stabilise in a very short time. Weinstein cites several studies that reveal rates varying between 22% returning to normal by 3 months and 88% within 2 months (Weinstein 1987). These differences may also be due to variations in definition. In addition, these studies also demonstrate that while many 'unstable' hips will stabilise shortly after birth, others will go on to subluxate, dislocate or retain dysplastic features. There is no evidence to predict those that will go on to improve and those that will deteriorate.

Late detection of dislocation has been attributed to failure of screening, however several authors cite studies suggesting that children who appear normal at birth on clinical examination can go on to develop a dislocation (Chan, Cundy, Foster, Keane, & Byron-Scott 1999; Weinstein 1987) (French & Dietz 1999) (Jones 1994). One hypothesis is that the acetabulum was shallow at birth but was not a cause of instability until the child was weightbearing. Another is that the repeated stresses of the maneuvers in the screening examination predispose the joint to later instability. Ultrasonographic examination reveals abnormalities not detected clinically, hence the higher rate of detection and also the higher rate of spontaneous return to normal observed in more recent studies.

Children with untreated DDH can go on to develop significant pathology and disability – degenerative joint disease, loss of joint space, cyst formation, etc and may have back pain, limb length inequality, knee deformity and pain, scoliosis and gait abnormalities.

Prevalence

The mean incidence of CDH requiring surgery in northern European populations prior to the introduction of neonatal screening was 1.25 per 1000. Following introduction of the Ortolani-Barlow screening tests, the incidence of neonatal hip instability was reported to be between 2.5 and 20 per 1000 (Hall 1996). This rate has risen again with the advent of ultrasound screening to 4 – 52% (Bialik, Bialik, Blazer, Sujov, Wiener, & Berant 1999). The UK National Orthopaedic Surveillance Scheme reports the incidence of first operative procedure as 0.78 per 1000 live births (Godward & Desateux, 1998). These findings highlight the variation in definitions and diagnostic criteria used and the differences between the clinical and sonographic findings.

It is suspected that genetic and ethnic factors play a significant role as there is an enormous variability across racial groups. Incidence has been reported as high as 25-50 per 1000 in Lapps and North American Indians and 75 per 1000 in Yugoslavia, yet this condition is almost non-existent in Chinese and Blacks (Yiv, Saidin, Cundy, Tgetgel, Aguilar, McCaul, Keane, Chan, & Scott 1997) (Weinstein 1987) (French & Dietz 1999). The variations might also be due to environmental factors, diagnostic criteria used, experience of the examiner or the age of the child.

Due to the variations in terminology, it is important when considering incidence rates to recognise the differences between reporting of CDH requiring surgery and DDH resulting in hip instability. It has been estimated that 1 in 100 newborns will have some evidence of instability, but the true incidence of dislocation is between 1-2 per 1000, a figure similar to the rates of surgery (Weinstein 1987) (French & Dietz 1999).

Genetics relevant to screening issues

Genetic factors are suspected but none known at present

Diagnosis

Clinical examination

The Ortolani and Barlow maneuvers have been used to examine the hips of newborn infants for nearly 40 years. The Ortolani sign is the palpable sensation of the femoral head moving in and out of the acetabulum. If positive, it identifies a dislocated hip that is still reducible in the early weeks of life, but will require active treatment. Barlow's test identifies an unstable hip that is lying in the reduced position but can be passively dislocated.

Both tests are conducted with the infant in the supine position and the hips are examined one at a time. The Ortolani maneuver requires the infant's hips and knees to be flexed to 90° and the contralateral hip held still while the thigh of the hip being tested is gently abducted and pulled anteriorly. In a positive finding, there is a palpable and sometimes audible 'clunk' as the femoral head is brought from its dislocated posterior position and relocated into the acetabulum. In less well developed acetabulae, in which the joint is more unstable, the 'clunk' is less pronounced. This means that the maneuvers must be performed extremely carefully to avoid missing the subtle signs that indicate a more severe condition. In Barlow's test the hip is flexed and the thigh is adducted and pushed posteriorly. This causes the femoral head to dislocate posteriorly from the acetabulum. The diagnosis is confirmed by performing the Ortolani maneuver to relocate the hip.

If an unstable hip remains undetected for many weeks, it becomes more difficult to reduce the femoral head into the acetabulum and limitation of abduction becomes the prominent clinical finding. The Ortolani test becomes negative. Limitation of hip abduction, shortening of the femur, asymmetric skin folds and telescoping of the affected hip are thus late signs.

Ultrasound

Ultrasound examination was introduced for detecting developmental dysplasia of the hip in 1978. Two techniques are used, either alone, or in combination, depending on the screening protocol being implemented. The Graf method is a static examination of the hip, using a classification based on 4 'types' of hip joint. Type I is considered normal. In Type II the acetabular cup is shallow, this is considered to be immature in infants under 3 months, but abnormal in older babies. Type III is subluxated and Type IV is dislocated. Many screening programs also include a dynamic examination of the hip where evaluations are made using real time visualisation of the joint in different positions, or when stressed.

Defining a gold standard for diagnosis of DDH is difficult, some studies use ultrasound to assess clinical examination, others use specialist opinion, based on clinical examination, to assess ultrasound. This is complicated further by the variations in terminology; whether we are measuring CDH or DDH; and the fact that ultrasound will detect many more abnormalities than clinical examination but many of these will go on to spontaneous resolution, begging the question of whether they were clinically significant.

Treatment/Management

Management is aimed at reduction of any dislocation; retention and immobilization of the femoral head in the acetabulum; and maturation of the joint. The type of treatment used depends on the age at which the diagnosis was made.

Birth

Treatment of an unstable hip diagnosed at birth consists of maintaining the position of the hip in flexion and abduction for about 1-2 months. Held in this position the femoral head is reduced, the ligamentous structures around the joint can tighten and there is stimulation of normal formation of the hip socket. There are several ways to achieve this, the Pavlik harness is the most common, but the Frejka splint and other devices are also used. The treatment must be continued until the hip is stable and x-rays or ultrasound examinations are normal.

1-6 months

True dislocations may be diagnosed at this time and a hip spica cast or fixed orthosis is required to maintain the femoral head in the acetabulum. If reduction is achieved, treatment is continued until radiological parameters are normal. If reduction is not achieved, then a closed reduction (manipulation under general anaesthesia) is performed.

6-18 months

Surgical closed reduction (manipulation under anaesthesia) is the major method of treatment.

After 18 months

The older the child at diagnosis, the less likely a closed reduction will be successful. By 18 months, open surgical intervention is necessary to realign the hip. A hip spica cast is worn postoperatively for 6-8 weeks.

Complications of treatment

Avascular necrosis is recognised as a complication of the Pavlik harness, possibly due to improper application of the harness, but may also be due to the disease process itself.

Prevention

There are no interventions to prevent DDH

How might screening reduce the burden of suffering?

Early recognition of DDH allows simple effective treatments to be instigated before complications arise or more invasive procedures are required.

6.3.6.2 EVIDENCE

Tests

Clinical examination of the hips using the Ortolani and Barlow techniques is routinely performed as a screening test and ultrasound has also been utilised for screening in some settings. Determination of sensitivity and specificity has been difficult due to the lack of an obvious gold standard. Since there is no confirmatory test, the extent to which a diagnosis is correct or not cannot be accurately assessed.

Clinical examination by hospital house staff has been assessed against that of an experienced orthopaedic surgeon and found to have sensitivity of 30% and specificity of 98.4% (Jones 1994). In an Israeli study of 9199 infants, clinical examination was compared with ultrasound, and if a discrepancy existed, a positive diagnosis was reached by consensus of experienced clinicians - sensitivity was reported as 54% and specificity as 100% (Rosenberg, Bialik, Norman, & Blazer 1998). There are obvious limitations to these methodologies, but they are suggestive of low sensitivity.

There appears to be significant variability in the accuracy of clinical screening related to the skill of the operator. In a study of the total birth cohort in one obstetric clinic in Bulgaria over a 6 year period (n = 20,417), all children were examined by an experienced clinician (approximately 1000 were examined by a second physician covering annual leave). This region has only one paediatric orthopaedic unit that deals with all hip problems, and an audit of these records was used to detect late presentations of CDH. 124 children had abnormal findings on clinical examination, 96% were diagnosed at neonatal screening and 4% at secondary screening at 2-3 months. No new cases were diagnosed after 3 months, suggesting that all cases were detected by screening, or a sensitivity of 100% (Darmonov & Zagora 1996). This is in contrast to a UK audit of the National Orthopaedic Surveillance Scheme which revealed that CDH had been missed in routine clinical screening in 70% of cases (Godward & Desateux, 1998). The incidence of a first operative procedure for CDH (0.78 per 1000 live births) was similar to that reported before screening was introduced in 1969 (0.66-0.85/1000). Australian figures are more reassuring, although they demonstrate there is still significant room for improvement. An audit conducted in South Australia found that 24% of known cases of CDH had been detected late and required surgery (Chan et al. 1999). The incidence of surgery in this SA cohort was 0.46 per 1000 live births.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	Clinical exam	U/S	Criteria for a Screening Program Wilson & Jungner 1968	
Simple, quick and easy to interpret	✓	✓	Important health problem	✓
Acceptable to the public	✓	✓	Accepted treatment	✓
Accurate	?	?	Facilities for diagnosis and treatment	✓
Repeatable	?	?	Latent or early symptomatic stage	✓
Sensitive	?	?	Suitable test or examination	?
Specific	?	?	Test acceptable to the population	✓
			Natural history adequately understood	✗
			Agreed policy on whom to treat	✗
			The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

Clinical examination

Clinical examination of the hips is routinely performed in the neonate examination and again at the 6 week post-natal check.

Ultrasound examination

Some screening programs have included ultrasound examinations - at either a population level, or for a target group of infants identified as having risk factors for DDH. In most cases this is in addition to the clinical examination.

Numerous prospective studies considering the effectiveness of clinical examination and ultrasound have been published. This large body of literature is not critically appraised here since a systematic review on this topic has been undertaken and is due to be published soon.

The Canadian Task Force on Preventive Health has received a systematic review 'Screening of Term Newborns for Developmental Dysplasia of the Hip'. We are unable to cite this review prior to publication, however the author presented the research to the Pediatric Academic Societies' 1999 Annual Meeting. The author's objective was to compare the effectiveness of screening for DDH by serial clinical examination and ultrasound. The conclusion in the published abstract states 'Although both screening methods have reduced the rate of operative intervention equally, the high false positive rate, moderate reproducibility and the unclear relevance of asymptomatic sonographic DDH make ultrasound a poor screening tool' (Patel 1999).

Level II evidence is available from a Norwegian randomised controlled trial (Rosendahl et al 1994). This study of 11,925 infants, the total birth cohort in one maternity hospital between 1988 and 1990, found no advantage to adding ultrasound screening to clinical examination. Infants were allocated to one of three groups - the control group underwent clinical examination only, those in the selected screening group were given an ultrasound examination if they had risk factors for DDH (defined in this study as positive findings on clinical examination, breech position or close family history of DDH), and those in the third group were all screened with both clinical examination and ultrasound. Operation rates and incidence of late presentations of DDH were used to assess efficacy of screening.

The method of randomisation has some limitations as allocation to the intervention groups was based on bed availability in the wards housing the different study cohorts, and subjects were allocated to the control group when the ultrasonographer was on leave (periods of 1-3 weeks spread unsystematically throughout the year). While there are limitations to this method of randomisation, the authors felt that having patients in different wards avoided communication between groups that could create a bias in reporting of risk factors. The only apparent systematic bias in this method of randomisation was that an increased number of infants born by caesarean section were admitted to the unit allocated to the universal ultrasound arm, resulting in a higher

rate of breech presentations in this group. However, none of the cases of late presentation of DDH were children in this category. There were no differences between the groups on sex, positive clinical findings or total number of infants with risk factors.

An additional limitation of this study is that the incidence of subluxation or dislocation was half that of previous years (1.3 per 1000, previously 2.6). This unexpectedly low rate reduced the statistical power of the study.

Experienced paediatric clinicians conducted the clinical examinations and a single physician, blinded to clinical findings and risk factor status, performed the ultrasounds. Treatment was initiated if the hips were judged to be persistently dislocatable or dislocated on clinical examination or were classified as Graf type IIIa or worse on U/S, even if they were clinically stable. Clinically stable but sonographically immature or slightly dysplastic hips were followed by U/S and clinical examination every 4 weeks until normal or treatment was commenced due to failure of improvement.

The main findings of this study were:

- There was no statistically significant difference between the groups for operation rates or incidence of late presentations of DDH.
- U/S screening resulted in a higher rate of treatment with abduction splints when all infants were screened, but not when screening was restricted to neonates with recognised risk factors. (3.2% universal U/S, 2.0% selective U/S, 1.8% clinical exam only, $p < 0.0001$). Complications related to treatment were not reported.
- U/S screening resulted in a higher follow up rate with repeat U/S and clinical examinations (13%, 1.8%, 0% $p < 0.0001$). 97% of these infants went on to spontaneous resolution by 3 months of age.

Further information about the value of U/S in the diagnosis and management of DDH will soon be available from a RCT recently completed in the UK; data analysis is currently underway. (personal communication Anne Quinn) This study also included economic effectiveness - considering monetary costs to the family and health service as well as psychological costs to the family. The trial commenced in December 1994 and recruited 629 infants with a clinical diagnosis of unstable hips and followed them for 2 years. The study hypothesis is that the additional information provided by the U/S will reduce the likelihood of splinting treatment in the intervention group and that this reduction will be achieved without doubling the risk of late treatment. A secondary hypothesis is that the extra costs associated with U/S scanning will be balanced by a reduction in other costs.

Potential negative outcomes from screening for DDH must also be considered. Ultrasound examinations detect dysplasia that may well resolve spontaneously, resulting in unnecessary treatment which contributes to excess costs to the community and possible adverse sequelae for the patient. Treatment with splints carries a small but significant risk of avascular necrosis of the femoral head. An additional concern is that repeated clinical examination has been linked by some authors to late dislocation. Although a causal relationship has not established, the role of repeat follow up examinations should be considered in light of possible adverse outcomes, and further research conducted in this area.

Cost effectiveness

The NHS Centre for Reviews and Dissemination has published a review of a cost-effectiveness evaluation of ultrasound screening for CDH (Geitung, Rosendahl, & Sudmann 1996).

The authors conclusions are 'that although U/S screening would result in fewer cases of late-detected CDH, a general screening program applied to the total population of newborn infants was not cost-effective. However, screening for those identified as being at greater risk may bring additional benefits and be cost-effective. Moreover, if the screening program adopted only U/S testing and the clinical examinations were eliminated the program would become cost-effective'.

The reviewers point out some limitations of this study, but find that the conclusions are likely to be justified. The authors recommend further study to assess the quality of life in cases of late-detected CDH and in patients in later stages of their lives when the condition would become more costly to treat.

Quality of evidence

PAPER	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
Rosendahl et al 1994	Level II RCT	<p><i>1. Randomisation and concealment of allocation: C (unsatisfactory quality)</i></p> <p>Subjects were allocated by bed availability to the 2 intervention arms, and to the control group when the radiologist was on leave</p> <p><i>2. Blinding of outcomes:</i></p> <p>B (some limitations) Ultrasonographer blinded to clinical findings and risk factor status. No other details of blinding given</p> <p><i>3. Withdrawals and dropouts</i></p> <p>A (high quality) All subjects entering the trial were accounted for at its completion (there was some missing data – but this was not relevant to the calculations of outcomes)</p> <p><i>4. Number of subjects included in trial 11,925</i></p>	<p>Prevalence of late presentation of DDH and operative management was not statistically significant between groups (p=0.11)</p> <p>U/S screening resulted in a higher rate of treatment (p<0.0001)</p> <p>U/S screening resulted in a higher follow up rate (p<0.0001)</p> <p>The unexpectedly low incidence of DDH in this study may have reduced the statistical power to determine differences between the intervention and control groups</p>	<p>Prevalence of late presentation of DDH - 0.3% universal U/S, 0.7% selective U/S, 1.3% clinical exam only (p<0.11).</p> <p>Rate of treatment with abduction splints - 3.2% universal U/S, 2.0% selective U/S, 1.8% clinical exam only (p<0.0001).</p> <p>Follow up - 13% universal U/S, 1.8% selective U/S, 0% clinical exam only (p<0.0001).</p>	Appropriate outcome measures were used

Generalisability of evidence

Clinical examination for DDH is conducted in the same manner in most programs, however it is interesting to note that Australian studies suggest better detection rates. Ultrasound facilities and experience of operators in Australia should be equivalent to similar practices overseas.

6.3.6.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening by clinical examination

Fair evidence to recommend against screening by ultrasound

The neonatal and 6 week postnatal examinations offer routine access to newborns to detect congenital abnormalities. Although there are limitations to clinical examination as a screening tool in this setting, it may be the only opportunity to detect developmental dysplasia of the hip in the pre-symptomatic phase.

However, it appears that hip examination frequently falls short of *program* requirements for an effective detection system. Although Australia has fewer cases missed by screening and a lower incidence of surgery than the UK where a similar screening regimen is in place, there is still considerable room for improvement. Given the uncertainty of the natural history, the knowledge that some cases will resolve spontaneously and the hypothesis put forward by some authors that repeated examinations may increase instability, further research is required in this area.

Consideration of strategies to improve the sensitivity and specificity of clinical examination is also required.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

The evidence suggests that ultrasound is not useful as a universal screening tool, and may result in unnecessary treatment and increased costs. However ultrasound screening for children at high risk of developmental dysplasia of the hip may have some merit.

6.3.6.4. RECOMMENDATIONS

Universal ultrasound screening for developmental dysplasia of the hip is not recommended.

Although there is little firm evidence to support the value of clinical screening for developmental dysplasia of the hip, we recommend continuation of newborn and 6 week examination using the Ortolani and Barlow maneuver, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical and nursing staff
- clear pathways and/or guidelines for referral
- standardised follow up procedures for infants with abnormal findings on clinical examination
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records
- strategies to improve attendance at the 6 week postnatal check (particularly for disadvantaged groups)
- opportunistic examination of children not examined in the newborn period

Measures to increase the effectiveness of a screening program for developmental dysplasia of the hip should be implemented and the incidence of surgery and late presentations monitored. If these rates are not reduced by improvements in the screening program, the benefits of clinical screening need to be examined further.

6.3.6.5. FURTHER RESEARCH

- Evaluation and continuing quality improvement of programs
- monitoring of the incidence of surgery and late presentations of CDH through population based data collection. If there is no reduction in these outcomes following improvements to the screening program, the role of clinical screening should be further evaluated

- Further examination of the role of ultrasound screening in infants *at risk* of developmental dysplasia of the hip. Results from the UK RCT, when available, should inform this process
- Teanby et al recommend a prospective multicentre trial of untreated dysplasia diagnosed by ultrasound. This would allow treatment to be advocated for certain types of dysplasia if the trial confirms persistence or progression of dysplasia (Teanby & Paton 1997)

6.3.6.6. SEARCH TERMS

MeSH headings: Hip dislocation, congenital

Keywords: Congenital hip dislocation

6.3.6.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Reference group: Colin Morley, Caroline Briggs

Clinical experts: Ian Torode Dept of Orthopaedics, Royal Children's Hospital, Melbourne

6.3.6.8. REFERENCES

Bialik, V., Bialik, G., Blazer, S., Sujov, P., Wiener, F., & Berant, M. 1999, "Developmental dysplasia of the hip: a new approach to incidence", *Pediatrics*, vol. 103, no. 1, pp. 93-99.

Chan, A., Cundy, P. J., Foster, B. K., Keane, R. J., & Byron-Scott, R. 1999, "Late diagnosis of congenital dislocation of the hip and presence of a screening programme: South Australian population-based study", *Lancet*, vol. 354, no. 9189, pp. 1514-1517.

Darmonov, A. V. & Zagora, S. 1996, "Clinical screening for congenital dislocation of the hip", *Journal of Bone & Joint Surgery American Volume.*, vol. 78-A, no. 3, pp. 383-388.

Godward, S., Dezateux C. Surgery for congenital dislocation of the hip in the UK as a measure of outcome of screening. *Lancet* 1998; 351:1-9.

French, L. M. & Dietz, F. R. 1999, "Screening for developmental dysplasia of the hip", *American Family Physician*, vol. 60, pp. 177-188.

Geitung, J. T., Rosendahl, K., & Sudmann, E. 1996, "Cost-effectiveness of ultrasonographic screening for congenital hip dysplasia in new-borns", *Skeletal Radiology*, vol. 25, no. 3, pp. 251-254.

Hall, D.M.B. 1996, *Health for All Children*, 3rd edn, Oxford University Press, New York.

Jones, D. A. 1994, "Principles of screening and congenital dislocation of the hip", *Annals of the Royal College of Surgeons of England*, vol. 76, no. 4, pp. 245-250.

Patel, H. "Is screening for developmental dysplasia of the hip in newborns worthwhile?", Pediatric Academic Societies', San Francisco.

Rosenberg, N., Bialik, V., Norman, G., & Blazer, S. 1998, "The importance of combined clinical and sonographic examination of instability of the neonatal hip", *International Orthopaedics*, vol. 22, pp. 185-188.

Rosendahl K, Markestad T, Lie RT. Ultrasound screening for developmental dysplasia of the hip in the neonate: the effect on treatment rate and prevalence of late cases. *Pediatrics* 1994;**94**:47-52.

Teanby, D. N. & Paton, R. W. 1997, "Ultrasound screening for congenital dislocation of the hip: a limited targeted programme", *Journal of Pediatric Orthopedics*, vol. 17, pp. 202-204.

Weinstein, S. L. 1987, "Natural history of congenital hip dislocation (CDH) and hip dysplasia.", *Clinical Orthopaedics & Related Research* no. 225, pp. 62-76.

Yiv, B. C., Saidin, R., Cundy, P. J., Tgetgel, J. D., Aguilar, J., McCaul, K. A., Keane, R. J., Chan, A., & Scott, H. 1997, "Developmental dysplasia of the hip in South Australia in 1991: prevalence and risk factors", *Journal of Paediatrics & Child Health*, vol. 33, no. 2, pp. 151-156.

6.3.7a. CONDUCTIVE HEARING IMPAIRMENT

6.3.7a.1. BACKGROUND

Clinical picture

Childhood conductive hearing impairment is usually due to otitis media with effusion (OME). OME is characterised by the presence of inflammation and fluid (effusion) in the middle ear cavity with an intact tympanic membrane. The fluid may be serous (thin, watery liquid), mucoid (thick, viscous liquid), or purulent (pus-like).

Opinion is sharply divided as to the burden of suffering caused by OME. In contrast to acute otitis media, which also features effusion with or without hearing impairment, pain, fever and malaise are not usually present. The only certain feature associated with OME is mild conductive deafness in about 50% of ears while effusion is present. The mean hearing loss over four frequencies is 20-25 dB (range 0-60dB) which may be present for days, weeks or months. A very small proportion of children may have persistent moderate hearing impairment (the range in which children with sensorineural hearing impairment would be fitted with hearing aids).

In recent years OME has been thought to cause deficits in expressive and receptive language, behaviour, academic achievement, intelligence (both performance and verbal), social skills, balance, sleep, coordination, and a range of other features. The major postulated pathway by which OME would cause language, behavioural and academic problems is via the associated hearing impairment; OME without hearing impairment would not be expected to lead to these adverse outcomes. Despite this, most studies have reported outcomes for OME rather than for associated hearing impairment. Although the mean hearing loss at any point in time for a child with OME is 20-25 dB, this translates to a much smaller proportion of children with persistent OME who also have persistent hearing impairment. A further difficulty is that we do not know the level of severity of hearing impairment beyond which adverse outcomes become more likely, and this may vary by language function. Most screening programs for permanent childhood hearing impairment aim to detect only hearing impairments greater than 35-40 dB HL on the basis that outcomes for hearing impairments milder than this are insufficiently understood and the harms and benefits of intervention are not known. If such a cutpoint were applied to children with OME, only a small proportion would be studied prospectively. Inclusion of the larger number of children with OME *without* persistent hearing impairment could greatly reduce the change of finding true effects for those worst affected. However, one would still expect to be able to detect trends towards adverse outcomes of hearing impairment particularly if using severity of hearing impairment as a predictor variable in multivariate analysis. It has frequently been suggested that the fluctuating nature of the hearing loss associated with OME may contribute to adverse outcomes over and above the severity of the hearing loss itself, but this has not been supported by the little rigorous study available (Peters, Grievink, van Bon & Schilder 1994).

Good prospective cohort studies of outcomes attributable to OME are now available. We summarise here those identified through our search that (a) studied whole cohorts identified at a specified age (b) achieved acceptable retention rates, (c) used objective measures of ascertainment and (d) examined frequently and regularly (varying across studies from fortnightly to three monthly) according to schedules specified by the researchers rather than by parents. In essence, these studies (summarised in the next two paragraphs) have failed to report significant adverse outcomes for persistent OME.

The Pittsburgh study reported on a birth cohort of 2253 children followed for the first 24 months of life with intensive (at least monthly) pneumatic otoscopy supplemented by tympanometry, in order to quantify the amount of OME experienced by American babies (Paradise, Rockette, Colborn, Bernard, Smith, Kurs-Lasky & Janosky 1997). The prevalence of OME in this cohort was very high (see "Prevalence", below). The most severe and persistent cases of OME were randomised to early or late treatment with tympanostomy tubes (see "Treatment/Management", below). Outcome data have been reported for children not meeting criteria for randomisation, many of whom nonetheless experienced large amounts of OME over the first 24 months. Parents of 2156 children reported the MacArthur Communicative Development Inventory-Words and Gestures when the child was 1 year of age and the MacArthur Communicative Development Inventory-Words and Sentences when the child was 2 years of age. Only one of five subscales of vocabulary production at age 2 was weakly correlated with cumulative percentage of days with middle ear effusion, which accounted for only a negligible percentage of the variance in scores on this scale. None of four subscales at age 1 was associated with middle ear effusion (Feldman et al. 1999). Parents of 2278 children completed the Parenting Stress Index at 1, 2 and 3 years of age and Child Behavior Checklists at 2 and 3 years of age. No significant relationships, or trends to significant relationships, were found between parents' ratings of parent-child stress or of their children's behaviour problems and the cumulative duration of OME (Paradise, Feldman, Colborn, Campbell, Dollaghan, Rockette, Janosky, Kurs-Lasky, Bernard & Smith 1999). These findings are similar to those reported by Roberts et al who have followed two much smaller cohorts of

disadvantaged children from infancy in long day care settings to adolescence. This group have reported no adverse effects on academic achievement, intelligence, or language, although some relationships were seen with narrative and discourse skills and attention at age 4 years (Roberts, Sanyal, Burchinal, Collier, Ramey & Henderson 1986; Roberts, Burchinal, Davis, Collier & Henderson 1991; Feagans, Sanyal, Henderson, Collier & Appelbaum 1987; Roberts, Burchinal, Collier, Ramey, Koch & Henderson 1989; Arcia & Roberts 1993; Roberts, Burchinal & Campbell 1994; Roberts, Burchinal & Clarke-Klein 1995; Roberts, Burchinal, Medley, Zeisel, Mundy, Roush, Hooper, Bryant & Henderson 1995).

Three high quality population-based studies have followed older children. In the Nijmegen study, academic outcomes were reported for 7-8 year old children with and without persistent OME at age 2-4 years (Peters et al 1994). Tests of reading and intelligence, and teacher ratings of reading, writing and arithmetic, did not differ between OME and OME-free groups, though tests of spelling were poorer in the OME group. Lous followed children originally aged 3-4 years at 8.5-9 years, reporting similar receptive language and intelligence for OME and OME free groups (Lous, Fiellan-Nikolajsen & Jeppesen 1988). In a further study, Lous studied relationships between OME and reading achievement in an unselected cohort of 366 8-year-old children (Lous 1993). The children underwent 10 tympanometry and 5 pure tone audiometry assessments during their first year at school. At the beginning of the second grade they all had a Silent Reading Word Test (OS-400). There was a small correlation between silent reading score and type B tympanograms, but not with hearing levels. Type B tympanogram explained only 2% of the total variance in outcomes.

Common to many of these studies is the large amount of variance due to social gradients and the quality of the home environment.

A very small number of children have significant congenital or progressive hearing impairment which neither fluctuates nor resolves. Many of these children have other associated problems, such as head and neck malformations, which assist in diagnosis. These children require appropriate long term management of their persistent hearing impairment (eg surgery, hearing aids).

Prevalence

OME is a fluctuating condition which is postulated to exert its effects via prolonged hearing loss, and therefore cumulative prevalence is of greater interest than point prevalence. The best study of cumulative prevalence in infants to date is the Pittsburgh study described above. The mean cumulative proportions of days with OME were 20.4% in the first year of life and 16.6% in the second. The proportions developing ≥ 1 episode of OME between 2 months of age (the starting point) and ages 6, 12 and 24 months were 48%, 79%, and 91% respectively. The main factors contributing to time spent with OME were socioeconomic index and amount of exposure to other children. Breastfeeding and smoking were both weakly predictive in the first year of life only (Paradise et al 1997). Similarly, an English study of 95 infants assessed with tympanometry at monthly intervals from birth to three years found that 17 (17.9%) children had unilateral or bilateral OME for more than half of their first three years of life, and 33 (35%) had tympanograms suggestive of OME at more than a third of observations.

Cumulative prevalence of OME over the ensuing years of life has not been so intensively documented. Between 2 and 6 years approximately 10-20% of children will have OME at any one time, with a sharp drop in prevalence after 6 years of age (Anonymous 1992).

Australian studies indicate similarly high prevalence. A recent Australian study of 252 children under 4 years of age attending Darwin day care centres demonstrated the presence of OME on average 4.4 times per child over 12 fortnightly tympanometry examinations (37% of all examinations conducted) (Skull et al. 1999). A 1997 study of 1563 school entry children in Victoria found 25% to have a high likelihood of OME in at least one ear based on abnormal screening audiometry and tympanometry in late winter and early spring. At repeat audiometry at 8-12 weeks, 52 (3.3%) had unilateral or bilateral type B tympanogram with normal hearing, 57 (3.6a%) had unilateral and 47 (3%) had bilateral hearing loss (Heathershaw & Wake 2000).

Genetics relevant to screening issues

None for the general population, although the tendency to develop OME is probably genetically mediated. OME is almost universally present in some genetic disorders (Downs syndrome) and some disorders with strong heritability (cleft palate) for whom ongoing surveillance may be required.

Diagnosis

The gold standard for diagnosis of OME is myringotomy or tympanocentesis, in which presence of middle ear fluid is determined directly via a small incision in the tympanic membrane. In practice, this occurs principally in the setting of tympanostomy tube insertion, ie in children with an extremely high prior probability of presence of effusion. In the clinical setting, OME is diagnosed most reliably by tympanometry or by pneumatic otoscopy (see below). Its presence is frequently inferred when mild hearing impairment is noted clinically or on audiometry, since it is far and away the most common cause of mild hearing

impairment in childhood, or in the presence of recurrent acute otitis media. However, audiometry is not a test specific for OME.

Australian health practitioners frequently diagnose OME by the appearance of the tympanic membrane on otoscopy without the pneumatic attachment, which is less reliable than pneumatic otoscopy or tympanometry.

Treatment/Management

Treatment/management may be directed either at resolution of middle ear effusion and/or hearing impairment, or at remediation of presumed adverse effects of OME.

There are numerous management strategies for OME. These have been well summarised in systematic reviews (Reidpath, Glasziou, & Del Mar 1999; Anonymous 1992; Anonymous 1994). Basically, they include:

- *antibiotics* – long course (10-30 days) antibiotic and prophylactic antibiotic treatment may improve the rate of resolution of OME, but only in the short term (<30 days). Limited effectiveness and concern about the rise in resistant infections reduce the attractiveness of this approach, though anecdotally many children continue to receive repeated courses of antibiotics over a period of months leading up to ENT referral.
- *autoinflation* – pooled estimates in a recent systematic review suggest overall benefit in children older than 4 years, particularly use of nasal balloons. Combining data from three studies, those using nasal balloon were 3.5 (95% CI 2.03-6.14) times more likely to improve than those not using the balloon. However, the quality of studies included was poor and considerable heterogeneity was present (Reidpath, Glasziou, & Del Mar 1999).
- *tympanostomy tubes* – tympanostomy tubes are highly acceptable to parents and professionals and offer dramatic improvement in hearing for those with hearing impairment. However, pooled estimates from multiple good randomised controlled trials suggest that this benefit is relatively short lived, with mean hearing improvement in operated vs unoperated ears approximately 12 dB HL at 6 months and 6 dB at 12 months. Three randomised controlled trials of language outcomes have been reported to date, two of which recruited on the basis of persistence of OME, while the third selected only those with severe persistent OME associated with significant hearing loss.

Four RCTs of tubes in children with persistent OME have been reported:

- i. The Nijmegen study showed no benefit of tube insertion in children aged 2-4 years with persistent OME detected by tympanometric screening (Rach et al. 1991). Only a small proportion of children with persistent OME took part in the RCT, and hearing loss was not a criterion for entry.
- ii. Paradise et al conducted an RCT nested within the Pittsburgh prospective cohort study in which 6353 infants were enrolled at <2 months of age and monitored prospectively for middle ear effusion at least monthly throughout the first two years of life. 429 children with the most persistent OME were randomised to receive early tubes (after three months of continuous OME) or delayed tubes (after 9 months if still necessary). 402 of these were assessed developmentally at age 3 years, and were also compared with a random sample of 241 children representing a spectrum of middle ear effusion from none to persistent OME just short of meeting randomisation criteria. There was no significant difference, nor any trend towards a difference, between the early tubes group (n=206) and the delayed tubes group (n=196) on any of the outcome measures (the McCarthy General Cognitive Index or any of its three subscales, the Peabody Picture Vocabulary Test-Revised, the number of different words and mean length of utterance in morphemes, and the Percentage of Consonants Correct-Revised). Further, neither of the randomised groups differed significantly from the children with less severe OME (Paradise, Feldman, Campbell, Dollaghan, Colborn, Bernard, Rockette, Janosky, Pitcairn, Sabo, Kurs-Lasky & Smith 2001).
- iii. In Maw's study, 186 children aged 1-4 years with OME with persistent hearing impairment of 25-70 dB for at least 3 months were randomised to early surgery (within six weeks) or watchful waiting for nine months (Maw et al. 1999). At baseline, children in the control group scored 0.3 SDs lower than cases on the Hearing & Speech subscale of the Griffiths mental development scales, and this difference, though not significant, persisted at 9 months. At 9 months both verbal comprehension and expressive language skills in the watchful-waiting group were approximately 0.4 SDs lower on the Reynell scales than in the early surgery group. This was a significant difference between the groups (p=0.03 for comprehension and p=0.04 for expressive skills), but analyses were not controlled for baseline language skills. 18 months after randomisation, 85% of children in the watchful-waiting group had received surgery and groups did not differ significantly (Maw, Wilks, Harvey, & Golding 1999).

- iv. Rovers studied the effectiveness of tympanostomy tubes on language development in a multicentre RCT, nested within a well-ascertained prospective cohort study. Infants aged 1-2 years with OME persisting for at least 4 months after a hearing check at 9 months or age were randomised to tubes (n=93) or no tubes (n=94). Over the 12 month followup period, no differences were found in expressive or comprehensive language between the two groups in multivariate analysis (Rovers, Straatman, Ingels, van der Wilt, van den Proek & Zielhuis 2000).

Results of a further RCT is awaited from the UK MRC Institute of Hearing Research, concentrating on changes in parent-reported health-related quality of life and behaviour following tube insertion.

In uncontrolled trials parents have strongly supported tubes, reporting marked improvements in language, behaviour and well being and high levels of satisfaction with their decision to have tubes inserted (Rosenfeld, Bhaya, Bower, Brookhouser, Casselbrant, Chan, Cunningham, Derkay, Gray, Manning, Messner & Smith 2000).

- *steroids* – appear to have some effectiveness in a small number of RCTs. With little information about the harm-benefit ratio, the 1993 US Guideline recommended against their use for OME in children under 4 years.
- *short-term hearing aids* – limited information about this strategy suggests that hearing aids can be well tolerated in preschool children with OME, with parents reporting appreciable improvement in their hearing.
- *watchful waiting* – given the fluctuating nature of OME, some evidence that children experiencing more OME suffer more frequent (rather than longer) episodes, and the precipitous decline in OME after the age of 6 years, watchful waiting for many children may be the management of choice.

How might screening reduce the burden of suffering?

The rationale for screening is that, if OME causes lasting adverse developmental and behavioural outcomes which could be averted by active management, then screening might be worthwhile. However, OME does not in general lead to adverse outcomes, at least using standard measures of language and developmental outcomes. Further, treatment of OME does not lead to improvements in language and developmental outcomes. Thus, on a priori evidence, screening does not appear to be justified.

Regardless of long term outcomes, moderate hearing loss over a number of months presumably imposes an immediate burden of suffering. However, screening at specified age points will pick up only a small minority of these children because of the fluctuating nature of OME, and it is difficult to quantify this burden in young children.

6.3.7a.2 EVIDENCE

Tests

Pneumatic Otoscopy (New Zealand Health Technology Assessment 1998)

Pneumatic otoscopy is widely used in primary practice in the US and has high sensitivity and specificity in trained hands (approximately 90% sensitivity and 75% specificity). Although cheap, quick and not difficult to perform, it requires individual one-on-one training and validation against normal and abnormal ears. Because there is no external pass/fail criterion, it requires individual interpretation. Ideally therefore individuals would be regularly “calibrated” in their interpretation of findings for individual ears against an experienced user, tympanometry or myringotomy. It does not appear to be widely used in Australasia either by general practitioners or by paediatricians, though is used by ENT specialists. Otoscopy without the pneumatic attachment is widely used, but is insensitive and nonspecific for most cases of OME. Otoscopy therefore seems unlikely to be useful for screening for OME. Neither otoscopy nor pneumatic otoscopy are tests of hearing impairment.

Tympanometry (New Zealand Health Technology Assessment 1998)

Tympanometry has high sensitivity and specificity in the detection of OME, both in clinical settings against the absolute gold standard of myringotomy and in community settings against the relative gold standard of pneumatic otoscopy by trained examiners. The finding of a Type B tympanogram has a sensitivity between 78-90% and specificity of 63-94%. If Type B and C2 tympanograms are combined, sensitivity increases to close to 100% but specificity falls to approximately 75%. A Type A tympanogram has high specificity in ruling out OME.

Tympanometry is not a test for hearing impairment. Between 49% and 66% of children with a type B tympanogram have a hearing impairment ≥ 25 dB, compared to only 2-13% of children without a Type B tympanogram.

Tympanometry is quick and painless, generally acceptable to infants and young children and their parents, requires only brief cooperation, and is suitable for screening purposes. Small handheld microtympanometers are being increasingly used in primary practice. Although tympanometry itself is an objective test, results require interpretation and thus may be less accurate when used by inexperienced or untrained testers.

Audiometry

Audiometry is not a test for OME, but detects and quantifies the hearing impairment associated with it. It requires skill, time, quiet conditions, and cooperation on the part of the child. It is suitable for use in screening situations with children older than about 4 years. Screening “sweep” audiograms screens at defined thresholds (eg 30, 20, 20, 20 dB at 500, 1000, 2000 and 4000 Hz) are frequently used in preschool or school entry programs. Such low thresholds are likely to be inaccurate in the presence of the usual level of ambient noise encountered in schools.

Because of the greater level of skill and interpretation required, universal screening audiometry has not been adopted for younger children, although children of any age can be formally tested in ideal conditions.

Parent concern about hearing loss

Because of doubts about the effectiveness of universal preschool or school entry screening, it has been suggested that parent questionnaires or reported concerns might be a good alternative which could “screen in” those most at risk of adverse outcomes of OME for back up screening or formal testing. At least one state (Victoria) has already introduced this system, although without prior validation of the approach. Therefore this issue is covered in some depth in this review.

Parent concern about their child’s hearing is an important risk factor for sensorineural hearing impairment, though its sensitivity, positive predictive value and specificity appear low even in the presence of severe or profound persistent hearing impairment. Careful studies also show that parent concern is a poor predictor of the mild hearing impairment associated with OME. Parent report of signs/symptoms of AOM, OME and hearing impairment ≥ 20 dB was insensitive in a prospective study of infants aged 0-27 months, even with the reinforcement of three monthly formal audiometry (Anteunis et al. 1999). A widely-used standard questionnaire containing questions about language and health, including concern about hearing, was found to have sensitivity of 56% and specificity of 52% in 685 4-5 year old South Australian preschool children against concurrent or persistent hearing impairment by screening audiometry, the authors concluding that the questionnaire used was not valid in detecting conductive hearing impairment (Hammond et al. 1997). At ages 5-7 years, findings were similar in a large population-representative Victorian school entry cohort of 1563 children. Despite a wide range of hearing-related questions, individual questions had low sensitivity (4-50%) and were poorly predictive of hearing loss (PPV 4-26%). By failing children who responded positively to any of the 15 most sensitive questions, sensitivity in the detection of persistent hearing loss improved to 89%. However, specificity dropped to 24% (PPV= 4%) meaning that 73% of children would require testing to detect the 3% with persistent hearing loss (Heathershaw & Wake 2000). One English parent questionnaire, the Childhood Middle Ear Disease and Hearing Questionnaire, had reasonable sensitivity (mainly due to questions about previously identified middle ear disease) in the detection of those school entry children who ultimately went on to have tubes inserted. This may however be a somewhat circular argument (Hind, Atkins, Haggard, Brady, & Grinham 1999).

Even for children with chronic OME, including those undergoing tube insertion, parent perception of hearing loss is a poor predictor of hearing threshold or change in threshold (Stewart et al. 1999) (Rosenfeld, Goldsmith, & Madell 1998).

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✗
Accepted treatment	✗
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	?
Suitable test or examination	✓
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	✗

Programs

We conclude above that OME does not in general lead to adverse outcomes, at least using standard measures of language and developmental outcomes, and neither does treatment of OME does not lead to improvements in language and developmental outcomes. Thus, on a priori evidence, screening does not appear to be justified. Nonetheless, there is a substantial though flawed literature addressing screening for OME.

In 1998 the New Zealand Health Technology Authority reviewed screening programs for the detection of otitis media with effusion and conductive hearing loss in preschool and new entrant school children. No RCTs were identified that examined screening programs for OME and conductive hearing loss. Most studies have tried to ascertain accuracy of screening in detecting children with OME, with very few addressing health or developmental outcomes. Essentially, no studies demonstrated health benefits arising from the programs. Readers are referred to the NZHTA website for a more detailed review of the screening programs summarised in the following paragraphs (<http://nzhta.chmeds.ac.nz>).

Numerous different algorithms using tympanometry and/or audiometry have been developed, each with its own schedule of tests and referral criteria. Lous carried out an intensive research program involving 500 children from age 3 years screened with tympanometry on four occasions over six months. Children with abnormal tympanograms were referred only after an additional abnormal finding on the second (Type B) or third (C2/B) screen. Sensitivity, specificity and positive predictive value of different algorithm varies greatly, resulting in a highly variable referral load and resource use (Lous 1987 as cited in New Zealand Health Technology Assessment 1998). Of five screening algorithms involving tympanometry on one or more occasions, the algorithm from the Danish study by Lous had the best results in terms of reasonable sensitivity (80%), excellent specificity (95%) and reasonable PPV (50%) in terms of detecting persistent OME (hearing impairment was not an outcome).

Three cohort studies used diagnostic confirmation by an ENT specialist as the outcome measure. These studies demonstrated how big an effect different screening algorithms may have on referral rates. In a Canadian one-stage program (O'Mara, Isaacs, & Chambers as cited in New Zealand Health Technology Assessment 1998) screening 1840 3 and 4 year old children attending day care with audiometry, 2% (35) of children failed the initial screen at 35 and 40 dB. Of these, 15 eventually saw an ENT surgeon, 11 had their hearing formally tested to confirm hearing impairment in eight, of whom only five (0.3%) were new cases with OME. In a Swedish one-stage program (Augustsson, Nilson, & Engstrand 1990 as cited in New Zealand Health Technology Assessment 1998) screening 2330 4 year old children with audiometry, 8% (187) of children failed the initial screen at 25 and 40 dB. At ENT referral, 40% had OME, 4% had sensorineural hearing loss, 26% had problems that were already known about, 28% did not have a problem, and 3% were not seen. 31 had persistent OME of whom 27 (1.5%) of the total population were treated. In the Danish study described above, 46 of 500 (9.2%) 3 year old children were identified with persistent OME using intensive screening tympanometry (hearing impairment was not an outcome).

Two cohort studies were identified examining health outcomes in potentially sustainable screening programs for OME. Neither demonstrated a health outcome benefit, but both were felt to have design flaws precluding firm conclusions. In the Nijmegen study, (Zielhuis, Rach & van den Broek 1989 as cited in New Zealand Health Technology Assessment 1998) children were tested with tympanometry at home at three monthly

intervals from age 2-4 years. Those with bilateral Type B tympanograms on two successive occasions were referred for ENT opinion via their GPs, and if clinically indicated entered a study in which they were randomised to tubes or no tubes. There were no significant differences in language test outcomes between the two groups. However, this study was compromised by the extremely high non-compliance rate in accepting referral and management, illustrating some of the problems faced when referring asymptomatic children detected through screening. The other study was felt to contain insufficient information to draw conclusions, though did not suggest any benefit from screening at school entry (Feldman, Milner, Sackett & Gilbert 1980).

A New Zealand study reported significant time delays through the referral chain for children failing the 5 year old screen (Claridge, Ford, Shluter, Wild & Macleod 1995 as cited in New Zealand Health Technology Assessment 1998).

No studies were identified that examined the effectiveness of opportunistic screening for OME and conductive hearing loss in the primary care setting.

Excluding the research studies described earlier in which children were “screened” extremely frequently for ascertainment purposes within a larger research project, no formal studies of screening programs for conductive hearing loss in infants or children less than two years were identified in the literature. However, detection of conductive hearing loss is a by-product of distraction testing for sensorineural hearing loss at 7-9 months of age. In the Victorian Infant Hearing Screening Program 5262 (8%) of the 65,624 infants born in 1993 were referred for failure of two distraction tests a month apart. Of the 3299 for whom data were available, 684 (20.7%) were confirmed with conductive hearing loss.

Cost effectiveness

There is no evidence that screening for OME is effective. Therefore consideration of cost-effectiveness is redundant.

Generalisability of evidence

There is high quality evidence showing that the presence of OME can be reliably detected through screening programs in children from at least 2 years of age, and that hearing impairment can be reliably detected through screening programs in children from approximately 4 years of age. However, the quality of studies examining effectiveness of screening programs is much lower, partly because of variability between studies but principally because the main outcomes of interest have not been examined (ie health and developmental gain). Given what is now known about outcomes of OME and of its treatment, even if high quality screening programs were introduced and evaluated they would seem unlikely to lead to substantial population health gain.

The prevalence of OME is similar across Western countries. Evidence about outcomes of OME and of treatment of OME appears to be generalisable, except possibly for particular populations at higher risk of long term suppurative middle ear disease. Conversely, studies of screening programs are very difficult to compare (and therefore generalise) because of the great variety of screening algorithms used in reported programs.

Quality of evidence

Two controlled screening studies examining the effectiveness of screening programs for otitis media with effusion were located. One compared the hearing of a school entry cohort screened on two occasions with that of a non-screened cohort 6-12 months later, but contained insufficient information to judge what had happened as a consequence of screening in the screened cohort so is not graded (Feldman et al 1980). The quality of the Nijmegen study is graded below.

Level of evidence

No controlled trials of screening programs for OME have been conducted.

Quality

No controlled trials available to grade for quality

Statistical precision

Not applicable.

Size of effect

Not applicable.

6.3.7a.3 CONCLUSIONS

Good evidence to recommend against screening

Comments:

Strong evidence indicates that long term developmental effects of otitis media with effusion detected through screening are minimal. Otitis media with effusion is a fluctuating condition, so that one off screening makes no sense. If screening were to occur it should be repeated many times for each child. However, treatment of children detected through repeated screening has been shown to provide no long term developmental or academic benefit.

However, no research program has yet adequately researched this issue for the small number of children with significant persistent conductive hearing impairment, as opposed to those with persistent otitis media with effusion regardless of hearing status.

6.3.7a.4 RECOMMENDATIONS

Universal screening programs to detect otitis media with effusion are not recommended.

6.3.7a.5 FURTHER RESEARCH

Long term effects of persistent conductive hearing loss associated with moderate or greater hearing loss

Long term effects of persistent conductive hearing loss on other aspects of language such as auditory processing

6.3.7a.6 SEARCH TERMS

Hearing

Auditory perceptual disorders

Hearing disorders

Hearing tests

Hearing loss

Auditory perception

Audiology

Deafness

6.3.7a.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Katrina Williams – Reference Group

David Hall – Reference Group

Jan Pollard – Chief Audiologist, Royal Children’s Hospital

Peter Morris – NHMRC Public Health Research Fellow, Menzies School of Health Research

6.3.7a.8 REFERENCES

- Anonymous. 1992, "The treatment of persistent glue ear in children", *Effective Health Care Bulletin*, vol. 4.
- Anonymous. 1994, "Managing otitis media with effusion in young children. American Academy of Pediatrics The Otitis Media Guideline Panel", *Pediatrics*, vol. 94, no. 5, pp. 766-773.
- Anteunis, L. J., Engel, J. A., Hendriks, J. J., & Manni, J. J. 1999, "A longitudinal study of the validity of parental reporting in the detection of otitis media and related hearing impairment in infancy", *Audiology*, vol. 38, no. 2, pp. 75-82.
- Arcia, E., Roberts, J. E. 1993, "Otitis media in early childhood and its association with sustained attention in structured situations", *Journal of Development and Behavioral Pediatrics*, vol. 14, pp. 181-183.
- Feagans, L., Sanyal, M., Henderson, F., Collier, A., Appelbaum, M. 1987, "Relationship of middle ear disease in early childhood to later narrative and attention skills", *J Pediatr Psychol*, vol. 12, pp. 581-594.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. 1999, "Parent-reported language and communication skills at one and two years of age in relation to otitis media in the first two years of life", *Pediatrics*, vol. 104, no. 4, pp. e52.
- Feldman, W., Milner, R. A., Sackett, B., Gilbert, S. 1980, "Effects of preschool screening for vision and hearing on prevalence of vision and hearing problems 6-12 months later", *Lancet*, vol. 2, no. 8202, pp. 1014-1016.
- Hammond, P. D., Gold, M. S., Wigg, N. R., & Volkmer, R. E. 1997, "Preschool hearing screening: evaluation of a parental questionnaire", *Journal of Paediatrics and Child Health*, vol. 33, no. 6, pp. 528-530.
- Heathershaw, K. & Wake, M. 2000, "Effectiveness of a parent questionnaire in the detection of mild to moderate hearing loss in school entry children.", *Pediatric Research*, vol. 47, no. Part 2, pp. 1157A (196).
- Hind, S., Atkins, R. L., Haggard, M. J., Brady, D., & Grinham, G. 1999, "Alternatives in screening at school entry: comparison of the childhood middle ear disease and hearing questionnaire (CMEDHQ) and the pure tone sweep test", *British Journal of Audiology*, vol. 33, pp. 403-414.
- Lous, J. 1993, "Silent reading and secretory otitis media in school children", *International Journal of Pediatric Otorhinolaryngology*, vol. 25, pp. 25-38.
- Lous, J., Fiellau-Nikolajsen, M., Jeppesen, A. L. 1988, "Secretory otitis media and language development: a six-year follow-up study with case-control", *International Journal of Pediatric Otorhinolaryngology*, vol. 15, pp. 185-203.
- Maw, R., Wilks, J., Harvey, I., & Goldring, J. 1999, "Early surgery compared with watchful waiting for glue ear and effect on language development in preschool children: a randomised trial", *Lancet*, vol. 353, pp. 960-963.
- New Zealand Health Technology Assessment 1998, *Screening programmes for the detection of otitis media with effusion and conductive hearing loss in pre-school and new entrant school children: a critical appraisal of the literature*, New Zealand Health Technology Assessment, Christchurch, 3.
- Paradise, J. L., Feldman, H. M., Campbell, T. F., Dollaghan, C. A., Colborn, D. K., Bernard, B. S., Rockette, H. E., Janosky, J. E., Pitcairn, D. L., Sabo, D. L., Kurs-Lasky, M., Smith, C. G. 2001, "Effect of early or delayed insertion of tympanostomy tubes for persistent otitis media on developmental outcomes at the age of three years", *New England Journal of Medicine*, vol. 344, pp. 1179-1187.
- Paradise, J. L., Feldman, H. M., Colborn, D. K., Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Janosky, J. E., Kurs-Lasky, M., Bernard, B. S., & Smith, C. G. 1999, "Parental stress and parent-rated child behavior in relation to otitis media in the first three years of life", *Pediatrics*, vol. 104, no. 6, pp. 1264-1273.
- Paradise, J. L., Rockette, H. E., Colborn, D. K., Bernard, B. S., Smith, C. G., Kurs-Lasky, M., & Janosky, J. E. 1997, "Otitis media in 2253 Pittsburgh-area infants: prevalence and risk factors during the first two years of life", *Pediatrics*, vol. 99, no. 3, pp. 318-333.
- Peters, S. A., Grievink, E. H., van Bon, W. H., Schilder, A. G. 1994, "The effects of early bilateral otitis media with effusion on educational attainment: a prospective cohort study", *Journal of Learning Disabilities*, vol. 27, pp. 111-121.
- Rach, G. H., Zielhuis, G. A., van Baarle, P. W., & van den, B. P. 1991, "The effect of treatment with ventilating tubes on language development in preschool children with otitis media with effusion", *Clinical Otolaryngology*, vol. 16, no. 2, pp. 128-132.

- Reidpath, D. D., Glasziou, P. P., & Del Mar, C. 1999, "Systematic review of autoinflation for treatment of glue ear in children", *BMJ*, vol. 318, no. 7192, p. 1177.
- Roberts, J. E., Burchinal, M. R., Campbell, F. 1994, "Otitis media in early childhood and patterns of intellectual development and later academic performance", *Journal of Pediatric Psychology*, vol. 19, pp. 347-367.
- Roberts, J. E., Burchinal, M. R., Clarke-Klein, S. M. 1995, "Otitis media in early childhood and cognitive, academic and behavior outcomes at 12 years of age", *Journal of Pediatric Psychology*, vol. 20, pp. 645-660.
- Roberts, J. E., Burchinal, M. R., Collier, A. M., Ramey, C. T., Koch, M. A., Henderson, F. W. 1989, "Otitis media in early childhood and cognitive, academic, and classroom performance of the school-aged child", *Pediatrics*, vol. 83, pp. 477-485.
- Roberts, J. E., Burchinal, M. R., Davis, B. P., Collier, A. M., Henderson, F. W. 1991, "Otitis media in early childhood and later language", *Journal of Speech and Hearing Research*, vol. 34, pp. 1158-1168.
- Roberts, J. E., Burchinal, M. R., Medley, L. P., Zeisel, S. A., Mundy, M., Roush, J., Hooper, S., Bryant, D., Henderson, F. W. 1995, "Otitis media, hearing sensitivity, and maternal responsiveness in relation to language during infancy", *Journal of Pediatrics*, vol. 126, pp. 481-489.
- Roberts, J. E., Sanyal, M. A., Burchinal, M. R., Collier, A. M., Ramey, C. T., Henderson, F. W. 1986, "Otitis media in early childhood and its relationship to later verbal and academic performance", *Pediatrics*, vol. 78, pp. 423-430.
- Rosenfeld, R. M., Bhaya, M. H., Bower, C. M., Brookhouser, P. E., Casselbrant, M. L., Chan, K. H., Cunningham, M. J., Derkay, C. S., Gray, S. D., Manning, S. C., Messner, A. H., Smith, R. J. 2000, "Impact of tympanostomy tubes on child quality of life", *Archives of Otolaryngology Head and Neck Surgery*, vol. 126, pp. 585-592.
- Rosenfeld, R. M., Goldsmith, A. J., & Madell, J. R. 1998, "How accurate is parent rating of hearing for children with otitis media?", *Archives of Otolaryngology Head and Neck Surgery*, vol. 124, no. 9, pp. 989-992.
- Rovers, M. M., Straatman, H., Ingels, K., van der Wilt, G. J., van den Proek P., & Zielhuis, G. A. 2000, "The effect of ventilation tubes on language development in infants with otitis media with effusion: a randomized trial", *Pediatrics*, vol. 106, no. 3, pp. e42.
- Skull, S. A., Morris, P. S., Yonovitz, A., Attewell, R. G., Krause, V., Leach, A. J., Anand, A., Scott, J., Reid, S., & Roberts, L. A. 1999, "Middle ear effusion: rate and risk factors in Australian children attending day care", *Epidemiol. Infect.*, vol. 123, no. 1, pp. 57-64.
- Stewart, M. G., Ohlms, L. A., Friedman, E. M., Sulek, M., Duncan, N. O., Fernandez, A. D., & Bautista, M. H. 1999, "Is parental perception an accurate predictor of childhood hearing loss? A prospective study", *Otolaryngology Head and Neck Surgery*, vol. 120, no. 3, pp. 340-344.

6.3.7b. PERMANENT CHILDHOOD HEARING IMPAIRMENT

6.3.7b.1. BACKGROUND

With the advent of new technologies for detection and management and the development of strong support for universal neonatal hearing screening, early detection of permanent childhood hearing impairment has become perhaps the single most pressing childhood screening/surveillance issue under debate in Australia. It has considerable economic implications.

Clinical picture

Hearing impairment is often described as the pure tone average hearing threshold in the better ear across either three (0.5, 1 and 2 kHz) or four (0.5, 1, 2 and 4 Hz) frequencies and expressed as dB HL (decibels hearing loss). While less than 15dB HL has been considered an optimal level of hearing for “normal” language development in children, (Northern JL 1991) the lower limits above which detriment to speech and language occur is not known. 35-40 dB HL in the better ear is usually considered to be educationally significant, with 40 dB often cited as the criterion by which prevalence of hearing impairment is assessed (Birtles et al. 1998).

Permanent childhood hearing impairment (PCHI) does not represent a single condition or defect. It may be described in many dimensions – for example, by severity, by the site of abnormality, by presence or absence of risk factors, by the timing of commencement, or by aetiology. Many of these classifications overlap. Patterns of hearing impairment may also vary, for example some children having predominantly high-frequency hearing loss and others high and low frequency impairments with central frequencies relatively preserved.

Severity - although cutpoints vary and are arbitrary, classifications such as the following are widely used: mild, 25-39 dB HL; moderate, 40-69 dB HL; severe, 70-94 dB HL; and profound, ≥ 95 dB HL (Wessex Universal Neonatal Hearing Screening Trial Group 1998).

Site of abnormality – PCHI may be due to outer or middle ear abnormality, cochlear abnormality (by far the most common), or it may be retrocochlear (auditory nerve through to auditory cortex). Neural (usually retrocochlear, though may be intracochlear) abnormalities are relatively uncommon though the exact prevalence is not known. Davis believes that this comprises less than 1% of all PCHI, though is likely to be higher in neonatal intensive care unit (NICU) infants (Davis et al. 1997). However, neural losses were present in 11% of Rance’s cohort of 109 children with moderate or greater PCHI identified through an early identification program based upon risk factors (Rance et al. 1999). PCHI may also be conductive, caused by congenital malformations of the external or middle ear. Impairments may be due to structural or functional abnormalities, an example of the latter being connexin deficiency.

Risk factors – There are a number of risk factors for PCHI. These include admission to NICU, birthweight <1.5kg, perinatal infection, craniofacial abnormalities, birth asphyxia, chromosomal abnormalities, very high levels of jaundice, family history of PCHI, and certain drugs such as aminoglycosides. Approximately 40% of children with PCHI have been admitted to special care nurseries and NICUs (Wessex Universal Neonatal Hearing Screening Trial Group 1998) (Watkin 1996) Approximately 60% of children with PCHI in the Trent region of England born 1985-1993 had one or more of just three risk factors (NICU/special care nursery for 48 hours or more, family history of PCHI, or cranio-facial abnormality (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). A large number of rare syndromes include PCHI as a feature.

Timing of commencement - PCHI may commence prenatally, perinatally, or postnatally (the latter two accounting for 6.7% and 6.1% respectively in the Trent Region 1985-1993 (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997) from a range of aetiologies. Regardless of when it commences, more than 6% of PCHI may be progressive, with onset of deterioration in one study after 4 years in more than 70% of children with progressive loss (Berettini et al. 1999). PCHI is often referred to as “congenital” (ie that which may be detected with neonatal screening) or “progressive or acquired”, which may be due to a specific insult such as meningitis or head trauma and which could not be detected until later infancy or early childhood.

Aetiology - Forty percent of PCHI is due to environmental factors (eg premature birth, meningitis). The remaining 60% of cases are now considered genetic, the largest single category of PCHI being the isolated genetic mutation leading to non-syndromic deafness. While most of these children have congenital PCHI, a substantial minority have progressive loss, particularly the 15% who have autosomal dominant deafness (Marazita et al. 1993; Morton NE 1991).

Natural history

Bilateral hearing losses of moderate or greater degree are known to have educationally-significant impacts on the infant’s speech, language and general development. Once detected, children with PCHI almost invariably

receive some form of oral or sign language habilitation and hearing augmentation, so that the natural history of PCHI is now only rarely played out in full. Even so, the consequences of severe and profound PCHI are enormous, with children on average suffering markedly poorer language, speech and academic outcomes than their peers. Social isolation combined with economic and educational disadvantage is a common outcome (Steel 2000). Despite the severe disability imposed by deafness for the majority of children, some individuals with PCHI achieve highly. Many in the Deaf Community do not feel that their quality of life is diminished by their hearing impairment, with strong and supportive social milieu and excellent educational outcomes from programs that use sign rather than oral languages.

Intuitively, one might expect to see two outcome gradients: relationships between severity of hearing impairment (ie the more severe the impairment, the poorer the language and educational outcomes) and between age of diagnosis and outcome (ie the earlier the diagnosis and remediation, the better the outcomes). There is a dearth of research examining either of these gradients. To date, the best evidence supports a relationship between outcome and age of diagnosis (see below) but not severity of hearing impairment. However, some children do poorly and others do well despite apparently identical levels of hearing loss and management. The reasons for this are poorly understood, though may in part reflect some children's underlying predispositions to poor language development.

The ideas of "sensitive" or "critical" periods versus the extent of neural plasticity are central to defining the need for early detection, and have been recently reviewed by Davis (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). A critical period may be defined as "...a period during which the action of a specific stimulus is required for normal development of the system, and during which the organism is maximally vulnerable to environmental manipulation" (Eggermont 1986 as cited in Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). There is no doubt that information from the linguistic environment is an essential element for normal language acquisition to occur, and that our ability to acquire a first language diminishes with age (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). The best current population-based evidence points to better outcomes for those diagnosed and habilitated in the first six months of life (Yoshinaga-Itano et al. 1998). However, there is also considerable evidence of plasticity, and some children treated much later (for instance, with cochlear implants) nonetheless acquire early language in an apparently normal though delayed sequence, even though their language outcomes may never reach those of their hearing peers. Whether there is an age (or any other determinant) beyond which language acquisition is universally *deviant*, rather than *delayed*, is currently under study. Davis concludes that with respect to PCHI "the evidence as to critical/sensitive periods remains far from consistent or conclusive" but "points to early sensitive periods for aspects of language acquisition, that suggest earlier intervention to be better than later intervention, other factors being equal." Furthermore, there is some evidence of "substantial and long-term detrimental effects of the lack of sensory input on neuronal pathways that suggest that the earlier the lack of input is overcome, the less detrimental the effects will be" (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997).

The consequences of mild or unilateral PCHI are not clear, since most reported outcome studies have not been conducted on population-based samples. Many children with mild PCHI may not be diagnosed and therefore not included in clinical samples, so that cohort studies of pre-identified individuals may be biased by including the individuals who come to attention (who may be those most likely to be experiencing difficulties). The only representative school-based population outcome study reported unilateral, mild bilateral, or high frequency sensorineural hearing loss in 66 (5.4%) of 1218 children in grades 3, 6 and 9. 37% had repeated a grade, compared to about 5% of the total school population. Grade 3 children, but not Grade 6 or 9 children, had significantly lower scores on almost all scales of educational performance as measured by the CTBS/4 (reading vocabulary, reading total, language mechanics, basic battery, word analysis, spelling and science) than matched controls. Student-reported functional status scores (COOP) in Grades 3 and 6 were poorer on almost all subtests, reaching significance for school support, stress, and self-esteem scales for Grade 9 children. Teacher reports also tended to be poorer but the response rate was unacceptably low (<50%) and thus open to bias. No data were presented as to cause of hearing impairment or the proportion of children in which it was already diagnosed (Bess, Dodd-Murphy, & Parker 1998).

Prevalence

The exact prevalence of permanent childhood hearing impairment is unknown but reported prevalence varies from 1.2-5.7 per 1000 live births. The prevalence of moderate or greater PCHI appears to be just under 2/1000, with variability between populations probably representing differences in definition (eg whether mild losses are included) and ascertainment than major differences in prevalence. By age six years 1.8/1000 children of Victorian children born in 1993 were fitted with hearing aids for congenital hearing impairment, classified as mild in 27%, moderate in 30%, severe in 18%, profound in 11%, and unclassified for the remainder. In the same period 0.15/1000 were fitted by age six years with hearing aids due to hearing impairment classified as acquired or progressive (personal communication, Victorian Infant Hearing Screening Program). It should be noted that newborn screening will not in general detect children with

acquired or progressive losses, nor the relatively high proportion of infants with mild hearing impairment since the target condition for most screening programs is hearing impairment of ≥ 35 dB HL.

Genetics relevant to screening issues

For the majority of children with hearing impairment the underlying cause is genetic. Over 60 known chromosomal loci are associated with non-syndromic deafness, for which over 15 genes have already been identified. Presumably, many more remain to be identified. A wide range of molecules is encoded by these genes (channel components, motor molecules, extracellular matrix components and transcription factors) leading to the same final outcome of hearing impairment. An even greater number of single gene defects is associated with the many syndromes associated with deafness (Steel 2000).

Mutations in the gene coding for connexin 26 (Cx26) account for about half of the autosomal recessive non-syndromic deafness corresponding to 10-20% of all prelingual hearing impairment (Estivill et al. 1998). Deafness is usually severe or profound, most marked in the higher frequencies, and is not associated with phenotypic abnormalities in other body systems. More than 50 Cx26 mutations have been identified since 1995. One mutation, a deletion of a guanine residue at nucleotide 35 (35delG) resulting in a premature stop codon, has been found at high prevalence in most ethnic groups studied so far (Estivill & Gasparini 2000). In the Australian population the carrier frequency has been estimated to be approximately 1:100. Early studies suggested that all children with recessive connexin deficiency had severe-profound bilateral hearing impairment maximal in the upper frequencies. However, recent careful phenotype-genotype studies have revealed that hearing impairment in some individuals is mild and severity may vary between individuals sharing apparently the same genetic defect within families. Further, some hearing-impaired individuals have a high-frequency loss audiogram characteristic of connexin deficiency but in the presence of only a single connexin mutation, suggesting either that a single mutation can cause deafness or that interaction with other genetic defects may be occurring (Wilcox et al. 2000).

Of the 60% of PCHI now considered genetic, approximately 25% of mutations are in the connexin 26 gene. A molecular screen for mutated deafness genes therefore poses significant challenges due to the many genes involved, but work is already commencing on microarray DNA technologies that may overcome these problems within the next five years. Such tools hold the potential to significantly alter the current balance in arguments for and against universal neonatal hearing screening (UNHS), particularly for the large majority of children currently classified as "low risk" by conventional criteria.

It seems highly likely that genetics and the environment interact even when there seems to be a clear environmental determinant. For example, individuals may be genetically predisposed to hearing loss induced by noise, drugs, or infection (Steel 2000).

Diagnosis

Diagnosis is by determining hearing threshold in each ear at a range of frequencies using definitive audiologic methods. Visual reinforcement audiometry can be used to determine ear specific thresholds in infants over the age of 6 months, and other age-appropriate methods including play audiometry may be used with toddlers and young children. Methods for determining threshold for infants under the age of 6 months, or in older infants and children with intellectual disabilities are based upon auditory evoked potential tests, primarily the auditory brainstem response (ABR). True sensitivity and specificity values for this test are more than 97% (Hyde et al, 1991).

In the past, clinical and other specialised assessments have determined the exact cause in fewer than 50% of cases. Advances in genetic testing are likely to soon substantially raise this figure.

Treatment/Management

Surgical intervention may be of benefit for some forms of permanent conductive PCHI, for instance in children with atresia. Effective aids to hearing are available, although they do not restore normal hearing. Hearing aids can be fitted in young infants, while cochlear implants are now implanted in infants as young as 1 year. Children with profound PCHI typically derive little benefit from conventional hearing aids and cochlear implantation may be the preferred device for providing an aid to hearing. The other cornerstone of management is habilitation of language development, using aural-oral (spoken language) or visual-manual (sign language) modes.

Does early detection make a difference to outcomes for children with PCHI? There are no randomised controlled trials of the benefits of early identification. Four studies provide some weak (Level III) evidence that earlier identification is associated with better outcomes but all are based on observational data. Such studies can result in biased measures of effectiveness, since other confounding factors may be associated with earlier intervention (Medicare Services Advisory Committee 1999).

The highest quality study examining benefit of early diagnosis is a retrospective cohort study. Yoshinaga-Itano et al found that children with hearing impairment detected in the first 6 months of life demonstrated

significantly better expressive and receptive language quotients (LQ) than children identified after 6 months on parent-reported measures taken at 13-36 months of age. Similar outcomes were obtained across all ranges of hearing impairment from 27-110 dB, the only influencing variable being age at diagnosis (<6 months vs >6 months) (Yoshinaga-Itano, Sedey, Coulter, & Mehl 1998). Children received early intervention services within an average of two months after identification. Similar benefits were seen for children without cognitive impairment (mean total LQ 91 vs 70, early vs later diagnosis) and for the whole group (mean total LQ 79 vs 64). Between the ages of 6 and 34 months, incremental delay in detection appeared to result in little additional deficit. Early-diagnosed children had better cognition, less severe hearing impairment, and were from families with higher maternal education; although controlled for in analysis, these variables remain a possible source of bias in this study. Markides reported similar findings in a population-based study in which the speech of children fitted with hearing aids in the first six months of life was more intelligible at age 8-12 years than of children fitted later. Again, incremental delay in detection appeared to result in little additional deficit (Markides 1986). Ramkawan examined spoken language skills in 16 children aged 27-79 months, finding better outcomes with earlier intervention, (Ramkalawan & Davis 1992) as did Robinshaw in a small study of five children fitted with hearing aids by six months (Robinshaw 1995). The Yoshinaga-Itano and the Markides studies did not report methods of detection, methods of acceptance into the cohort or outcomes for children not included in the cohort, ie some issues of particular importance to population-based screening programs were not addressed. Further studies are urgently needed, particularly to examine whether these findings continue to hold for those diagnosed through screening efforts who otherwise would have been diagnosed later.

Questions as to the effectiveness of different management approaches have received little study. Despite this, families and professionals typically hold strong preferences as to how the hearing impairment of individual children should be managed, which may prevent comparative study of the effectiveness of different options.

Prevention

Currently, there is no way of preventing the majority of cases of PCHI. Some acquired causes of deafness are preventable, such as meningitis, head injury, and rubella. Improved neonatal intensive care may have already reduced PCHI due to prematurity (Meyer et al. 1999). Genetic screening in utero for some "deafness" genes is technically not difficult, but the ethics and desirability of offering termination to parents of potentially deaf fetuses are far from clear.

How might screening reduce the burden of suffering?

If screening leads to early and accurate identification of hearing impairment, in turn leading to early intervention which in turn has positive effects, then it should also lead to improved language and communication. Theoretically it could also lead to improvement on a range of related outcomes such as quality of life, psychosocial adjustment, academic achievement and employment prospects (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). However, it is not yet known what proportion of the burden of disability and suffering due to PCHI can be averted by very early diagnosis.

Potential harms include cost and anxiety generated by false positive screens, identification and possibly unnecessary management of conductive hearing impairment, and the stresses of early adjustment to an impairment that does not yet appear to parents to be causing a disability or handicap for their child.

6.3.7b.2. EVIDENCE

Tests

Two technologies are widely available for newborn hearing screening: evoked otoacoustic emission (OAE) and automated auditory brainstem response (ABR) tests. Each technology has its advantages and disadvantages. The only study to date evaluating their relative benefits and costs in the context of hospital-based UNHS indicated equivalent test performance for OAEs and ABRs in terms of true sensitivity and specificity for detecting mild or greater hearing losses in newborns (Norton 1999). OAEs and automated ABR screening tests provide "pass" or "fail" outcomes, and definitive testing is needed to define hearing thresholds and therefore severity of hearing impairment.

Otoacoustic emissions (OAE)

OAEs are responses arising from the cochlear outer hair cells, and can be measured in the external ear canal in response to transient (click or tone-burst) stimuli (Kemp 1978) (Norton 1999). OAEs are essentially the "sounds" created by the organ of hearing mechanics as it transduces sound. OAEs are generally absent when pure tone hearing loss is greater than 35 dB HL so that binary decision making can be employed (OAE present = pass, OAE absent = refer or retest). OAEs have been used extensively in hospital-based programs and at least one community based UNHS program. It is a quick, non-invasive test that can be completely

automated and is highly sensitive to loss of outer hair cell integrity, the part of the organ of hearing most susceptible to disease or lesion causing congenital hearing loss. However, it also has high false positive rates (up to 38% of babies in the first 24 hours of life, and no lower than 5% even in experienced hands after this age). In addition, OAE tests by definition miss inner hair cell or neural losses.

One systematic review of the sensitivity and specificity of OAE testing was identified (Medicare Services Advisory Committee 1999). It concluded that the quality of studies was not good, principally because of failure to describe sample selection or failure to evaluate OAE independently of the reference test "leaving open the possibility of bias". No adverse outcomes were reported in any of the trials included in the review. The studies in the review compared OAE with either ABR (eight studies) or other forms of audiological testing (five studies). For studies using ABR as the comparator, OAE sensitivity ranged from 50-100%, specificity from 52-95%, false positive rates from 0-50% and false negative rates from 5-48%. For studies using other comparators OAE sensitivity ranged from 39-94%, specificity from 68-94%, false positive rates from 6-45% and false negative rates from 6-32%. Time estimates for OAE in newborns ranged from 3-17.5 minutes. Pass rates improved over the first 48 hours of life, but in one study of children tested at 12 weeks fail rates had risen to 9.6% because of the rising prevalence of otitis media with effusion.

Automated auditory brainstem responses (ABR)

ABRs are the far-field recorded brain evoked potentials generated by the auditory nerve and the auditory pathway within the brainstem in response to brief sounds (clicks or tonebursts) and have been employed in the assessment of hearing sensitivity of infants since 1974. More recently, automated ABR has been developed specifically for UNHS in which computerised automated detection algorithms judging presence or absence of a response obviate the need for expert observers. In general, false positive rates are lower than for OAEs, and may be less than 0.3% for a two stage protocol (Mason & Herrmann 1998) with high sensitivity even in the first 24 hours of life. The disadvantages are that it requires the placement of three EEG leads on the infant's scalp, so that it can be perceived as somewhat more "invasive" or disruptive than an OAE test requiring an earprobe alone, and takes slightly longer (12-30 minutes (Medicare Services Advisory Committee 1999)) to complete a test.

No systematic review of the sensitivity and specificity of automated ABR screening was identified. Davis reviews selected studies and reports specificities of 90-99.7% (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). Sensitivity is harder to calculate, in part because of lack of gold standard testing and the possibility that "false negatives" may in fact be acquired or progressive losses. Figures therefore range from 52% to 100%, with many studies reporting approximately 90% sensitivity. When a diagnostic ABR test was used as the gold standard, Herrman reported specificity of 96% and sensitivity of 98% from 1187 neonates across five studies (Herrman, Thronton, & Joseph 1995).

SSEP

Auditory steady state evoked potentials (SSEP) are obtained by using amplitude and frequency modulated tones, rather than a click, to evoke a response from the auditory nerve and brainstem auditory system. They are related to ABRs, in that the neural generators of the response to high frequency (>60 Hz) modulation appear to be the same. The advantage of SSEPs over click-ABR is that hearing thresholds can be estimated using tonal frequencies in the audiometric range of 250-8000 Hz. Rickards et al (Rickards 1994) developed normative threshold data for newborn infants, and Rance et al (Rance 1995, Rance 1998) have further demonstrated the efficacy of using SSEPs to estimate thresholds in infants, children and adults who have hearing loss. Screening applications of SSEP have been piloted (Cone-Wesson et al 1997), but need large-scale trials to show that the technology developed for diagnostic purposes could be applicable in the screening context.

Genetic screening

Studies have identified the population prevalence of several genes known to cause non-syndromic deafness. The most common of these is the 35delG mutation leading to deficiency in connexin 26, believed to be responsible for 10-20% of non-syndromic prelingual deafness. This mutation is readily and inexpensively identifiable using blood or buccal samples and would be suitable for an analysis of blood obtained for neonatal heelprick testing, leading to early hopes that screening programs might soon be feasible. However, the severity of deafness associated with identical recessive genetic defects varies and some hearing-impaired individuals have only a single connexin mutation (*see above*). Therefore until the implications of the connexin-26 mutation/normal phenotype are better understood, counselling of affected families is difficult. At present, these uncertainties preclude use of connexin 26 testing for population screening outside the research context. Its implications for secondary testing of children diagnosed with hearing impairment are currently being researched (Wilcox, Saunders, Osborn, Arnold, Wunderlich, Kelly, Collins, Wilcox, Gardner, Kamarinos, Cone-Wesson, Williamson, & Dahl 2000). The development of microarray DNA technologies may soon make feasible screening of the single newborn blood spot for multiple genes associated with PCHI.

A potential attraction of genetic screening is that virtually all Australian neonates provide the heelprick blood test that would be required for genetic testing. However, this approach has not yet been subjected to trial.

Distraction testing

Behavioural screening tests for hearing impairment have been in place in some countries (eg Australia, New Zealand and the UK) for many years. Essentially, an infant aged approximately 8 months is required to localise frequency-specific stimuli (eg low and high frequency sounds) presented at quiet levels (approximately 35-40 dB).

Davis et al reviewed studies of the performance of the 7-9 month distraction test (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). They conclude that coverage can be high (80-95%), but sensitivities vary widely (18-88%), tending to be better for severe and profound losses. More recent studies with larger samples and better population ascertainment tend towards the lower range of sensitivity. Failure rate is 5-10%, almost all of which are false positives (though many of these will have conductive hearing loss) which has major service implications. With UNHS, incremental yield falls to very low levels. Median age of identification via distraction testing varies from 12-20 months and is dependent on severity of hearing loss.

Distraction testing plays a major part in the Victorian Infant Hearing Screening Program, which is one of the few programs to have carried out a population-based assessment of its effectiveness. All babies other than those referred for ABR tests because of risk factors (3.6% in 1993) are eligible for a modified distraction test (DT) carried out by Maternal and Child Health nurses at 7-9 months of age. The distraction test is repeated 4-6 weeks later if pass criteria are not met, to reduce numbers of false positives due to transient conductive losses. The modified test, with more stringent pass-fail criteria than previously and a standardised protocol for delivery, was introduced in 1992 with careful attention to training and quality of testing across the state. In 1993, coverage was approximately 79%. Based on six year population-based follow up of the 1993 birth cohort to end 1999, program sensitivity was approximately 37% with a PPV of 0.5%. For a small subset of 27 children with hearing aids who entered the diagnostic system after nine months of age from whom detailed parent information was obtained, 12 (44%) passed the test, 11 (41%) failed, and the result was not recalled by 4 (15%) of the parents.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test: Neonatal ABR/OAE	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓ but limited evidence
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Characteristics of a Screening Test: Distraction test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✗
Acceptable to the public	✓
Accurate	✗

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓ (hearing augmentation) ? (mode of early intervention)
Facilities for diagnosis and treatment	Variable; even where good, delays in commencement common
Latent or early symptomatic stage	✓
Suitable test or examination	✓
Test acceptable to the population	✓
Natural history adequately understood	✗
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?
Comment: logistics of administration for Australia not yet known	

Repeatable	X
Sensitive	X
Specific	✓

Programs: infant hearing screening

Two major reviews of infant hearing screening were identified. “A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment” (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997) was commissioned by the UK NHS Research & Development Health Technology Assessment (HTA) Programme. “Universal Newborn Hearing Screening: A Systematic Evidence Review” (McPhillips, Thompson, & Davis 2000) was commissioned by the US Agency for Healthcare Research and Quality (AHRQ) for the US Preventive Services Taskforce. This work has so far been presented only in abstract and poster form at scientific meetings. In addition, one nested case-control study has been reported examining outcomes for deaf children born in hospitals with screening vs those born in hospitals without (Yoshinaga-Itano, Coulter & Thomson 2000).

A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment (Davis et al, 1997)

This review takes an epidemiological approach with particular emphasis on the UK situation. Rather than posing a single question (does UNHS improve outcomes?) addressed in an evidence-based framework, it attempts to synthesise evidence as to whether early detection is important, and if so what would currently seem to offer the best alternative in the real-world setting. Its objectives were to review the available literature on the screening of permanent childhood hearing impairment; to provide commissioners and providers of health care with information about how to deliver a more uniform service, better outcomes, and more cost-effective screening; and to identify areas for further research and service development. Coverage and feasibility, service structures, impact and acceptability of screens, and implementation issues were important aspects of this review. It incorporated a health economics study of hearing screening costs in the UK. Articles were not selected on the basis of levels of evidence, and studies were not graded in any way.

Essentially, this review concluded that the best evidence available supports strong efforts to systematically detect moderate or greater PCHI, while acknowledging that its feasibility and its benefits at a population level are as yet unproven. It then undertook a theoretical review of the likely costs, efficiency, responsiveness, equity and benefits of nine possible options for the UK situation (see *Cost Effectiveness* below) concluding that the weight of evidence strongly supports the introduction of UNHS, supplemented by a targeted infant distraction test at about 7 months of age. It provided a detailed discussion of the very considerable service, organisational, and information technology issues to be faced before successful national implementation of the desired quality could be achieved.

Since this review, the UK has gone a considerable way towards national implementation of UNHS, without supplementary distraction testing.

Universal Newborn Hearing Screening: A Systematic Evidence Review (McPhillips et al, 2000)

Using an “Analytic Framework” addressing four key questions (see below), 864 abstracts were obtained from MEDLINE, CINAHL and PsychINFO; 700 were excluded and 164 included (56 reviews, 108 abstracts). Abstracts were screened in two stages: (1) abstracts were reviewed and either included or excluded based on relevance to the topic; (2) articles were then reviewed for established inclusion criteria by two reviewers. Inclusion criteria were relevance to topic, English language and presence of a comparison group. Finally, two reviewers abstracted the included articles using a standardised form.

Four key questions were addressed within the review’s Analytic Framework.

A. Is there direct evidence (randomised or case-control study) that screening improves outcomes?

No prospective controlled studies were identified reporting the impact of UNHS on language or cognitive outcomes.

How accurate is the screening test and what are the adverse effects?

A total of 71 articles pertained to the accuracy of screening, including five state programs, three hospital series, and one controlled trial. Reported coverage rates ranged from 81-99%. The PPV of screening ranged from 2.2-11.7%, and the NNS to identify one case of bilateral hearing impairment ranged from 303-925. Between 10 and 52% of those failing screening were reported to be lost to follow-up.

The Wessex Controlled Trial was the only study located comparing the yield of screening in high risk and low risk populations (Wessex Universal Neonatal Hearing Screening Trial Group 1998). In this trial a

screening program was rotated between four hospitals, with a total of 21,279 infants screened. The low risk yield was 7 of 19,555 (NNS = 2,794), while the high risk yield was 20 of 1,724 (NNS = 86).

The literature search for potential adverse effects of screening revealed two small surveys demonstrating good parental acceptance of screening and low parental anxiety related to test results. Little information was found about adverse effects including the effects of false-positive tests, the effects of early identification of mild/unilateral losses, and surgical treatment of transient effusions.

Does early detection lead to early intervention?

In the Wessex Controlled Trial, (Wessex Universal Neonatal Hearing Screening Trial Group 1998) 15 of 27 infants (55%) in the screened group were managed by 10 months, compared to only 7 of 26 (27%) in the unscreened group. The Rhode Island study reported a mean age of amplification of 13.1 months in 1993 (prior to commencing screening) compared to 5.7 months in 1996 (after commencement of screening) (Vohr et al. 1998) though this followup, being short, may have lead to an overestimate of longer-term effectiveness.

Does early intervention improve outcomes?

Of 40 articles pertaining to the efficacy of early treatment, most had design limitations. The best evidence identified was the single retrospective cohort study of Yoshinaga-Itano (Yoshinaga-Itano, Sedey, Coulter, & Mehl 1998) described above, indicating a positive effect of treatment by age 6 months on language outcomes not modified by severity of hearing impairment.

The review concludes that:

no single study assessed the effect of screening on future language and communication abilities

PPV is low, particularly in population-based studies, but possibly acceptable

early treatment may improve language abilities

there is insufficient evidence on adverse effects of screening, particularly given the low PPV.

Infant hearing impairment and universal hearing screening. The Colorado Newborn Hearing Screening Project: effects on speech and language development for children with hearing loss (Yoshinaga-Itano et al, 2000)

This study is the first to compare outcomes for deaf children born in hospitals with newborn screening to those of deaf children born in hospitals without. A “natural experiment”, it utilised the facts that (a) in the US State of Colorado there was a brief period when approximately half the birthing hospitals had implemented UNHS while the other half were yet to implement, and (b) most deaf children, regardless of birthing hospital, were referred to a single early intervention research program which prospectively collected outcome data. Twenty-five pairs were apparently matched on severity of hearing loss (mild-profound), cognition (quotients 75-132) and age (9-61 months) though this is not explicitly stated. Total, Expressive and Receptive Language Quotients were derived from the Minnesota Child Development Inventory.

Language quotients for infants born in “screening” hospitals vs “nonscreening” hospitals were as follows:

total language quotient	82.2 vs 62.1	p<0.001
receptive language quotient	81.5 vs 64.6	p<0.001
expressive language quotient	82.9 vs 63.1	p<0.001

Results were also collapsed into normal ($LQ \geq 80$), borderline ($70 \leq LQ < 79$), and delayed ($LQ < 70$) groups. 24% of the screened vs 68% of the non-screened groups were delayed. 56% of the screened vs 24% of the non-screened were in the normal range. Those born in screening hospitals were 2.33 (95% CI 1.07, 5.09; $p < 0.05$) times more likely to have language in the normal range compared to peers born in hospitals without a screening program.

The Victorian Infant Hearing Screening Program

The Victorian Infant Hearing Screening Program (VIHSP) has evaluated the population effectiveness of targeted neonatal screening plus universal DT. This program was implemented in 1992 with the aims of detecting and habilitating all Victorian children (birth rate approximately 60,000 pa) with PCHI >40dB by the age of 12 months. All babies assessed as being at high risk in the neonatal period are referred to audiologists for screening as early as possible; all other babies are screened by Maternal & Child Health nurses at eight months of age using a modified distraction test (see above). MCHNs routinely enquire about hearing concerns at all major well-baby checks. Early results appeared good, with an immediate significant and sustained rise in the numbers of infants detected and aided by 6 months of age, from an annual rate of 0.7 pre-VIHSP to 8.3 post-VIHSP. However, a population-based followup with high ascertainment of the entire 1993 birth cohort compared to the entire 1989 birth cohort to 6 years of age eroded these results. Median age at diagnosis was lower post-VIHSP for severe impairment (11.2 months 1993 vs 22.4 months 1989) and

moderate impairment (25.1 months vs 36.4 months), but not for profound (9.9 months vs 10.3 months) or mild (44.5 months vs 39.6 months) impairments. Overall, median age of diagnosis of moderate or greater PCHI fell from 20.3 months to 15.7 months. Being assessed as being in a high-risk category had a PPV of 0.011 with 91 babies needing to be screened to detect one child with PCHI requiring aiding. Being referred for two failed DTs had a much lower PPV of 0.005, with 183 babies needing to be referred to detect one child with PCHI requiring aiding.

Genetic screening

No papers have yet reported a trial of genetic screening for early detection of PCHI.

Harms

Though little researched, neonatal screening for PCHI appears to be overall well accepted by parents with few declining testing when offered and the vast majority supporting testing across a number of studies. Watkin studied 288 parents whose babies had undergone neonatal screening using OAE. After the first test, anxiety was low and 97% felt that the test was worthwhile. Fifteen per cent reported some anxiety and <1% marked anxiety. Of the 57 who had retests because they did not pass the screen, two (3.5%) were very worried but still positive about the screen (Watkin et al. 1998). DeUzategul surveyed parents of infants failing initial screening who were referred for additional testing. Of those responding, 22% of parents reported that they were angry, 42% confused, 52% afraid, 37% depressed, 31% frustrated, 27% shocked, 42% sad, 19% guilty, 16% lonely, and 11% that they felt blame. Follow up after definitive testing was not reported (deUzategul & Yoshinaga-Itano 1997). Clemens surveyed 49 mothers whose infants had failed UNHS via automated ABR (53% of those referred for a second stage outpatient screen, 64% of those successfully completing it). More than 90% of these mothers felt that UNHS was a good idea, but 9% reported that they had treated their child differently (eg talking loudly, clapping to check hearing) before outpatient rescreening and 14% reported lasting anxiety even after passing their outpatient screen (Clemens, Davis, & Bailey 2000). One study examining the effect of age of hearing aid fitting on a standardised scale of Quality of Family Life made the surprising finding that, overall, very early diagnosis was in fact associated with considerably *poorer* long term family quality of life (Hind & Davis 2000). On stratification, this finding was limited to parents who were dissatisfied with services at diagnosis and first hearing aid fitting, emphasising the importance of high quality diagnostic services, support, counselling and management services that must back up screening programs for hearing impairment. However, even in those satisfied with the service provided, early diagnosis was not associated with better long term Quality of Family Life scores.

Programs: school screening

School entry hearing screening, once a cornerstone of the school health system in many countries, has been discontinued in many areas in Australia. Originally aimed at detecting SNHL, such programs have more recently aimed primarily to detect conductive hearing loss. In a recent school-based cross sectional study of 1218 American children (43% response) in grades 3, 6 and 9, 5.4% were found to have minimal sensorineural hearing loss defined as unilateral SNHL (≥ 20 dB HL) or bilateral SNHL (20-40 dB HL) or high frequency SNHL (>25 dB HL at two or more frequencies above 2 kHz) (Bess, Dodd-Murphy, & Parker 1998). The number of these who had previously been diagnosed was not stated. These figures are much higher than those from a recent representative cross sectional study of hearing in Victorian school entry children (Heathershaw & Wake 2000). In this study, only one child of the 1563 tested had previously undiagnosed mild bilateral SNHL, while seven had undiagnosed unilateral losses (one profound, one severe, three mild high frequency, and one mild low-mid frequency) with an overall prevalence of undiagnosed minimal SNHL as defined by Bess of only 0.45%. How much of this difference is due to a lower prevalence of detectable progressive and acquired loss in school entry children than in Bess's older children cannot be determined from the information provided.

The yield of school based screening once an effective neonatal or preschool screening program is in place is not known. Almost all cases of moderate or greater bilateral SNHL are detected prior to school entry even without UNHS, so that screening would be expected to mainly detect mild, unilateral high frequency, and progressive hearing losses. The low cut points needed to detect mild SNHL would also detect large numbers of conductive hearing impairments, for which the harm/benefit ratio is not known. The effects of mild, high frequency, unilateral and progressive SNHL are not well understood but are concerning, especially as progressive hearing impairment due to noise exposure may be increasing (Bess, Dodd-Murphy, & Parker 1998).

Cost effectiveness

Several cost analyses of neonatal hearing screening programs have been published. Many studies have reported the costs of hearing screening, usually in the context of UNHS programs in relatively large hospitals rather than on a true population basis. Some reports have also attempted to estimate savings in (for example)

educational costs that might result from earlier detection. This is more difficult, since neither the proportion of disability that can be averted by early diagnosis nor the possible reduction in required educational input has been well quantified.

Costs of screening vary. Mason reported costs over a five year period of universal automated ABR screening of US\$17 per infant in a large newborn nursery with 96% coverage before discharge, and cost to identify each true bilateral hearing loss of US\$17,750 (Mason & Herrmann 1998) Maxon et al reported that the average 1993 cost of the TEOAE-based Rhode Island UNHS program was US\$26.05 per infant, including all salaries, fringe benefits, hospital overhead, supplies and equipment amortised over a five year period, initial screen, re-screens, tracking and scheduling appointments, and program management and coordination, but excluding costs of actual diagnostic assessments (Maxon et al. 1995). This compared with only US\$7.42 per baby for a small Utah hospital in 1996 accounting for similar costs (Weirather 1997). Mehl and Thomson in 1998 estimated the costs of UNHS in Colorado as US\$18.30-\$25.60 per infant screened, with the average cost per confirmed diagnosis US\$9,600 (Mehl & Thomson 1998). Vohr assessed the economic outcomes of three different UNHS programs (TEOAE vs AABR vs a two-step protocol of TEOAE + AABR) in five sites, using prospective activity-based costing techniques. She concluded that the cost per infant (total pre- and post-discharge cost) and cost per identified child did not differ significantly among programs, and that refer rates and costs were most sensitive to personnel type (dedicated vs part time) and infant length of stay, particularly for two-step and TEOAE programs (Vohr et al. 1999).

In the UK, Stevens et al surveyed ten centres in England and Wales to determine the current costs of hearing screening in the first year of life (Stevens et al. 1998) Centres surveyed tended to be in urbanised areas and to have an active interest in hearing impairment. Mean service costs for targeted neonatal screening, UNHS, and the health visitor distraction test at 1994 costs were UK£5,052, £13,881 and £24,519 for standardised districts of 1000 live births. No attempt was made to cost items of equipment, consumables, or other non-direct staff costs used by each service.

Few formal cost-effectiveness analyses have been reported. Assuming that half of hearing-impaired children realise some ultimate savings in school-based costs because of newborn screening and early amplification, Mehl and Thomson estimate that screening costs for the state of Colorado (birth rate 54,000 per annum) would be recovered after ten years (Mehl & Thomson 1998). Kemper carried out a theoretical cost-effectiveness analysis of targeted vs universal neonatal screening, assuming the use of TEOAE followed up by automated ABR for screen positives (Kemper & Downs 2000). Other assumptions included a true positive rate for bilateral CHI >40 dB of 1.1/1000; OAE sensitivity of 80% and specificity of 92%; ABR sensitivity of 98% and specificity of 96%; and cost of accurate ascertainment of risk factors of US\$1/infant. Short term costs (including costs of screening and follow up testing) were included. Long term costs such as costs of treatment and any potential savings resulting from earlier management were not included since they were considered unknown, nor were indirect costs such as parental wages lost and transportation costs. They concluded that for every 100,000 newborns, universal screening would detect 86 of 110 cases of CHI, at a cost of US\$11,650 per case identified. Targeted screening would detect 51 of 110 cases, at US\$3,120 per case identified. Universal screening would result in 320 false positives compared to only 16 for targeted screening. Universal screening would cost US\$23,930 more than targeted screening per case detected. However, if the costs of ascertaining risk factors were more than US\$5.34 rather than \$1 per infant, universal screening would become cheaper, while if it were as high as US\$15/infant, targeted screening would cost US\$18,900 more per case identified than universal screening.

Davis undertook a theoretical review of the likely costs, efficiency, responsiveness, equity and benefits of nine possible options for the UK situation (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). These comprised:

1. no screening
2. universal distraction testing (DT) at 6-8 months
3. targeted neonatal screening plus universal DT
4. targeted neonatal screening plus technologically-improved universal DT
5. targeted neonatal screening plus universal parent questionnaire at 6-8 months
6. targeted neonatal screening only
7. targeted neonatal screening plus targeted DT
8. UNHS
9. UNHS plus targeted DT.

Results are summarised in the Table. Costs did not include estimated long term costs of habilitation, or any excess in these costs if habilitation were delayed. In these models, targeted neonatal screening was the most

efficient at £10,000 per case identified but would identify only 0.5/1000 and miss up to 0.6/1000 children with PCHI. UNHS had the highest likely yield of 0.9/1000 at the next lowest cost of £15,000 per case identified, rising to a potential yield of 1.0/1000 at £17,000 per case if DT testing were retained for those missing their neonatal screen or at high risk of progressive hearing loss.

No studies have yet estimated the possible cost-effectiveness of genetic testing programs (which would be able to use existing newborn blood spot collections, for which coverage is already close to 100%) combined with at-risk testing.

Current screening approaches

Practices vary widely across and within countries at the present time. The idea of UNHS has gained political momentum and many countries are aiming at its implementation on a large scale. In the US, the 1993 NIH Consensus Statement endorsed the concept of UNHS, and in 1998 the US Early Hearing Detection and Intervention Act mandated UNHS for all American babies prior to discharge from hospital, with block grant funding to the individual states to implement programs. Currently, 27 states have mandated UNHS, and it is supported by many influential national organisations such as the American Speech, Language and Hearing Association, the American Academy of Audiology, and the American Academy of Pediatrics. In the UK, practices and standards of practices vary widely from region to region (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997). The UK National Screening Committee has endorsed the principle of UNHS and progress is being made towards national implementation. The 1998 European Consensus Statement on Neonatal Hearing Screening also endorsed UNHS as a goal.

In Australia, a number of approaches are being trialed. The two-tier Victorian Infant Hearing Screening Program is described above. West Australia implemented a trial of UNHS in February 2000, the program funded by the state government (supplemented by a research grant to examine outcomes). Newborn hearing screening is now carried out in five Perth metropolitan hospitals (contributing approximately 40% of the 25,000 babies born in Western Australia each year) with the aim of detecting bilateral hearing loss and commencing intervention by 6 months of age. Infants are screened with OAEs, complemented by an automated ABR prior to discharge if a pass response is not obtained in both ears (approximately 25% of children tested in the first 48 hours at the end of the first 3 months of the program). Children not passing ABR (approximately 8% at the end of the first 3 months) are then followed up at 3-4 weeks post-discharge, and only if they fail this screen are they then referred to an audiologist. No results are yet available from this trial.

Despite the widespread interest in hearing screening, to date this effort has not been coordinated at a national level. A number of state-based multidisciplinary groups are currently working toward implementation of UNHS, with some areas concentrating on screening of at-risk neonates particularly those in NICUs and others working directly towards UNHS. The Australian Public Health Association (PHA) developed a policy statement in 1999 which was put on hold until 2000 pending this review. The Audiological Society of Australia is currently developing a national policy statement.

Australia poses particular logistic challenges, with a large number of hospitals for a relatively low birth rate. In the US and the UK very high coverage (>95%) has been achieved with UNHS in relatively populous areas in which births are concentrated in a small number of hospitals. However, many Australian women do not live within relatively close range of major hospitals. For example, in 1997-8 7% of Victoria's 60,000 annual confinements took place in 74 hospitals averaging fewer than 5 confinements per week, 11% in 18 hospitals averaging 5 to <10 per week, and 20% in 15 hospitals averaging 10 to <20 per week (Victorian Inpatient Minimum Dataset). Twenty per cent of confinements are for two days or less including the delivery, so that the post-delivery time available for testing would be short. The many small hospitals will tend to have higher false positive rates (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997) and therefore more referrals for definitive testing in areas least likely to have adequate follow-up services. Most other states have much larger geographic areas with which to contend. While there is no doubt that the sensitivity of *individual* tests is high, achieving a hospital-based UNHS system with adequate *program* sensitivity will therefore pose formidable logistic challenges. Community based testing could potentially achieve adequate coverage while obviating the salary costs of testing on a daily basis, but has as yet received little research attention.

The rapid development of the genetics of PCHI may offer competitive alternatives in the next five years. For example, good risk ascertainment (inpatient screening of all neonates admitted to NICUs or SCBUs for more than 48 hours, coupled with careful community nurse screening and referral for other risk factors) should detect close to 50% of children with PCHI. Coupled with universal genetic testing at birth using the existing neonatal blood spot collections, for which coverage is already close to 100%, this could potentially detect similar numbers of children with PCHI as a UNHS system with 80% coverage and 90% sensitivity. It could also have the added advantage of detecting children at high risk of progressive loss who would not be detected by neonatal hearing screening.

Quality of the evidence

Quality issues for the two systematic reviews are discussed in turn.

A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment (Davis, Bamford, Wilson, Ramkalawan, Forshaw, & Wright 1997):

Level of evidence

III-3 (comparatives studies with historical controls)

Quality

1. Clearly defined question and inclusion criteria

B (some limitations) A clear aim or research question, but no list of clear inclusion criteria was reported

2. Comprehensive search

B (some limitations) More than 1 electronic database was used BUT the search appears to have been restricted to the English language literature only

3. Critical appraisal of the validity of studies reviewed

B (some limitations) Critical appraisal was undertaken but explicit criteria were not stated

4. Consistency of results

B (some limitations) Heterogeneity of size of effect but trend obvious (ie: all studies indicate positive effects, but of differing magnitudes)

5. Number of studies included in review

210 references covering a broad range of subtopics. No formal meta-analysis or pooling of results.

Statistical precision - not applicable

Size of effect (estimates from knowledge of literature – not meta-analysis)

(a) Universal neonatal hearing screening

Coverage >90% possible for UNHS

Screen specificity >90%

Screen sensitivity for individuals with PCHI >40 dB 80-100%

Screen sensitivity for programs ? ≈80%

Yield 1-1.3/1000; decreases subsequent incremental yields of distraction testing to very low levels

(b) At risk neonatal hearing screening

Potential yield 60%

Achievable yield likely to be ≤45-50% because of difficulty implementing full coverage for all risk indicators

Relevance of evidence – high

Universal newborn hearing screening: a systematic evidence review (McPhillips, Thompson, & Davis 2000):

Level of evidence

III-2 (a single comparative study with concurrent controls and allocations not randomised, and a single cohort study)

Quality

1. Clearly defined question and inclusion criteria

A (high quality) Clear aim or research question AND a list of clear inclusion criteria was reported

2. Comprehensive search

B (some limitations) More than 1 electronic database was used BUT the search was restricted to the English language literature only

3. Critical appraisal of the validity of studies reviewed

B (some limitations) Critical appraisal was undertaken but explicit criteria not stated (this may be addressed when the published report becomes available)

4. Consistency of results

B (some limitations) Insufficient studies to assess heterogeneity of size of effect

5. Number of studies included in review

864 abstracts were obtained from MEDLINE, CINAHL and PsychINFO; 700 were excluded and 164 included (56 reviews, 108 abstracts). Ultimately, this included:

(a) One controlled trial of accuracy of screening, the Wessex study (Wessex Universal Neonatal Hearing Screening Trial Group 1998)

(b) No controlled trials of effect of early treatment on language outcomes; one retrospective cohort study demonstrating positive effect of early treatment on language outcomes, by Yoshinaga-Itano (Yoshinaga-Itano, Sedey, Coulter, & Mehl 1998)

Statistical precision

(a) Yield of screening, based on Wessex study (Wessex Universal Neonatal Hearing Screening Trial Group 1998):

All severities:

Refer by age 6 months, UNHS vs no UNHS:	p=0.006, RR 2.9 (95% CI 1.4, 6.3)
Confirm by age 10 months, UNHS vs no UNHS:	p=0.22, RR 1.8 (95% CI 0.8, 3.9)
Manage by 10 months, UNHS vs no UNHS:	p=0.08, RR 2.4 (95% CI 1.0, 6.0)

Moderate/severe only:

Refer by age 6 months, UNHS vs no UNHS:	p=0.0002, RR 7.7 (95% CI 2.3, 26.0)
Confirm by age 10 months, UNHS vs no UNHS:	p=0.0042, RR 7.21.8 (95% CI 1.6, 32.0)
Manage by 10 months, UNHS vs no UNHS:	p=0.002, RR 13.0 (95% CI 1.7, 102.0)

Profound only:

Refer by age 6 months, UNHS vs no UNHS:	p=0.61, RR 0.6 (95% CI 0.1, 2.2)
Confirm by age 10 months, UNHS vs no UNHS:	p=0.29, RR 0.4 (95% CI 0.1, 1.6)
Manage by 10 months, UNHS vs no UNHS:	p=0.61, RR 0.6 (95% CI 0.1, 2.2)

(b) Effects on language outcomes (based on Yoshinaga-Itano study (Yoshinaga-Itano, Sedey, Coulter, & Mehl 1998))

In infants identified ≤ 6 month vs > 6 months, total group means adjusted for cognitive ability:

total language quotient	79.0 vs 63.8	p<0.001
receptive language quotient	79.6 vs 64.6	p<0.001
expressive language quotient	78.3 vs 63.1	p<0.001

In infants with cognitive quotient ≥ 80 identified ≤ 6 month vs > 6 months, group means adjusted for cognitive ability:

total language quotient	91.3 vs 70.2	p<0.001
receptive language quotient	92.2 vs 71.7	p<0.001
expressive language quotient	90.5 vs 68.7	p<0.001

In infants with cognitive quotient < 80 identified ≤ 6 month vs > 6 months, group means adjusted for cognitive ability:

total language quotient	59.6 vs 51.7	p=0.05
receptive language quotient	60.4 vs 51.8	p=0.06
expressive language quotient	58.8 vs 51.7	p=0.09

Size of effect

Accuracy of screening (based on Wessex study (Wessex Universal Neonatal Hearing Screening Trial Group 1998)): 71/100,000 more babies with PCHI ≥ 40 dB referred before 6 months of age during periods of

hospital-based UNHS than periods with health visitor distraction testing alone. 87% coverage of births in participating hospitals, which in turn represented 73-92% of all births in relevant districts (ie 63-80% population coverage). If we accept the study's assumption of 80% sensitivity (the stated observed versus expected yield of PCHI), this equates to 50-64% program sensitivity.

(b) Effects on language outcomes (based on Yoshinaga-Itano retrospective study (Yoshinaga-Itano, Sedey, Coulter, & Mehl 1998)):

- see Statistical Precision, above

Relevance of evidence – high

6.3.7b.3. CONCLUSIONS

Fair evidence to recommend universal neonatal hearing screening

Insufficient evidence to make a recommendation for or against genetic screening

Insufficient evidence to make a recommendation for or against school entry screening

Good evidence to recommend against distraction testing

Comments:

Hearing screening before discharge for all NICU neonates and preferably all neonates admitted to special care nurseries for more than 48 hours is now accepted best practice.

There is excellent evidence that it is feasible to test the hearing of *all* newborns within maternity hospitals, and that this results in the detection of many children with significant hearing impairment much earlier than would otherwise be the case. There is fair evidence to support benefits of such screening in terms of language outcomes, but this has not been well quantified, and evidence to support benefits on academic outcomes and quality of life is awaited. Testing appears to be well accepted by parents but this requires further study.

However, there is as yet little evidence of *program* sensitivity when applied at population level. We await information on population benefits in terms of median age of diagnosis and proportions of children diagnosed late. We also await evidence of incremental benefits in language, academic and quality of life outcomes at a population level. US programs have reported relatively high levels of loss to follow up.

Population-based programs are likely to stand or fall on the logistics and quality of the testing system, and on the logistics and quality of the follow up system for screen positives. Given the expense of universal neonatal hearing screening, such issues need to be addressed *before* decisions are made to undertake universal screen for whole populations. In particular, Australia is large and has many small communities geographically removed from support services. These will be a particular issue for program delivery and may necessitate adapted or alternative systems if adequate outcomes are to be obtained.

Genetic tests are highly sensitive and specific for approximately 15-20% of deaf children with single gene mutations, most of whom do not have identified risk factors for deafness. This proportion is likely to double over the next five years. Logistic issues have already been extraordinarily well dealt with for universal neonatal heelprick testing. Should universal hearing screening prove very difficult to fully implement in Australia, an alternative could be to test the hearing of all NICU and special care babies who are at risk for deafness, backed up with genetic screening of all other babies through the existing heelprick system. If successful, this could offer major logistic and cost advantages. However, this is as yet only at the research stage.

6.3.7b.4. RECOMMENDATIONS

1. Hearing screening before discharge for all NICU neonates and preferably all neonates admitted to special care nurseries for more than 48 hours is now accepted best practice, and should become a high priority at State level.
2. The evidence supports some form of neonatal screening for hearing impairment. However, to be successful implementation packages require careful trialing for the situation in which they will be used. Pilot projects should be mounted to trial the feasibility of different models of infant screening, paying careful attention to processes, acceptability and costs, in both large and small communities. These should include settings other than major metropolitan hospitals (already under study in West Australia.) Models to consider could include at-risk screening with and without genetic screening, hospital based UNHS, and community based UNHS.

3. An adequately funded economic analysis should be undertaken urgently to examine costs of different models (including genetic screening) of infant hearing screening for Australian settings. This should take into account likely achievable coverage and program sensitivity for each model at a population level.
4. A national forum should be convened to discuss possible infant screening models in terms of logistics, coverage, acceptability, program maintenance and quality, minimum standards, data collection/reporting and outcome analysis at a population level for each state and for Australia as a whole. This should take results of the economic analysis into account. Models to consider could include at-risk screening with and without genetic screening, hospital based UNHS, and community based UNHS.
5. Final decisions about implementation of universal neonatal screening programs in Australia should be guided by the above information when available.
6. If the above processes lead to an agreement about proceeding with a specific model(s), planning and implementation would ideally be overseen by a national body set up for this purpose. This could draw on the experience of Breastscreen, the most recent new screening program implemented in Australia. Formal channels of communication with the UK, US, and Canada should also be established.
7. Careful population-based documentation of a wide range of outcomes for children with permanent childhood hearing impairment should commence now, against which to assess the benefits of new programs if/when introduced.

6.3.7b.5. FURTHER RESEARCH

Many questions about infant hearing screening remain unanswered or partly answered. Possible research questions for Australia include:

- What are the effectiveness and efficiency of different models of early hearing screening at population level, weighed against short and long term harms?
- What proportion of disability can be prevented by early detection and management, and for whom?
- What aspects of early intervention and management are most crucial, and for whom?
- How do we most effectively deliver appropriate and ethical counselling and education to affected families?
- What is the role of genetic analyses in screening and management programs?
- Does early detection and management of mild and/or unilateral sensorineural hearing loss improve outcomes (ie in infancy/preschool years)?
- What is the prevalence of mild, unilateral and high-frequency hearing impairment in Australian school aged children, and what are the educational impacts?
- Are there opportunities in childhood and adolescence to prevent progressive hearing loss?

6.3.7b.6. SEARCH TERMS

Hearing

Auditory perceptual disorders

Hearing disorders

Hearing tests

Hearing loss

Auditory perception

Audiology

Deafness

6.3.7b.7. ACKNOWLEDGMENTS

Drafts of this document were circulated to:

Prof Field Rickards, Department of Learning & Educational Development, University of Melbourne
 Prof Barbara Cone-Wesson, Audiology & Speech Sciences, Department of Otolaryngology, University of Melbourne
 Dr Kerry Saunders, Monash Medical Centre, Melbourne
 Ms Jan Pollard, Chief Audiologist, Royal Children's Hospital, Melbourne
 Dr Peter Baghurst, Public Health Research Unit, Women's and Children's Hospital, Adelaide
 Dr David Starte, Developmental Paediatrician, Chatswood Assessment Centre, NSW
 Mr Harvey Oates, Perth, West Australia
 Prof Adrian Davis, MRC Institute of Hearing Research, Nottingham, UK
 Prof David Hall, Institute of General Practice and Primary Care, Sheffield, UK

6.3.7b.8. REFERENCES

- Berettini, S., Ravecca, F., Sellari-Franceschini, S., Mattuecci, F., Siciliano, G., & Ursino, F. 1999, "Progressive sensorineural hearing loss in childhood.", *Pediatr Neurol*, vol. 20, pp. 130-136.
- Bess, F. H., Dodd-Murphy, J., & Parker, R. A. 1998, "Children with minimal sensorineural hearing loss: prevalence, educational performance, and functional status", *Ear & Hearing*, vol. 19, pp. 339-354.
- Birtles, G., Hodgson, F., Lovegrove, R., Mutton, P., Rosen, J., Starte, D., & Williams, K. 1998, *Early Identification of Hearing Impairment in Children in NSW: Working Party Report*, Parent Council for Deaf Education Inc., NSW.
- Clemens, C. J., Davis, A. S., & Bailey, A. R. 2000, "The False-Positive in Universal Newborn Hearing Screening", *Pediatrics*, vol. 106, no. 1, pp. 1-5.
- Davis, A., Bamford, J., Wilson, I., Ramkalawan, T., Forshaw, M., & Wright, S. 1997, *A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment*. 1.
- deUzcategul, C. A. & Yoshinaga-Itano, C. 1997, "Parents' reactions to newborn hearing screening.", *Audiology Today*, vol. 24, p. 27.
- Estivill, X., Fortina, P., Surrey, S., Rabionet, R., Melchionda, S., D'Agruma, L., Mansfield, E., Rappaport, E., Govea, N., Mila, M., Zelante, L., & Gasparini, P. 1998, "Connexin-26 mutations in sporadic and inherited sensorineural deafness", *Lancet*, vol. 351, no. 9100, pp. 394-398.
- Estivill, X. & Gasparini, P. Connexins and Deafness. Connexin Deafness Homepage . 4-2-2000. The Deafness Research Group. <http://www.iro.es/cx26deaf.html> 1999.
- Ewing, I. R. & Ewing, A. W. G. 1944, "The ascertainment of deafness in infancy and early childhood", *Journal of Laryngology and Otology*, vol. 59, pp. 309-333.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. 1999, "Parent-reported language and communication skills at one and two years of age in relation to otitis media in the first two years of life", *Pediatrics*, vol. 104, no. 4, p. e52.
- Heathershaw, K. & Wake, M. 2000, "Effectiveness of a parent questionnaire in the detection of mild to moderate hearing loss in school entry children.", *Pediatric Research*, vol. 47, no. Part 2, p. 1157A (196).
- Herrman, B. S., Thornton, A. R., & Joseph, J. M. 1995, "Automated infant hearing screening using the ABR: development and validation.", *American Journal of Audiology*, vol. 4, pp. 6-14.
- Hind, S. E. & Davis, A. "Outcomes for Children with Permanent Hearing Impairment.", in *A Sound Foundation Through Early Amplification. Proceedings of an International Conference.*, 17 edn, Richard C. Seewald, ed., pp. 199-212.
- Hyde, M. D., Malizia, K., Riko, K., & Alberti, P. W. 1991, "Audiometric estimation error with the ABR in high risk infants", *Acta Oto-Laryngologica*, vol. 111, no.2, pp. 212-219.
- Kemp, D. T. 1978, "Stimulated acoustic emissions from the human auditory system.", *Journal of the Acoustical Society of America*, vol. 64, pp. 1386-1391.
- Kemper, A. R. & Downs, S. 2000, "A cost-effectiveness analysis of newborn hearing screening strategies", *Archives of Pediatrics & Adolescent Medicine*, vol. 154, pp. 484-488.

- Marazita, M., Rawlings, B., Remington, E., Aros, K., & Nance, W. 1993, "Genetic epidemiology studies of early-onset deafness in the US school age population.", *Am J Med Genet*, vol. 46, pp. 486-491.
- Markides, A. 1986, "Age at fitting of hearing aids and speech intelligibility", *British Journal of Audiology*, vol. 20, no. 2, pp. 165-167.
- Mason, J. & Herrmann, K. 1998, "Universal infant hearing screening by automated auditory brainstem response measurement", *Pediatrics*, vol. 101, no. 2, pp. 221-228.
- Maxon, A. B., White, K. R., Behrens, T. R., & Vohr, B. R. 1995, "Referral rates and cost efficiency in a universal newborn hearing screening program using transient evoked otoacoustic emissions.", *Journal of The American Academy of Audiology*, vol. 6, no. 4, pp. 271-277.
- McPhillips, H. A., Thompson, D. C., & Davis, R. L. 2000, "Universal newborn hearing screening: a systematic evidence review.", *Pediatric Research*, vol. 47, no. Part 2, p. 209A (1234).
- Medicare Services Advisory Committee 1999, *Oto-acoustic emission audiometry*, Commonwealth of Australia, Canberra.
- Mehl, A. L. & Thomson, V. 1998, "Newborn hearing screening: the great omission", *Pediatrics*, vol. 101, no. 1, p. E4.
- Meyer, C., Witte, J., Hildmann, A., Hennecke, K.H., Schunck, K.U., Maul, K., Franke, U., Fahnenstich, H., Rabe, H., Rossi, R., Hartmann, S., & Gortner, L. 1999, "Neonatal screening for hearing disorders in infants at risk: incidence, risk factors, and follow-up", *Pediatrics*, vol. 104, no. 4, pp. 900-910.
- Morton, N.E. 1991, "Genetic Epidemiology of hearing impairment.", *Ann NY Acad Sci*, vol. 630, pp. 16-31.
- Northern, J.L. 1991, "Hearing and Hearing Loss in Children," in *Hearing in Children*, 4 edn, Williams and Wilkins, Baltimore, USA, pp. 1-31.
- Norton, S. J. 1999, "Identification of neonatal hearing impairment. Evaluation of TEOAE, DPOAE and ABR test performance.", *ASHA*, vol. 41, no. 5, p. 211.
- Ramkalawan, T. W. & Davis, A. C. 1992, "The effects of hearing loss and age of intervention on some language metrics in young hearing-impaired children", *British Journal of Audiology*, vol. 26, no. 2, pp. 97-107.
- Rance, G., Clark, G.M., Beer, D., Cone-Wesson, B., Shepherd, R.K., Dowell, R., King, A.M., & Rickards, F.W. 1999, "Clinical findings for a group of infants and young children with auditory neuropathy.", *Ear & Hearing*, vol. 20, pp. 238-252.
- Robinshaw, H. M. 1995, "Early intervention for hearing impairment: differences in the timing of communicative and linguistic development.", *British Journal of Audiology*, vol. 29, pp. 315-334.
- Steel, K. P. 2000, "New interventions in hearing impairment", *British Medical Journal*, vol. 320, pp. 622-625.
- Stevens, J. C., Hall, D. M. B., Davis, A., Davies, C. M., & Dixon, S. 1998, "The costs of early hearing screening in England and Wales", *Archives of Disease in Childhood*, vol. 78, no. 1, pp. 14-19.
- Vohr, B., Oh, W., Stewart, E. J., Bentkover, J. D., Gabbard, S., Lemons, J., Papile, L. A., & Pye, R. 1999, "An economic evaluation of infant hearing screening methods.", *Pediatrics*, vol. 104 (Supplement), p. 737.
- Vohr, B. R., Carty, L. M., Moore, P. E., & Letourneau, K. 1998, "The Rhode Island Hearing Assessment Program: Experience with statewide hearing screening (1993 - 1996)", *Journal of Pediatrics*, vol. 133, no. 3, pp. 353-357.
- Watkin, P. M. 1996, "Outcomes of neonatal screening for hearing loss by otoacoustic emission.", *Archives of Disease in Childhood*, vol. 75, p. F158-F168.
- Watkin, P. M., Baldwin, M., Dixon, S., & Beckman, A. 1998, "Maternal anxiety and attitudes to universal neonatal hearing screening", *British Journal of Audiology*, vol. 32, no. 1, pp. 27-37.
- Weirather, Y. P. 1997, "Cost analysis of TEOAE-based universal newborn hearing screening", *Journal of Communication Disorders*, vol. 30, no. 6, pp. 477-493.
- Wessex Universal Neonatal Hearing Screening Trial Group 1998, "Controlled trial of universal neonatal screening for early identification of permanent childhood hearing impairment", *The Lancet*, vol. 352, pp. 1957-1964.

Wilcox, S. A., Saunders, K., Osborn, A. H., Arnold, A., Wunderlich, J., Kelly, T., Collins, V., Wilcox, L. J., Gardner, R. J. M., Kamarinos, M., Cone-Wesson, B., Williamson, R., & Dahl, H. H. 2000, "High frequency hearing loss correlated with mutations in the GJB2 gene.", *Human Genetics*, vol. 106, pp. 399-405.

Yoshinaga-Itano, C., Sedey, AL., Coulter, D., & Mehl, A. L. 1998, "Language of early- and later-identified children with hearing loss.", *Pediatrics*, vol. 102, pp. 1161-1171.

Yoshinaga-Itano, C., Coulter, D., & Thomson, V. 2000, "Infant hearing impairment and universal hearing screening. The Colorado Newborn Hearing Screening Project: effects on speech and language development for children with hearing loss", *Journal of Perinatology*, vol. 20, pp. S132-S137.

6.3.8. HYPERTENSION

6.3.8.1. BACKGROUND

Clinical picture

Hypertension in adults is defined as a systolic or diastolic reading greater than a specified level. This definition is unworkable in childhood as the normal values vary with the age, sex and height of the child. Childhood hypertension is defined as average systolic and/or diastolic blood pressures greater than the 95th percentile on at least 3 occasions.

Systemic hypertension is rare in children and is usually the result of an underlying disease process (ie secondary hypertension). Primary or essential hypertension, where no underlying cause is found, is generally a disease of adulthood, but may be present in older children and adolescents. Many factors including family history, diet, obesity and stress may affect development of essential hypertension. In adolescents, use of alcohol, tobacco and illegal drugs should also be considered.

Secondary hypertension in childhood is usually related to renal or endocrine disease, coarctation of the aorta or medication. Children with coarctation of the aorta may have hypertension of the upper limbs with lower blood pressure in the lower limbs.

Defining hypertension in children as blood pressure greater than the 95th percentile suggest that 5% of children should be affected. However, since the definition specifies that this must be sustained on at least 3 occasions, it is easier to understand that the prevalence is only 1%. Blood pressure readings decrease with repeated measurements, perhaps due to familiarity with the test and less associated anxiety, or regression to the mean.

Natural history

The natural history of essential hypertension originating in childhood is not well understood. Many children, but not all, continue to have hypertension as adults. The course of the illness and the long-term outcomes for children with secondary hypertension will depend on the nature of the underlying condition. Many children with underlying disease will already have organ damage by the time the condition is diagnosed.

Parental hypertension is a risk factor for high blood pressure in children. Hypertension in childhood is a strong predictor of the disease in adults. Long standing hypertension in adults is linked to congestive heart failure and ischaemic heart disease; aneurysm formation; cerebral haemorrhage and thrombo-embolic stroke; peripheral vascular disease; renal failure and retinopathy.

Prevalence

Hypertension is reported to affect approximately 1% of children, while the incidence of symptomatic hypertension is 0.1% (Hall 1996). In a study of 9355 Australian kindergarten children, Hogg et al found the prevalence of essential hypertension to be 0.3/1000 and secondary hypertension 0.4/1000 (Hogg, Harris, Lawrence, Henning, Wigg, & Jureidini 1998).

Genetics relevant to screening issues

None known

Diagnosis

Clinical examination

Accurate measurement of blood pressure depends on the patient, the equipment and the operator. The diagnosis of hypertension should only be made based on resting blood pressure levels, hence the child must be suitably quiet, comfortable and, as far as possible, at ease. This is complicated further by recent studies suggesting that the fluctuations in children's blood pressure throughout the day make even resting readings unreliable (Gillman & Cook 1995; O'Sullivan, Derrick, Griggs, Foxall, Aitkin, & Wren 1999). The equipment, either sphygmomanometers or automated devices, must be the appropriate size for the patient and regularly calibrated. The operator should be experienced at taking blood pressure measurements in children.

Careful attention to the size of the cuff is required – if too short or too narrow the readings will be artificially high (sometimes producing major over-estimates of blood pressure). If too large, the level may be underestimated (though the error is usually small when an oversized cuff is used). The width of the cuff should be at least two thirds of the length of the upper arm (measured from the acromion to the olecranon) and should not obscure the cubital fossa. The bladder should encircle the upper arm as completely as possible to ensure uniform compression.

Systolic pressure is designated by auscultation of the first Korotkoff sound. Diastolic pressure lies between the muffling (the Korotkoff 4th phase or K4) and eventual disappearance of the Korotkoff sounds (5th phase or K5). There is some debate in the literature about which is the most reliable to use in children and which is the best indicator of later hypertension in adulthood (Gillman & Cook 1995; Sinaiko 1996).

Blood pressure is known to fluctuate throughout the day, hence measuring at the same time of the day would be optimal, but if this is not possible, the time of day the reading is taken should be recorded.

The standard mercury sphygmomanometer is subject to digit preference and observer bias resulting from knowledge of previous readings. Other devices are available that overcome this, however they are considerably more expensive and no reference standards for their use in childhood are available.

Treatment/Management

Treatment can be pharmacologic or non-pharmacologic. Interventions for non-pharmacologic management of essential hypertension are based on education and advice about the risk factors eg weight loss, decreased sodium intake, increased exercise, decreased use of alcohol and tobacco, etc. Antihypertensive drugs may be used in patients with primary or secondary hypertension, however levels at which to start treatment are not well defined in childhood. Other treatments will be specific to any underlying conditions.

Prevention

General advice to undertake regular exercise, eat a well balanced diet and maintain a healthy weight may prevent raised BP levels in children predisposed to essential hypertension.

How might screening reduce the burden of suffering?

Cardiovascular disease is the commonest cause of death in developed countries. Long standing hypertension is linked to congestive heart failure and ischaemic heart disease; aneurysm formation; cerebral haemorrhage and thrombo-embolic stroke; peripheral vascular disease; renal failure and retinopathy. These sequelae result in significant mortality, morbidity and disability in the population. The onset of hypertension in childhood is known to be linked to adult hypertension and thus to the complications above. However, the incidence of hypertension in childhood is extremely small, so would only be expected to contribute in a small way to the overall burden of suffering.

6.3.8.2 EVIDENCE

Tests

Clinical examination is the only screening test used for hypertension. However, its validity and reliability as a diagnostic test in this age group is questionable, making it unsuitable for screening.

Hogg et al set out to measure the incidence of abnormalities in urinalysis and blood pressure in Australian preschool children (Hogg et al. 1998). Of the 335 children with abnormal blood pressure readings invited to return, 264 attended for review and 7 were found to have a clinical abnormality – a yield of 2.5%. Sensitivity, specificity and PPV are only available for their combined regimen of urinalysis plus BP measurement – 81%, 87%, and 20% respectively.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✗
Repeatable	✗
Sensitive	?
Specific	?

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	?
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✗
Test acceptable to the population	✓
Natural history adequately understood	✗
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

Blood pressure measurement at certain childhood milestones, or during ‘well child’ examinations, have been recommended in various screening programs.

No evidence was found to demonstrate that routine screening of blood pressure in children or adolescents was effective in improving health outcomes. While it is known that hypertension in children is linked to the disease in adulthood, there are also many questions about the validity and reliability of blood pressure readings in children; the significance of diurnal variations; and the levels at which to commence treatment.

Studies of hypertension screening in adults have demonstrated negative effects that occur as a result of making patients aware of their condition – depression, absenteeism, marital difficulties, etc. One study screening children aged 10-18 years recorded absenteeism and found no difference between those in the screened group found to be hypertensive and matched controls from neighbouring schools (Stenn, Noce, & Buck 1981).

Cost effectiveness

No information regarding the cost-effectiveness of screening for hypertension was found in the course of this review.

Quality of evidence

No evidence found

Generalisability of evidence

No evidence found

6.3.8.3. CONCLUSIONS

Fair evidence to recommend against screening

Comments:

Childhood hypertension and measurement of blood pressure in children do not fit criteria for screening.

We were unable to find any evidence that measurement of blood pressure as a screening procedure during childhood improves health outcomes, or that it is a valid or reliable screening test.

6.3.8.4. RECOMMENDATIONS

Implementation of screening programs to detect hypertension in well children is not recommended.

6.3.8.5. FURTHER RESEARCH

What is the association between childhood hypertension and adult morbidity?

Can active management of childhood hypertension reduce adult morbidity and mortality?

If screening for hypertension in children is found to reduce the burden of disease, can the reliability of the screening test be improved and what are the appropriate cutpoints?

6.3.8.6. SEARCH TERMS

MeSH headings: Hypertension

Keywords: Hypertension

6.3.8.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Reference group: Garth Alperstein, David Hall

Clinical experts: Dr Jim Wilkinson, Director, Dept of Cardiology, Royal Children's Hospital, Melbourne

Dr Colin Jones, Director, Dept of Nephrology, Royal Children's Hospital, Melbourne

6.3.8.8. REFERENCES

Gillman, M. W. & Cook, N. R. 1995, "Blood pressure measurement in childhood epidemiological studies", *Circulation*, vol. 92, pp. 1049-1057.

Hall, D.M.B. 1996, *Health for All Children*, 3rd edn, Oxford University Press, New York.

Hogg, R. J., Harris, S., Lawrence, D. M., Henning, P. H., Wigg, N., & Jureidini, K. F. 1998, "Renal tract abnormalities detected in Australian preschool children", *Journal of Paediatrics and Child Health*, vol. 34, no. 5, pp. 420-424.

O'Sullivan, J. J., Derrick, G., Griggs, P., Foxall, R., Aitkin, M., & Wren, C. 1999, "Ambulatory blood pressure in schoolchildren", *Archives of Diseases in Childhood Fetal Neonatal Edition*, vol. 80, pp. 529-532.

Sinaiko, A. I. 1996, "Current Concepts: Hypertension in Children", *New England Journal of Medicine*, vol. 335, no. 26, pp. 1968-1073.

Stenn, P. G., Noce, A., & Buck, C. 1981, "A study of the labelling phenomenon in school children with elevated blood pressure", *Clinical Investigative Medicine*, vol. 4, pp. 179-181.

6.3.9. IRON DEFICIENCY

6.3.9.1 BACKGROUND

Clinical picture

Iron deficiency is the most common nutritional deficiency in children and adults in both developed and developing countries (Roeser 1992). However it is most commonly seen in children (children aged 9-18 months are at particular risk), women of childbearing age (U.S. Department of Health and Human Services 1998) and in individuals from lower socioeconomic backgrounds where good diet is less common (Booth & Aukett 1997).

Iron is present in all cells in the human body. Its functions include the transportation of oxygen around the body, the facilitation of oxygen use and storage in the muscles, acting as a transport medium for electrons within the cells and forming part of enzyme reactions in some tissues. Most iron is found in red blood cells as haemoglobin, but some is also found in myoglobin, intracellular respiratory enzymes and stored as ferritin or haemosiderin. The body's ability to absorb iron from the diet is dependent on: (1) the amount of iron already stored in the body (more iron is absorbed when stores are low), (2) the rate of red blood cell production (increased production increases iron absorption), (3) the amount and kind of iron in the diet (eg iron in meat is more readily absorbed than iron in vegetables), and (4) the presence of absorption enhancers and inhibitors in the diet (meat and vitamin C are enhancers and inhibitors include polyphenols, tannins, phytates and calcium) (U.S. Department of Health and Human Services 1998).

Iron deficiency occurs across a spectrum from iron depletion (minor iron deficiency), where stores are depleted but there is enough functional iron in the body to maintain normal functioning, through to anaemia (severe iron deficiency) where both the stores and functional iron are depleted to the extent that normal functions are compromised (U.S. Department of Health and Human Services 1998).

Iron deficiency interferes with the vital functions performed by iron in the cells of the human body and can have severe consequences (U.S. Department of Health and Human Services 1998). Iron deficiency can contribute to lead toxicity by enhancing lead absorption in the tissues and has been implicated as a cause of behaviour problems (U.S. Department of Health and Human Services 1998), developmental delay, and poor growth in children (Booth & Aukett 1997).

In children, clinical indicators of iron deficiency may include behavioural changes (lethargy, irritability, lack of concentration), cognitive and psychomotor deficits, decreased immune function (recurrent infections), loss of appetite and pica. In addition, anaemia causes pallor and, in extreme cases, heart failure. However many cases of iron deficiency in children are asymptomatic (Australian Iron Status Advisory Panel).

Natural History

Newborn stores of iron last until the child is 4-6 months old. This, combined with the efficient absorption of iron from breast milk and the supplementation of iron in infant formulae, means that iron deficiency in children under 4-6 months of age is rare. However, once newborn iron stores are depleted, the child must meet the body's iron needs through dietary intake (National Health and Medical Research Council 1996). Iron-rich foods include red meat, other meats and seafood, some cereals (particularly bran and wholemeal pasta), some beans and pulses, and some fruits and vegetables (particularly dates and raisins) (National Health and Medical Research Council 1996). If there is insufficient iron in the diet or if other problems prevent dietary iron from being absorbed into the body, a child's iron stores will become depleted. The depletion of iron stores occurs in three stages with the end result being anaemia: (1) the red-cell distribution width becomes abnormal as iron is depleted from the bone marrow; (2) serum iron level is reduced, reflecting a loss of transport iron; (3) mean corpuscular volume (MCV) is reduced and there is an increase in red-cell protoporphyrin concentration reflecting an iron-deficient erythropoiesis (Oski 1993).

Prevalence

The prevalence of iron deficiency in infants and children has decreased in developed countries over recent decades due to the introduction of iron fortified formulas and foods (eg breakfast cereals) (U.S. Department of Health and Human Services 1998).

The prevalence of iron deficiency varies with age. The peak prevalence is 18 months, although many children become iron deficient much earlier (Wharton 1999). This has implications for the timing of screening programs.

Prevalence estimates for iron deficiency vary from approximately 2% up to almost 50%, depending on the case definition, study population and the country (with higher prevalences reported in less developed countries) (Looker, Dallman, Carroll, Gunter, & Johnson 1997; De Maeyer, & Adiels-tegman 1985 as cited in Martins, Logan, & Gilbert 2000).

A study conducted between September 1992 and June 1994 of 678 children aged 9-62 months in Sydney found the prevalence of iron depletion¹ was 10.5%, iron deficiency² was 2.8% and iron deficiency anaemia³ was 1.1%. Children aged 24-35 months had the highest prevalence of iron deficiency anaemia (3.0%). Children aged 9-23 months had the highest prevalence of iron depletion and iron deficiency (18.7% and 5.4% respectively) (Karr, Alperstein, Causer, Mira, Lammi, & Fett 1996).

A 1993 study of 234 children (191 Caucasian and 43 Asian) aged 6-24 months in Adelaide found the prevalence of iron deficiency⁴ was 25% in Caucasian and 14% in Asian children. The prevalence of iron deficiency anaemia⁵ was 12% for Caucasian and 6% for Asian children (Oti-Boateng, Seshadri, Petrick, Gibson, & Simmer 1998).

In the 3rd National Health and Nutrition Examination Survey in the USA (NHANES III; 1988-1994) the prevalence of iron deficiency was 9% for children aged 1-2 years (3% had iron deficiency anaemia), decreasing to 3% in children aged 3-5 years (<1% had iron deficiency anaemia) and to 2% in children aged 6-11 years (<1% had iron deficiency anaemia). The prevalence was higher in children living in poverty and in children of African-American or Mexican-American origin (Looker et al. 1997).

Early 1990s UK estimates put the prevalence of 1.5-2 year old children with haemoglobin levels <10µg/L at 12% and found 28% with low ferritin levels. Children aged 6-24 months in low socioeconomic populations were estimated to have a prevalence of iron deficiency anaemia between 25% and 40% (Gregory, Collins, Davies, Hughes & Clarke 1995 as cited in Booth & Aukett 1997). Another UK study indicated that 23% of the 1075 8 month old infants sampled could be classified as anaemic using the WHO definition (haemoglobin <11µg/L) but found no relationship with socioeconomic status (Edmond, Hawkins, Pennock, & Golding 1996).

Genetics relevant to screening issues

None known

Diagnosis

Iron deficiency can be divided into three phases which should be seen as a continuum:

1. Low iron stores, with normal erythropoiesis.
2. Impaired erythropoiesis, resulting in reduced MCV (hypochromic microcytic red cells), but without anaemia.
3. Iron deficiency anaemia (hypochromic microcytic anaemia).

The diagnosis of each phase requires a combination of investigations, usually the full blood count (FBE) with blood film and the serum ferritin. Other tests including the serum iron, transferrin saturation are less useful. The most common and challenging diagnostic group are the patients with reduced MCV but normal haemoglobin who must be differentiated from those with thalassaemia traits.

Serum ferritin is the most accurate measure of iron stores. A reduced ferritin is usually diagnostic of iron deficiency. Ferritin is an acute phase reactant, so in the setting of infection or inflammation, ferritin may be normal even in the presence of iron deficiency (false negative). Reduced MCV suggests impaired erythropoiesis, and usually occurs prior to the development of anaemia. A low MCV with a normal ferritin usually represents a thalassaemia trait. Haemoglobin is the least sensitive test for iron deficiency as haemoglobin levels will only determine the presence or absence of anaemia, the end stage of iron deficiency (personal communication with Paul Monagle). Even so, haemoglobin level remains the most common criterion for screening because it is easy to measure. The criterion for diagnosing iron deficiency anaemia in children aged between 6 months and 4 years is usually haemoglobin level <11µg/L (Oski 1993).

In cases of iron deficiency without anaemia the haemoglobin level has not yet fallen to a level meeting the criteria for iron deficiency anaemia. Iron deficiency without anaemia is usually diagnosed on the basis of serum ferritin, erythrocyte protoporphyrin, or MCV levels rather than haemoglobin levels. In children aged 6 months to 4 years, levels considered diagnostic of iron deficiency are serum ferritin <10µg/L, erythrocyte protoporphyrin >2.5 micro g/g haemoglobin and MCV <72fl (Oski 1993).

¹ defined as a plasma ferritin of <12 g/L

² defined as a plasma ferritin of <12g/L plus a MCV of <70fl for children aged 9-24 months or <73 fl for children aged 24-60 months

³ defined as a haemoglobin of <11µg/L

⁴ defined as haemoglobin concentration >11µg/L and serum ferritin <15µg/L and/or transferrin >3g/L, iron saturation <12% and serum iron <8µmol/L

⁵ defined as iron deficiency with haemoglobin concentration <11µg/L

A further diagnostic technique is to commence children who meet the criteria for iron deficiency (haemoglobin $<11\mu\text{g/L}$ or equivalent and no other identifiable cause) on a trial of iron supplementation. As this is the standard treatment for nutritional iron deficiency, children are monitored and those not responding after a reasonable amount of time (often one month) are investigated for other causes (Choi & Reid 1998).

Treatment/Management

As iron deficiency in children is usually the result of dietary inadequacies, a dietary assessment is usually the first component of management. Following a dietary assessment, the health care practitioner can advise parents on ways to increase their child's consumption of foods rich in iron and those that enhance iron absorption, while decreasing the consumption of foods that hamper iron absorption. Initially children may also be prescribed iron supplements to replete their iron stores.

A recent study found children with severe, chronic iron deficiency in infancy (defined as haemoglobin $\leq 10\mu\text{g/L}$ and high erythrocyte protoporphyrin values) continue to have lower mental and motor functioning at 10 years of age, despite showing an average haemoglobin increase of $\geq 1\mu\text{g/L}$ when treated with iron therapy in infancy (Lozoff, Jimenez, Hagen, Mollen, & Wolf 2000). A systematic review examining the effects of iron therapy on psychomotor development and cognitive functioning in children less than three years of age with iron deficiency is currently under way and may shed more light on the effectiveness of iron therapy as a treatment for iron deficiency (Martins, Logan, & Gilbert 2000).

Prevention

NHMRC recommended dietary intake (RDI) for iron for Australian children is designed to prevent iron deficiency. The RDI (expressed as mean daily intake) for breast fed infants aged up to 6 months is 0.5mg, for bottle fed infants aged up to 6 months is 3mg, for infants 7-12 months of age 9mg, and for children 1-7 years of age 6-8mg per day (National Health and Medical Research Council 1996). An important aspect of prevention is educating parents about the changing dietary needs of their growing child and the types of foods that are rich in iron or which encourage iron absorption and also those that restrict iron absorption. Informing parents of the two most common factors associated with iron deficiency may also be a useful preventative activity. These two factors are (1) being fed on cows' milk prior to 12 months of age, and (2) continuing solely on milk (either breast or cows' milk) after 12 months of age, without the introduction of solids (personal communication with Paul Monagle).

Feeding children iron fortified formula and iron fortified foods (eg breakfast cereals) has been suggested as a prevention strategy. One randomised controlled trial of 493 children nine months of age being fed cows' milk (randomised to cows milk, formula containing 0.9mg/L of iron, or fortified formula containing 1.2mg/L of iron) found higher haemoglobin and plasma ferritin levels at 18 months in the children fed iron fortified formula (compared to cows milk and non-fortified formula) and fewer anaemic children. However, they found no beneficial effect of iron fortified formula on mental or psychomotor development or growth (Morley, Abbott, Fairweather-Tait, MacFadyen, Stephenson, & Lucas 1999). In another randomised controlled trial 283 healthy bottle-fed infants aged birth to two months from very low income families were allocated to receive either regular formula (1.1 mg/L iron) or fortified formula (12.8mg/L iron) which was continued until weaning (encouraged at 9 months). Measures taken at 6, 9, 12 and 15 months of age indicated that at all time points infants receiving the iron fortified formula had higher haemoglobin, MCV, ferritin and transferrin saturation levels and lower erythrocyte distribution width and free erythrocyte protoporphyrin levels. At all time points, the proportion of children classified as anaemic (haemoglobin $<11\mu\text{g/L}$) was significantly lower in the fortified formula group. There appeared to be some beneficial effect of iron fortified formula on psychomotor development, and to a lesser degree, mental development (Moffatt, Longstaffe, Besant, & Dureski 1994).

How might screening reduce the burden of suffering

Screening theoretically might lead to effective therapy which might reverse the physical, behavioural and developmental effects of iron deficiency. However, it is unclear whether early detection through screening and therefore earlier treatment of iron deficiency would have an increased beneficial effect on these problems. In particular, there remains some controversy over the existence of a causal link between iron deficiency and developmental deficit (Logan 1999). A systematic review due to be released shortly examining the effects of iron therapy on psychomotor development and cognitive functioning in children with iron deficiency may provide further evidence to resolve this debate (Martins, Logan, & Gilbert 2000). In addition, iron replacement therapy is unpalatable. It is also potentially toxic, with overdose causing serious illness and sometimes permanent disability or death.

6.3.9.2 EVIDENCE

Tests

Haematological and biochemical tests

Haematological tests measure the characteristics of red blood cells including haemoglobin concentration, haematocrit, mean cell volume and red blood cell distribution width. They are generally more common and less expensive than biochemical tests but are less likely to detect early changes in iron status (U.S. Department of Health and Human Services 1998).

Biochemical tests measure erythrocyte protoporphyrin concentration, serum ferritin concentration and transferrin saturation. Tests measuring these properties are able to detect earlier changes in iron status than are haematological tests (U.S. Department of Health and Human Services 1998). For example, measuring haemoglobin alone will allow severe cases of iron deficiency to be detected but will not detect children with low iron stores who are not yet anaemic. In contrast, measurement of ferritin can detect children with low iron stores. Although ferritin can give false negative results, it does not produce false positive results (personal communication with Paul Monagle).

One study routinely screened for iron deficiency in 361 healthy children aged two months to 18 years of age who were attending a private paediatric clinic for a well-child visit, preschool physical or sports physical. The prevalence of iron deficiency was 3% using zinc protoporphyrin/haem ratio >80 , 28% using transferrin saturation $<16\%$, 4% using ferritin $<10\mu\text{g/L}$, 23% using haematocrit $<35\%$, and 7% using haemoglobin $<11\mu\text{g/L}$. The zinc protoporphyrin/haem ratio was the measure most highly correlated with haemoglobin concentration (the most common measure of iron deficiency). Ferritin concentration and serum transferrin saturation were also significantly correlated with haemoglobin concentration. Zinc protoporphyrin/haem ratio showed a specificity of 97.7% for the detection of iron deficiency, using ferritin levels as the gold standard. It also showed a pattern across age groups consistent with known iron deficiency prevalence patterns (higher mean ratio in children under three years of age and a marked divergence at age 15-18 years with girls showing a significantly higher mean ratio than boys). The downside to using the zinc protoporphyrin/haem ratio as a screening test is that elevated ratios can be caused not only by nutritional iron deficiency, but also anaemia of chronic disease, chronic infections, chronic inflammation or haemoglobinopathies. However the authors point out that all causes of elevated zinc protoporphyrin/haem ratio warrant clinical intervention or monitoring (Rettmer, Carlson, Origenes, Jack, & Labbe 1999).

A study of 321 infants aged 9-18 months attending one of four well-child clinics in Houston, US (servicing low income families) investigated the efficacy of capillary blood haematocrit (haematological test) versus serum ferritin (biochemical test) levels for the screening of iron deficiency. Infants were excluded if they had been unwell in the 2 weeks before the screening, had signs of an infectious disease, were taking oral iron medication or had previously being diagnosed as anaemic. Six infants (2%) were detected using a criterion of haematocrit $\leq 33\%$, 51 (16%) were detected using a criterion of serum ferritin $<10\mu\text{g/L}$, with no infants meeting both criteria and no correlation between the two measures. The authors argue that using the more traditional measure of haematocrit will result in many more undetected cases of iron deficiency than using a measure of serum ferritin (Kazal 1996).

In a study of 210 children (mean age 2.9 years, SD 2.0 years) having a blood test ordered by a primary care paediatrician for lead screening, each child's iron status was also assessed. The gold standard used for iron deficiency was transferrin saturation $<20\%$ and for iron deficiency anaemia was transferrin saturation $<20\%$ plus haemoglobin level $<11\mu\text{g/L}$. Using this standard 43 children (21%) were iron deficient, 24 (12%) of whom were also anaemic. Reticulocyte haemoglobin content (CHr) was found to be a better predictor of iron deficiency and iron deficiency anaemia than iron, transferrin, ferritin and circulating transferrin receptor, with 70% sensitivity and 78% specificity for iron deficiency and 83% sensitivity and 91% specificity iron deficiency anaemia at a threshold of 26pg. The authors discuss the cost implications of these findings for the USA, but this discussion is not relevant to the Australian context. (Brugnara, Zurakowski, DiCanzio, Boyd, & Platt 1999).

Haemoglobinometer

The haemoglobinometer (Hemocue) is an instrument that measures haemoglobin levels in the blood. It is simple to perform and relatively cheap, but tends to produce false negative results as haemoglobin alone is not a definitive measure of iron deficiency (Booth & Aukett 1997).

Diet questionnaire

An alternative approach might be a two stage screening program with the first stage being a parent reported questionnaire or verbal questioning enquiring about the quantity and regularity of foods consumed that contain iron, and foods that promote or limit iron absorption, to identify children at risk of iron deficiency. However, apart from the verbal questioning of parents reported in the program described below (Boutry &

Needlman 1996), we could find only one other measures of this kind which could be used as a first stage screening tool.

A study of 282 children aged 9-30 months in 1997-1998 assessed a parent completed diet history questionnaire as a first stage screening test to identify children at risk for iron deficiency anaemia. The 5 minute diet questionnaire covered four dietary domains (infant diet, beverage intake, solid food intake and participation in Special Supplemental Nutrition Program for Women, Infants and Children (WIC; a program specific to the US)), and three historical domains (birth history, child's health history and maternal/family health history). An affirmative answer on any one question within a domain was considered indicative of risk for iron deficiency anaemia. The gold standard was a venous blood test eliciting complete blood count and ferritin determination. Iron deficiency was defined as ferritin <10µg/L or MCV <70fL and red cell distribution width >14.5% and included children who were iron deficient and anaemic and iron deficient but not anaemic. Anaemia was defined as haemoglobin <11µg/L and included children who were iron deficient anaemic and iron sufficient anaemic. At least one affirmative (at risk) response was indicated on the questionnaire for every child. For the detection of iron deficiency anaemia, the sensitivity and specificity of the maternal/family history domain was 64% and 37% respectively, for the beverage intake domain was 91% and 14%, for the solid food intake domain was 91% and 29%, for the WIC participation domain was 64% and 60% respectively and for the remaining domains was <50% for both sensitivity and specificity. Logistic regression analyses found no group of items to be more sensitive in the identification of iron deficiency anaemia than those proposed by Boutry and Needlman in 1996 (refer programs section) (Bogen, Duggan, Dover, & Wilson 2000).

Apart from the diet questionnaire (which is suggested as a first stage screening tool to be followed by a blood sample for "at risk" children), screening tests for iron deficiency require a blood sample. One of the barriers to iron deficiency screening is the reported reluctance on the part of parents to consent to their child having a blood test. Various studies have reported high refusal rates by parents: 11% (range 0-32% between different clinics) (James, Laing, Logan, & Rossdale 1997) and 21% (Kazal 1996). In addition there are problems with insufficient blood being collected for testing; up to 29% of the samples collected in one study were unable to be analysed due to insufficient blood being collected (James et al. 1997). Even for the first stage screening diet questionnaire, 131 (24%) eligible subjects refused to participate and a further 69 (13%) were missed by recruiters, which may have implications for screening compliance (Bogen et al. 2000).

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✗ ¹
Acceptable to the public	✗
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✓
Test acceptable to the population	✗
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?

¹ simple and quick but prevalences vary widely depending on the measure of iron deficiency used, therefore not easy to interpret

Programs

A program of screening 13-14 month old children for iron deficiency anaemia in a general practice setting was introduced to 5 general practice clinics in the UK: 2 in deprived inner city areas with a high proportion of ethnic minority groups, 1 in a relatively high socioeconomic area and 2 in working class populations. All children attending their general practitioner for a routine MMR immunisation over a 2 year period were invited to participate in the screening program. The general practice nurses took a capillary blood sample from all participating children by finger prick. These samples were sent to a laboratory for analysis of

haemoglobin concentration. The threshold for detection of iron deficiency anaemia was haemoglobin level $<11\mu\text{g/L}$. Over the 5 practices 460 children were invited to participate in the screening program, 11% (range 0-32%) refused, and 5% (range 0-29%) had inadequate blood samples taken so laboratory analysis was not possible. Therefore the screening program reached 85% of children attending for MMR immunisations and approximately 73% of children in the population (as 15% of children are estimated not to attend their general practice for MMR immunisations; one of the deprived inner city practices withdrew from the program after 6 months due to time constraints, parent objections and technical problems with obtaining blood samples). The overall prevalence of iron deficiency anaemia identified in this program was 23% (range 0-35% between the 5 general practices) (James et al. 1997). As the screening test was also the diagnostic "gold standard" it is assumed that all the test results were accurate. The authors concluded that the main problem with this program was its relatively low coverage rate (73%), particularly the high refusal rate (11%).

In 1991-1993 in the US, a screening program for the identification of microcytic anaemia (a form of iron deficiency anaemia characterised by abnormally small red blood cells) was trialed in an urban primary care clinic servicing mainly low socioeconomic African-American families. Children were excluded from the screening program if they had a medical condition known to influence blood iron status. All remaining children attending for their first health supervision visit (between the ages of 13 and 60 months) were included (n=350). Parents were asked to verbally respond to three questions about their child's diet. The questions were:

- 1) Does your child eat meat, cereals or bread, vegetables, and fruit at least five times a week?
- 2) Does he or she drink more than 2 full cups or 16 oz of milk per day?
- 3) Does he or she eat chips, fried snacks, or sweets or drink more than 2 glasses of pop daily?

Children were assessed as having dietary deficiency if their parent responded negatively to question 1 or positively to question 2 or 3. All children then had a blood taken and tested for red cell indices and lead. Children who were assessed as dietary deficient were also tested for ferritin levels. A diagnosis of microcytic anaemia was made if the child's blood tests revealed haemoglobin concentration $<11\mu\text{g/L}$ and MCV $<73\text{fL}$. Children were diagnosed as low haemoglobin if their haemoglobin concentration was $<11\mu\text{g/L}$ and as low ferritin if their ferritin level was <15 microgram/L. Eight percent of children had microcytic anaemia, 12% had low haemoglobin. Using a diagnosis of microcytic anaemia as the reference standard, the diet questioning had a sensitivity of 71%, specificity of 79%, false positive rate of 21%, false negative rate of 29%, PPV of 22% and NPV of 97%. Using a diagnosis of low haemoglobin as the reference standard, the diet questioning had a sensitivity of 55%, specificity of 79%, false positive rate of 21%, false negative rate of 45%, PPV of 28% and NPV of 93% (Boutry & Needlman 1996).

The timing of a screening program to detect children with iron deficiency is an issue. Some children with normal haemoglobin levels at 12-18 months of age may become anaemic by the age of 2 years (Booth & Aukett 1997). Although the peak prevalence of iron deficiency is 18 months, many infants will develop iron deficiency much earlier; however some infants with acceptable iron stores at 15 months will have become iron deficient by 18 months (Wharton 1999).

In the USA, screening for anaemia is recommended as part of routine well child care. Lannon and colleagues surveyed charts of 1413 two year olds and 1396 four year olds attending 50 paediatric practices in North Carolina and Oregon and found a mean coverage of 61% (range 20-90%) across the practices (Lannon, Stuart, Margolis 1999).

Cost effectiveness

We identified no formal cost effectiveness studies of screening programs for iron deficiency anaemia. The cost effectiveness of a program is likely to be affected by the type of screening test used. As Brugnara and colleagues (Brugnara et al. 1999) estimated, a screening test deriving complete blood cell count and evaluation of iron, transferrin, ferritin and erythrocyte zinc protoporphyrin from a blood sample would cost approximately US\$154.33, compared to a screening test deriving only blood cell and reticulocyte counts from a blood sample, costing approximately US\$40.26 based on typical fees. However these estimates may not be generalizable to the Australian context and no estimates of costs in other countries have been published.

Quality of evidence

PAPER	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
James and colleagues	n/a	Cohort study of single diagnostic test, audit of screening program.	n/a	n/a	High non-response rate in some practices may limit generalisability
Bountry and colleagues	n/a	Retrospective review of patient records.	n/a	n/a	Largely African-American sample and restrictive inclusion criteria may limit generalisability to population screening of Australian children.

Two studies of screening programs for iron deficiency were identified. One study (Bountry and colleagues) involved only children from an identified high risk population (low socioeconomic status), while the other study (James and colleagues) involved children from a mix of socioeconomic groups but had a high non-participation rate. Four studies investigating screening tests were identified.

Generalisability of evidence

As the prevalence of iron deficiency is higher amongst children living in lower socioeconomic areas, the results of studies reporting on children from more affluent areas can not be generalized to children from lower socioeconomic areas, and vice versa. The accuracy of test results varies between different laboratories. In addition some of the studies listed as evidence had restrictive inclusion criteria (eg Bountry et al exclude children with an acute illness in the past month, premature birth, chronic illness, failure to thrive, hemoglobinopathies, recent medicinal iron intake or elevated lead) which may limit their generalisability to population screening.

6.3.9.3. CONCLUSIONS

Fair evidence to recommend against screening

Comments:

Iron deficiency does not meet criteria for screening in infants and children since the most appropriate age for screening is unclear, and the benefits are unclear and may be outweighed by the harms. In addition, coverage rates in existing programs and research projects have been poor.

However, there is fair evidence to recommend that health professionals maintain a high level of vigilance for iron deficiency and iron deficiency anaemia in children (particularly among children from low socioeconomic areas), and that iron status be assessed based on risk factors.

Iron deficiency may be more appropriately seen as a public health, rather than an individual, issue.

6.3.9.4. RECOMMENDATIONS

It is recommended that primary health care professionals provide parents with age-appropriate dietary information congruent with the Australian Dietary Guidelines to assist in the prevention of diet-related iron deficiency in infants and children.

Health care professionals working in areas of lower socioeconomic status, where the prevalence of iron deficiency is likely to be higher than average, should be alert to the possibility of iron deficiency in their child patients and assess iron status based on risk factors (uncorrected iron deficiency in the mother during pregnancy, prematurity, age <24 months, introduction of cow's milk as the main milk source before 12 months of age, consumption of >600ml of cow's milk per day).

Children living in areas prone to lead toxicity and (as per current NHMRC guidelines) children with blood lead levels >0.72µmol/L should be tested for iron deficiency.

6.3.9.5. FURTHER RESEARCH

We recommend population-based surveys of Australian children approximately every five years. Surveys should use appropriate sampling strategies to access both low and high risk children, and should be designed to inform public health nutritional policy for young children.

Public health studies into effective strategies to promote optimal dietary intake of iron among high risk and lower risk children. Outcomes should include (1) iron status and (2) developmental and behavioural outcomes.

Research into simple parent-report diet questionnaires that might be used to reliably identify children with poor dietary patterns, who are likely to be at greater risk for iron deficiency.

6.3.9.6. SEARCH TERMS

Anemia, iron deficiency

Iron metabolism disorders

Anemia

Anemia, neonatal

6.3.9.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Kay Gibbons – Head, Nutrition and Food Services Department, Royal Children’s Hospital, Melbourne.

Paul Monagle – Haemopathology Consultant, Royal Children’s Hospital, Melbourne.

Garth Alperstein – Reference Group

Lyndall Whitecross – Reference Group

6.3.9.8. REFERENCES

Australian Iron Status Advisory Panel. Iron deficiency in children; a comprehensive management guide.

Australian Iron Status Advisory Panel. <http://www.ironpanel.org.au/AIS/AISdocs/childdocs/Ccontents.html>, last updated 10-5-2000.

Bogen, D. L., Duggan, A. K., Dover, G., & Wilson, M. H. 2000, "Screening for Iron Deficiency Anemia by Dietary History in a High Risk Population", *Pediatrics*, vol. 105, no. 6, pp. 1254-1259.

Booth, I. W. & Aukett, M. A. 1997, "Iron deficiency anaemia in infancy and early childhood.", *Archives of Disease in Childhood*, vol. 76, no. 6, pp. 549-553.

Boutry, M. & Needlman, R. 1996, "Use of Diet History in Screening of Iron Deficiency", *Pediatrics*, vol. 98, no. 6, pp. 1138-1142.

Brugnara, C., Zurakowski, D., DiCanzio, J., Boyd, T., & Platt, O. 1999, "Reticulocyte Hemoglobin Content to Diagnose Iron Deficiency in Children", *Journal of the American Medical Association*, vol. 281, pp. 2225-2230.

Choi, Y. S. & Reid, T. 1998, "Anemia and red cell distribution width at the 12-month well-baby examination", *South. Med J*, vol. 91, no. 4, pp. 372-374.

Edmond, A. M., Hawkins, N., Pennock, C., & Golding, J. 1996, "Haemoglobin and ferritin concentrations in infants at 8 months of age", *Archives of Disease in Childhood*, vol. 74, no. 1, pp. 36-39.

James, J. A., Laing, G. J., Logan, S., & Rosedale, M. 1997, "Feasibility of screening toddlers for iron deficiency anaemia in general practice", *British Medical Journal*, vol. 315, pp. 102-103.

Karr, M., Alperstein, G., Causer, J., Mira, M., Lammi, A., & Fett, M. J. 1996, "Iron status and anaemia in preschool children in Sydney", *Australian and New Zealand Journal of Public Health*, vol. 20, no. 6, pp. 618-622.

Kazal, L.A. 1996, "Failure of Hematocrit to Detect Iron Deficiency in Infants", *The Journal of Family Practice*, vol. 42, no. 3, pp. 237-240.

Lannon, C., Stuart, J., & Margolis, P. 1999, "Mirror, mirror, on the wall; what's the best quality measure of them all?", Workshop proceedings from the 1999 Ambulatory Pediatric Association Meeting, San Francisco.

Logan, S. 1999, "Commentary: Iron deficiency and developmental deficit - the jury is still out", *British Medical Journal*, vol. 318, pp. 697-698.

Looker, A. C., Dallman, P. R., Carroll, M. D., Gunter, E. W., & Johnson, C. L. 1997, "Prevalence of Iron Deficiency in the United States", *Journal of the American Medical Association*, vol. 277, no. 12, pp. 973-976.

Lozoff, B., Jimenez, E., Hagen, J., Mollen, E., & Wolf, A. W. 2000, "Poorer Behavioral and Developmental Outcome More than 10 Years After Treatment for Iron Deficiency in Infancy", *Pediatrics*, vol. 105, no. 4, pp. 1-11.

Martins, S., Logan, S., & Gilbert, R. 2000, "Iron therapy and psychomotor development and cognitive function in children under the age of three with iron deficiency anaemia [protocol]", *The Cochrane Library (Oxford)*.

Moffatt, M. E., Longstaffè, S., Besant, J., & Dureski, C. 1994, "Prevention of iron deficiency and psychomotor decline in high-risk infants through use of iron-fortified infant formula: a randomized clinical trial.", *Journal of Pediatrics*, vol. 125, no. 4, pp. 527-534.

Morley, R., Abbott, R., Fairweather-Tait, S., MacFadyen, U., Stephenson, T., & Lucas, A. 1999, "Iron fortified follow on formula from 9 to 18 months improves iron status but not development or growth: a randomised trial", *Archives of Disease in Childhood*, vol. 81, pp. 247-252.

- National Health and Medical Research Council 1996, *Australian Dietary Guidelines for Children and Adolescents* Commonwealth of Australia.
- Oski, F. A. 1993, "Iron deficiency in infancy and childhood", *N.Engl J Med*, vol. 329,no. 3, pp. 190-193.
- Oti-Boateng, P., Seshadri, R., Petrick, S., Gibson, R. A., & Simmer, K. 1998, "Iron status and dietary iron intake of 6-24-month-old children in Adelaide", *Journal of Paediatrics & Child Health*, vol. 34, pp. 250-253.
- Roeser, H. P. 1992, "Eat foods containing iron." In National Health and Medical Research Council, *Dietary Guidelines for Australians*. Commonwealth of Australia.
- Rettmer, R. L., Carlson, T. H., Origenes, M. L., Jack, R. M., & Labbe, R. F. 1999, "Zinc Protoporphyrin/Heme Ratio for Diagnosis of Preanemic Iron Deficiency", *Pediatrics*, vol. 104,no. 3, pp. 1-5.
- U.S. Department of Health and Human Services Public Health Service Centres for Disease Control and Prevention. 1998, "Recommendations to Prevent and Control Iron Deficiency in the United States", *Morbidity & Mortality Weekly Report*, vol. 47,no. RR-3 , pp. 1-36.
- Wharton, B. A. 1999, "Iron Deficiency in Children: Detection and Prevention", *British Journal of Haematology*, vol. 106, pp. 270-280.

6.3.10. LEAD TOXICITY

6.3.10.1 BACKGROUND

Clinical picture

Lead toxicity or lead poisoning is the most common disease caused by an environmental substance. Individuals of any age can suffer lead toxicity but it is most harmful to children under 6 and particularly those under 3 years of age. Young children are at particular risk from lead exposure because the developing brain appears more vulnerable to a range of biological and environmental insults, including lead; their tendency to engage in hand-to-mouth activity and to put things in their mouth increases their exposure to lead from contaminated environments; and because children absorb a much higher proportion of ingested lead than adults (up to 50%) (Alperstein, Taylor, & Vimpani 1994).

Lead is a naturally occurring metal and is also used for industrial and domestic purposes (Donovan 1996). Lead from leaded petrol, lead-based paint and other lead-based products can be found in the air, dust and soil (Bodenhorn 1991). Increasing government regulation of lead based products (such as the recent phasing out of leaded petrol in Australia) has lessened the risk of lead exposure over recent decades. Lead can sometimes be a problem in drinking water if there is a considerable amount of lead in the pipes and the water itself is slightly acid, and in food supplied in cans, although lead soldered cans are rare in Australia (Donovan 1996). Most children are exposed to these contaminants to some extent, although children living in lower socioeconomic areas tend to have greater exposure. The development and functioning of almost all body systems are affected by the presence of lead in the body which can persist long beyond exposure due to the long half life of lead (Bodenhorn 1991). One of the most serious effects is on the developing nervous system, although lead also affects the blood, kidneys and other organs (United States Preventive Services Task Force 1996). Lead toxicity often coexists with iron deficiency which can enhance lead absorption (United States Department of Health and Human Services 1991).

The NHMRC iterates that human exposure to lead has no benefits and all demonstrated effects of lead exposure are adverse (NHMRC 1993). The higher the level of lead to which a child is exposed, the higher the likelihood of toxicity. This relationship is difficult to quantify in a one-off blood test, which indicates only the current level and not exposure over the longer term. Long term cohort studies which quantify exposure over time have indicated that lead exerts a dose-response effect on children's intelligence and a range of behavioural and psychosocial outcomes including risk of antisocial and delinquent behaviour, even after controlling for factors such as social status and home environment. For example, the Port Pirie Cohort Study reported that the IQ of children with an average blood lead concentration of 30µg/dL over the first four years of life was 4.4 points lower than children with an average blood lead concentration of 10µg/dL (Baghurst, McMichael, Wigg, Vimpani, Robertson, Roberts, & Tong 1992). However this equates to only a small effect size. The variance in IQ attributable to lead exposure has been reported to be smaller than that attributable to social factors (Wasserman, Staghezza-Jaramillo, Shrout, Popovac, & Graziano 1998), though this study did not report levels of parent lead exposure as a possible confounder. A meta analysis from a systematic review looking at the relationship between lead and IQ in children 5 years of age or older found an overall deficit of 1-2 IQ points associated with an increase in blood lead from 10 to 20µg/dL (Pocock, Smith & Baghurst 1994). A recent study using data on 6-16 year olds from the National Health and Nutrition Examination Survey (NHANES) III (1988-1994) found an inverse relationship between blood lead concentration (down to <2.5µg/dL) and cognitive function scores on measures of reading, maths, short-term memory and visual-constructional skills (Lanphear, Dietrich, Auinger, & Cox 2000). A case-control study of youths arrested and judged delinquent found delinquent children had significantly higher bone lead levels than controls (13.7ppm (95% confidence interval 18.5, 8.9) versus 2.0ppm (95% confidence interval 6.8, -2.0)). The authors concluded that early lead exposure is a risk factor for later delinquent behaviour (Needleman, McFarland, Ness, Tobin, & Greenhouse 2000). Cognitive changes resulting from lead exposure may be only minimally reversible (Tong, Baghurst, Sawyer, Burns, & McMichael 1998).

Natural History

Children with mild to moderate lead exposure (<45µg/dL) are usually asymptomatic (Alperstein et al 1994). When symptoms do appear they are usually non-specific, but may include malaise, lethargy, irritability and sometimes abdominal pain (Needleman 1988). More pronounced lead toxicity may produce mood changes, slightly lowered IQ or anaemia. However, these effects can be very subtle and difficult to recognise. Extreme lead toxicity can cause mental retardation and lead encephalopathy which leads to convulsions, coma and even death, but such high levels of lead toxicity are extremely rare (Canadian Task Force on the Periodic Health Examination 1994).

Prevalence

The prevalence of lead toxicity has declined sharply over the preceding decades as governments have regulated products containing lead with a resultant lowering in exposure. Whereas in 1991 there were estimated to still be 3-4 million children (approximately 1 in 6) affected by lead toxicity in the United States (Bodenhorn 1991) this has now fallen to in excess of one million. CDC estimated from NHANES II and III data that the proportion of children aged 1-5 years with lead levels ≥ 10 had fallen to 4.4% by 1991-4, down from 8.9% for 1988-1991 and 88.2% for 1976-1980. In 1991-2 a cross sectional study of 636 12-60 month old children attending well-care checkups in the northern California region found 4% to have blood lead levels of $10\mu\text{g/dL}$ or more (Haan, Gerson, Zishka, & Anne 1996).

The most recent national study of blood lead levels of Australian children was conducted in February and March 1995 by the Australian Institute of Health and Welfare. It reported 7.3% (115/1575) of 1-4 year old children to have blood lead levels $\geq 10\mu\text{g/dL}$. Of these 27 (1.7%) had blood lead levels of $15\text{-}24\mu\text{g/dL}$ and 4 (0.3%) had levels of $\geq 25\mu\text{g/dL}$ with $32.7\mu\text{g/dL}$ being the highest level recorded. The proportion of children with blood lead levels $\geq 10\mu\text{g/dL}$ differed by state: 4.6% (n=3/65) for the Australian Capital Territory, 4.8% (n=13/270) for Queensland, 6.1% (n=21/345) for Victoria, 7.5% (n=10/133) for South Australia, 7.8% (n=10/129) for Tasmania, 8.4% (n=36/431) for New South Wales, 10.5% (n=19/181) for Western Australia, and 14.3% (n=3/21) for Northern Territory. There were similar proportions of males (8.3%) and females (6.3%) with blood lead levels $\geq 10\mu\text{g/dL}$. Mean blood lead level was highest at 15-17 months of age. The proportion of children with blood lead levels $\geq 10\mu\text{g/dL}$ decreased with age: 10.2% for 1 year olds, 9.9% for 2 year olds, 6.4% for 3 year olds and 2.7% for 4 year olds. More Aboriginal and Torres Strait Islander children had high blood lead levels ($\geq 10\mu\text{g/dL}$) than other Australian children (18.5% vs 7.2%) (Donovan 1996).

A survey of 726 (76% response) children in central and southern Sydney in 1992-1994 found 16% of children to have blood levels ≥ 10 , 3.9% ≥ 15 , and 0.3% $\geq 25\mu\text{g/dL}$. For children aged 9 months-5 years living within 10km of the Sydney CBD, 25% had blood lead levels ≥ 10 and 7% $\geq 15\mu\text{g/dL}$ (Mira, Bawden-Smith, Causer, Alperstein, Karr, Snitch, Waller, & Fett 1996). In a 1995 study of 120 children from day-care centres and 44 hospital inpatients in Fremantle, WA, 26% had lead levels $\geq 10\mu\text{g/dL}$ (Willis, Rossi, Bulsara, & Slattery 1995). In a 1999 study of 22 children of lead industry workers and 15 non-exposed children, all children were found to have blood lead levels $< 15\mu\text{g/dL}$, although the average blood lead level of exposed children was higher than that of non-exposed children. One of the main problems encountered by this study was the high non participation rate (personal communication with Vikki Lynch, Victorian Department of Human Services).

Genetics relevant to screening issues

None known

Diagnosis

Lead toxicity is diagnosed by a blood test. Levels of blood lead considered dangerous have varied over the past few decades. The Centres for Disease Control and Prevention in 1991 stated that levels over $10\mu\text{g/dL}$ should be considered dangerous in children (United States Department of Health and Human Services 1991) and most countries now use this level as the standard threshold for diagnosis of lead toxicity. In Australia, the National Health and Medical Research Council's target is for all Australians to have a blood lead level below this threshold (National Health and Medical Research Council 1993).

With increased awareness of the need to limit the contamination of the environment during the refining of lead and the production of lead-based products severe cases of lead toxicity have been greatly reduced, but awareness of the adverse effects of even low levels of lead has increased. Now, most cases of lead exposure are milder and occur without obvious symptoms, so that diagnosis is rarely made on clinical ground. However, once a diagnosis is made, parents often report a history of mood or behaviour change and psychological testing often uncovers marginally below average IQ scores and behavioural problems (Needleman 1988).

Treatment/Management

Currently NHMRC recommends action when blood lead levels exceed $15\mu\text{g/dL}$ (National Health and Medical Research Council 1993). Based on the prevalence figures supplied in the 1995 Australian Institute of Health and Welfare report, less than 2% of children 1-4 years of age are likely to have levels $\geq 15\mu\text{g/dL}$ (Donovan 1996), and the prevalence may have fallen further since then.

NHMRC recommends graded management responses based on the blood lead level of the child. For children with blood lead levels of $15\text{-}24\mu\text{g/dL}$ ($0.72\text{-}1.16\mu\text{mol/L}$) it recommends an evaluation of exposure, parent and child education on exposure control, and repeat testing as appropriate. For children with blood lead levels of $\geq 25\mu\text{g/dL}$ ($\geq 1.20\mu\text{mol/L}$) it recommends a detailed medical history and examination, evaluation of exposure, arrangement of remediation/abatement of exposure source, parent and child education on exposure control, relocation if exposure control not possible, and repeat testing in three months. In children with blood

lead levels of $\geq 55 \mu\text{g/dL}$ ($2.65 \mu\text{mol/L}$) it recommends in addition to the response for children $\geq 25 \mu\text{g/dL}$, that children undergo urgent clinical assessment regarding immediate medical management (National Health and Medical Research Council 1993).

Severe lead toxicity ($\geq 70 \mu\text{g/dL}$) is likely to be accompanied by seizures, vomiting and/or altered state of consciousness. In such severe cases children require hospitalisation. For children with such dangerously high blood lead levels, chelation therapy is the recommended treatment to eliminate the lead (United States Department of Health and Human Services 1991). However, such severe cases are extremely rare, with the most recent Australian survey finding $32.7 \mu\text{g/dL}$ to be the highest recorded blood lead level (Donovan 1996).

Treatment for iron deficiency (which often accompanies and contributes to lead toxicity) is important where present, but should not be commenced until after chelation in children who are receiving dimercaprol (BAL) (United States Department of Health and Human Services 1991). Psychosocial, language and other problems require referral to appropriate agencies. Children identified with high lead exposure should have regular blood lead tests to monitor their blood lead levels.

Prevention

Undernourished individuals (particularly those with iron deficiency) absorb more lead than those who are well nourished, therefore good nutrition is one strategy that can be used in preventing lead toxicity; in particular, ensuring children consume adequate amounts of calcium, iron and zinc and limited amounts of fat (Alperstein & Vimpani 1994). Other strategies include reducing children's exposure to lead and lead-based products through the elimination of leaded petrol, removing children from homes containing leaded paint whilst renovations are occurring, washing fruits and vegetables before consumption, regular washing of pets, and where children live in areas of high soil lead, paving or planting grass over exposed dirt, regular hand washing (particularly prior to eating) and wet mopping of the home (Alperstein & Vimpani 1994).

How might screening reduce the burden of suffering?

Screening could allow identification of children with elevated blood lead levels who may otherwise go undetected. Probably more importantly, it could alert authorities to areas of high lead toxicity allowing preventive and environmental strategies to be developed and implemented where needed.

6.3.10.2 EVIDENCE

Tests

Erythrocyte protoporphyrin test

The most widely used test for lead toxicity screening and diagnosis from the late 1970s to the early 1990s, this test is now not recommended for use. Although cheap and simple to perform, requiring only a small amount of blood from a fingerstick, not prone to contamination and able to be analysed by non technical personnel using a portable hematofluorometer, the erythrocyte protoporphyrin test required a confirmatory blood lead test. The sensitivity of the erythrocyte protoporphyrin test to detect blood lead levels at $10 \mu\text{g/dL}$ or above is less than 50% (Parsons, Reilly, & Esemio-Jenssen 1997).

Blood lead test

The blood lead test is recommended by Centres for Disease Control and Prevention for screening and diagnosis of lead toxicity (United States Department of Health and Human Services 1991). Capillary (by fingerstick) or venous (by venipuncture) blood can be used for blood lead tests. Capillary blood is more prone to contamination by lead on the skin than venous blood, leading to a higher incidence of false positive results, but is less invasive for population screening. With a screening threshold of $10 \mu\text{g/dL}$, using venous blood as the "gold standard", one study found the capillary blood lead test to have a sensitivity of 100%, specificity of 93% (95% confidence interval 91-96%), total accuracy of 95% (95% confidence interval 94-97%), false positive rate (total number of false positives divided by total number of positives, multiplied by 100) of 13% and false positive proportion (total number of false positives divided by total number screened, multiplied by 100) of 5% (Parsons, Reilly, & Esemio-Jenssen 1997). A community-based study collected paired samples of capillary blood on filter paper and venous blood as the standard from 120 children aged 6 months to 6 years (97% African American) in a high risk area. With a screening threshold of $10 \mu\text{g/dL}$, the capillary blood test had a sensitivity of 94%, specificity of 99%, positive predictive value of 97% and negative predictive value of 98% (Holtrop, Yee, Simpson, & Kauffman 1998). Since PPV varies with prevalence, it would probably be lower in low risk areas.

Graphite furnace atomic absorption spectrometry (GFAAS) is one validated method of analysing blood samples for lead levels. This method is usually accurate to at least $1.5 \mu\text{g/dL}$ up to $40 \mu\text{g/dL}$ (Parsons, Reilly, & Esemio-Jenssen 1997) however the equipment required to perform GFAAS is expensive (Verebey, Rosen,

Schonfeld, Carriero, Eng, Deutsch, Reimer, & Hogan 1995). Second generation filter paper collection-based Delves cup-flame atomic absorption spectrophotometry (2nd-Gen FPDC) is a cheaper method of analysing blood lead samples and has been shown to correlate highly ($r=0.98$) with GFAAS results (Verebey et al. 1995). 2nd-Gen FPDC has the additional advantages of requiring only a couple of drops of blood from a finger prick making it simpler to collect from children, blood samples on filter paper remain stable for ≥ 6 months compared to 4-5 days for venous and capillary blood samples and are simpler to transport and process (Verebey et al. 1995). As with all tests requiring laboratory analysis, the accuracy of blood lead measurement differs between laboratories (Sargent & Dalton 1996). Venous blood sampling is the commonest and most reliable technique in clinical practice in Australia.

Centre for Disease Control and Prevention (CDC) risk assessment questions

Risk assessment questions have been proposed as a first stage screening test to identify children at risk for lead toxicity who should receive blood lead screening. The Centre for Disease Control and Prevention (CDC) recommends three screening questions identify children at risk for lead toxicity:

1. *Does your child live in or regularly visit a house that was built before 1950?* This question could apply to a facility such as a home day care centre, or the home of a babysitter or a relative.
2. *Does your child live in or regularly visit a house built before 1978 with recent or ongoing renovations or re-modelling (within the last 6 months)?*
3. *Does your child have a sibling or playmate who has or did have lead poisoning?* (Centre for Disease Control and Prevention 1997)

We identified no studies assessing these questions as screening tools. However, Question 3 is almost identical to one of the five 1991 CDC questions (United States Department of Health and Human Services 1991) that this set of three questions replaces, which had received some assessment. One study investigating the use of the earlier questions in children attending a well-care checkup found Question 3 to have a sensitivity of 28%, specificity of 81%, PPV of 6%, and NPV of 97% compared to the “true” diagnosis of venous blood test $\geq 10\mu\text{g/dL}$ (Haan et al. 1996). Another study using logistic regression found the odds of elevated blood lead levels associated with an affirmative answer to Question 3 was 2.7 (95% confidence interval 1.7, 4.2) (Dalton, Sargent, & Stukel 1996).

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✗
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	?
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✗
Suitable test or examination	✓
Test acceptable to the population	✗
Natural history adequately understood	✗
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

Most lead screening reports are from the US, which has a formal policy that all children should be screened, usually in the community setting by their primary health care professional. Community-based screening typically relies on capillary blood lead tests as the initial screening tool (with or without questioning of parents using the CDC questions or equivalent). Children who exceed the threshold for abnormal screening results (typically set at $10\mu\text{g/dL}$) are called back for a confirmatory venous blood lead tests.

The Centres for Disease Control and Prevention (CDC) recommends states develop their own screening programs based on local data. These may be universal, as recommended in the 1991 guidelines (all children aged 12-72 months should be screened regularly), or targeted. The CDC provides an example of a targeted screening program: screening all children for risk of lead toxicity at ages 1 and 2 years (and children aged 32-72 months who have never previously been screened). Those children found to be at high risk for lead toxicity should then be screened with a blood lead test. High risk children are considered those who live in a postcode area that has been identified as having $\geq 27\%$ of housing built prior to 1950 and children who receive services for the poor. In addition, children whose parents answer "yes" to any one of three suggested questions (described above) should be considered at high risk of lead exposure and undergo blood lead screening (Centre for Disease Control and Prevention 1997).

The effectiveness of this approach has not been evaluated. Kemper carried out a theoretical comparison of blood lead screening strategies in light of the 1997 CDC recommendations (Kemper, Bordley, & Downs 1998). Assuming universal uptake, universal venous screening was estimated to detect 100% of cases (ie lead levels $\geq 10\mu\text{g/dL}$), universal capillary screening 93-5%, targeted venous screening 77%, and targeted capillary screening 73% of cases. In the US, screening for lead toxicity is recommended as a part of routine well child care. However Lannon and colleagues surveyed charts of 1413 two year olds and 1396 four year olds attending 50 paediatric practices in North Carolina and Oregon, and found a mean coverage rate of 29% (range 0-90%) across the practices (Lannon, Stuart & Margolis 1999). Therefore Kemper's estimates are probably useless in determining effectiveness of the US program.

Australia has no screening program for lead toxicity. In 1993, NHMRC set a goal of a reduction of lead in all Australians to less than $15\mu\text{g/dL}$ by the end of 1998. It is not known whether this goal has been achieved. The 1993 NHMRC Australian guidelines for lead in blood and lead in ambient air called for representative surveys in communities known to be at high risk of lead toxicity, at frequencies to be determined by local authorities. Environmental strategies were recommended if high risk communities were found to have $\geq 95\%$ of children aged 1-4 years with lead levels < 25 but $> 5\%$ $\geq 15\mu\text{g/dL}$. In communities with $> 5\%$ of children with lead levels $> 25\mu\text{g/dL}$, more urgent environmental strategies would be required (National Health and Medical Research Council 1993). These strategies have not been evaluated.

Cost effectiveness

Two cost-effectiveness analyses have reached similar conclusions. Fingertick blood tests (if contamination is minimised) are cheaper than venipuncture blood tests to screen children for lead toxicity where the prevalence is low, however as the prevalence increases ($\geq 40\%$) this cost difference diminishes (Campbell, Paris, & Schaffer 1996). Targeted programs which use a risk assessment (based on area or individual children) to determine which children to screen are cheaper than universal programs which screen all children. However the cost-effectiveness of targeted programs which are more likely to miss cases has not been fully assessed (Campbell, Paris, & Schaffer 1996). Kemper and colleagues (Kemper, Bordley, & Downs 1998) compared blood lead screening strategies in light of 1997 CDC recommendations. Taking into account costs of follow up testing and treatment, in high prevalence populations universal venous screening minimised the cost per case (US\$490), while in low- and medium-prevalence populations, targeted screening using venous testing minimised the cost per case (US\$729 and \$556, respectively.) In all populations, venous testing resulted in a lower cost per case than capillary testing.

Quality of evidence

PAPER	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
Kemper and colleagues	n/a	theoretical comparison of program effectiveness in different prevalence groups	n/a	n/a	This was a theoretical study which did not incorporate consideration of the practical realities of screening. Therefore the evidence it provides is not likely to be relevant to a real-life situation.

The only study of screening programs for lead toxicity in children that was identified was a theoretical comparison of sensitivity, specificity and cost effectiveness of four different screening program strategies. Four studies reporting on screening tests for lead toxicity were also identified.

Generalisability of evidence

Most of the available evidence and recommended screening schedules come from the US and may not be relevant to the Australian setting. In particular, the risk factors between Australia and the US may differ (eg main sources of exposure, year in which lead free paint was introduced), in which case the risk assessment questions may not be generalizable.

6.3.10.3. CONCLUSIONS

Fair evidence to recommend against screening

Comments:

Lead toxicity does not meet criteria for screening in infants and children because it is poorly accepted by professionals and parents and lead exposure is low (and falling) for the great majority of Australian children. There is no information about the effectiveness of screening programs.

Lead toxicity may be more appropriately seen as a public health, rather than an individual, issue.

6.3.10.4. RECOMMENDATIONS

Implementation of screening programs to detect lead toxicity in well children is not recommended.

We recommend five yearly surveys of representative samples of children, particularly those aged 1-4 years, in high and low lead exposure areas to document the prevalence of lead toxicity. We note that a similar recommendation was made in the 1993 NHMRC Review of Child Health Surveillance and Screening.

Health care professionals working in areas where the prevalence of lead toxicity is likely to be higher than average should be alert to the possibility of lead toxicity in their child patients.

Public health preventive and management strategies are likely to be most effective in further reducing childhood lead levels.

6.3.10.5. FURTHER RESEARCH

None suggested.

6.3.10.6. SEARCH TERMS

Lead poisoning

Lead toxicity

6.3.10.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Paul Monagle – Haemopathology Consultant, Royal Children's Hospital, Melbourne.

Peter Baghurst – Women's and Children's Hospital, Adelaide

Meredith Carter – Reference Group

Paul Dworkin – Reference Group

6.3.10.8. REFERENCES

Alperstein, G., Taylor R., & Vimpani, G. 1994, *Lead alert – a guide for health professionals*, Commonwealth Environment Protection Agency, Australia.

Alperstein, G., & Vimpani, G. V. 1994, "Reducing the lead load in childhood", *Journal of Paediatrics and Child Health*, vol. 30, no. 1, pp. 3-5.

Baghurst, P. A., McMichael, A. J., Wigg, N. R., Vimpani, G. V., Robertson, E. F., Roberts, R. J., & Tong, S. L. 1992, "Environmental exposure to lead and children's intelligence at the age of seven years. The Port Pirie Cohort Study", *N.Engl J Med*, vol. 327, no. 18, pp. 1279-1284.

Bodenhorn, K. A. 1991, "Lead poisoning: the foremost preventable disease of childhood", *Journal of Pediatric Health Care*, vol. 5, no. 3, pp. 156-157.

- Campbell, J. R., Paris, M., & Schaffer, S. J. 1996, "A comparison of screening strategies for elevated blood lead levels", *Archives of Pediatrics & Adolescent Medicine*, vol. 150, no. 11, pp. 1205-1208.
- Canadian Task Force on the Periodic Health Examination 1994, *The Canadian Guide to Clinical Preventive Health Care*, Canadian Government Publishing, Ottawa.
- Centre for Disease Control and Prevention 1997, *Screening Young Children for Lead Poisoning: Guidance for State and Local Public Health Officials*, CDC, Atlanta.
- Dalton, M. A., Sargent, J. D., & Stukel, T. A. 1996, "Utility of a Risk Assessment Questionnaire in Identifying Children with Lead Exposure", *Archives of Pediatrics & Adolescent Medicine*, vol. 150, no. 2, pp. 197-202.
- Donovan, J. 1996, *Lead in Australian Children. Report on the National Survey of Lead in Children* Australian Institute of Health and Welfare, Canberra.
- Haan, M., Gerson, M., Zishka, B., & Anne, B. A. 1996, "Identification of Children at Risk for Lead Poisoning: An Evaluation of Routine Pediatric Blood Lead Screening in an HMO-Insured Population", *Pediatrics*, vol. 97, no. 1, pp. 79-83.
- Holtrop, T. G., Yee, H. Y., Simpson, P. M., & Kauffman, R. E. 1998, "A Community Outreach Lead Screening Program Using Capillary Blood Collected on Filter Paper", *Archives of Pediatrics & Adolescent Medicine*, vol. 152, no. 5, pp. 445-458.
- Kemper, A. R., Bordley, W. C., & Downs, S. M. 1998, "Cost-effectiveness analysis of lead poisoning screening strategies following the 1997 guidelines of the Centers for Disease Control and Prevention", *Archives of Pediatrics & Adolescent Medicine*, vol. 152, no. 12, pp. 1202-1208.
- Lannon, C., Stuart, J., & Margolis, P. 1999, "Mirror, mirror, on the wall; what's the best quality measure of them all?", Workshop proceedings from the 1999 Ambulatory Pediatric Association Meeting, San Francisco.
- Lanphear, B. P., Dietrich, K., Auinger, P., & Cox, C. 2000, "Subclinical lead toxicity in U.S. children and adolescents", Proceedings of the 2000 Pediatric Academic Societies and American Academy of Pediatrics Joint Meeting, Boston.
- Mira, M., Bawden-Smith, J., Causer, J., Alperstein, G., Karr, M., Snitch, P., Waller, G., & Fett, M. J. 1996, "Blood lead concentrations of preschool children in central and southern Sydney", *Med J Aust*, vol. 164, no. 7, pp. 399-402.
- National Health and Medical Research Council. 1993, "Revision of the Australian guidelines for lead in blood and lead in ambient air (extract from the 115th Session of Council, June 1993)", National Health and Medical Research Council, Canberra.
- Needleman, H. L. 1988, "The persistent threat of lead: medical and sociological issues", *Current Problems in Pediatrics*, vol. 18, no. 12, pp. 697-744.
- Needleman, H. L., McFarland, C., Ness, R., Tobin, M., & Greenhouse, J. 2000, "Bone lead levels in adjudicated delinquents: a case-control study", Proceedings of the 2000 Pediatric Academic Societies and American Academy of Pediatrics Joint Meeting, Boston.
- Parsons, P. J., Reilly, A. A., & Esernio-Jenssen, E. 1997, "Screening children exposed to lead: an assessment of the capillary blood lead fingerstick test", *Clinical Chemistry*, vol. 43, no. 2, pp. 302-311.
- Pocock, S. J., Smith, M., & Baghurst P. 1994, "Environmental lead and children's intelligence: a systematic review of the epidemiological evidence", *British Medical Journal*, vol. 309, pp. 1189-1197.
- Sargent, J. D. & Dalton, M. A. 1996, "Rethinking the Threshold for an Abnormal Capillary Blood Lead Screening Test", *Archives of Pediatrics & Adolescent Medicine*, vol. 150, no. 10, pp. 1084-1088.
- Tong, S., Baghurst, P. A., Sawyer, M. G., Burns, J., & McMichael, A. J. 1998, "Declining blood lead levels and changes in cognitive function during childhood: the Port Pirie Cohort Study", *JAMA*, vol. 280, no. 22, pp. 1915-1919.
- U.S. Department of Health and Human Services Public Health Service Centers for Disease Control. 1991, *Preventing lead poisoning in young children*.
- U.S. Preventive Services Task Force *Guide to Clinical Preventive Services*, 2 edn, US Department of Health and Human Services, Office of Public Health and Science, Office of Disease Prevention and Health Promotion.

Verebey, K., Rosen, J. F., Schonfeld, D. J., Carriero, D., Eng, Y. M., Deutsch, J., Reimer, S., & Hogan, J. 1995, "Blood collection and analytical considerations in blood lead screening in children", *Clinical Chemistry*, vol. 41, pp. 469-470.

Wasserman, G. A., Staghezza-Jaramillo, B., Shrout, P., Popovac, D., & Graziano, J. 1998, "The effect of lead exposure on behavior problems in preschool children", *Am.J Public Health*, vol. 88,no. 3, pp. 481-486.

Willis, F. R., Rossi, E., Bulsara, M., & Slattery, M. J. 1995, "The Fremantle Lead Study", *J Paediatr Child Health*, vol. 31,no. 4, pp. 326-331.

6.3.11. PHENYLKETONURIA (PKU)

6.3.11.1 BACKGROUND

Clinical picture

Phenylketonuria (PKU) is a congenital disease inherited as an autosomal recessive trait. The condition is caused by the absence or near absence of phenylalanine hydroxylase (PAH), the enzyme that converts phenylalanine (present in all dietary protein) into tyrosine in the liver. There are approximately 70 mutations that cause PKU (see genetic defect section) (Anonymous 1999). Disease severity varies according to mutation and is negatively correlated with the level of residual phenylalanine hydroxylase activity.

Individuals with PKU have an accumulation of phenylalanine in the blood exceeding 1200 μ mol/l (20mg/dL) on a normal diet. The level of phenylalanine in the blood of infants with PKU is usually normal at birth, rapidly rising thereafter.

Untreated, PKU leads to serious impairment in early neuronal development in all affected individuals. The most common effect of untreated PKU is severe mental retardation, with hyperactivity, spasticity, seizures, tremors, gait disorders, eczema and autistic-like behaviour also characteristic. Affected infants may show few symptoms in the early weeks of life but by 6-12 months delayed mental development is usually evident.

Early diagnosis is imperative to prevent impairment. In the US healthy infants are usually screened 1-3 days after birth. Premature or unwell infants are screened in the first week of life. In the UK screening usually occurs 6-11 days after birth. In Australia babies are screened in the first few days of life. If screening occurs before 48 hours of life (often necessitated by early release from hospital), there is slightly greater likelihood of false negative results; therefore a second screen is performed at the first home visit (usually within the first week of life).

Natural History

At birth, the level of phenylalanine in the blood of infants affected with PKU is usually normal. Immediately after birth the level of phenylalanine in the blood rises rapidly. If left untreated the phenylalanine level will remain high, impairing early neuronal development.

Prevalence

The highest incidence of PKU is in Caucasians, with 1 in 10,000-15,000 newborns affected. It is estimated that approximately 1 in 60 Caucasians carry one copy of the defective gene. In Australia, the incidence is approximately 1 in 10,000-14,000 with approximately 20-25 newborns diagnosed with the condition each year (Dietitians Working Party of the Australasian Society for Inborn Errors of Metabolism 1996). PKU also occurs in non-Caucasian newborns; the incidence in Northern China is similar to that in Europe, but the incidence is much lower in other non-Caucasian groups.

Genetics relevant to screening issues

The gene for phenylalanine is located on chromosome 12. Approximately 70 mutations have been identified to date, of which four are common. The most frequently occurring mutation is IVS12DS, G-A, +1 which is 52 amino acids shorter than normal (Marvit and colleagues 1987 as cited in Anonymous 1999) and causes classic PKU. The other common mutations are Arg408 to Trp which also causes classic PKU, Arg 158 to Gln which causes mild PKU and Arg261 to Gln which causes a variation on mild PKU known as benign hyperphenylalaninemia (Anonymous 1999).

Diagnosis

Traditionally, the principal diagnostic/screening tool for PKU has been the Guthrie test, a bacterial inhibition assay able to detect elevated levels of phenylalanine in the blood. There is no true cut-off for abnormal levels of phenylalanine in the blood. A test result exceeding 800 μ mol/l in an untreated child or 240 μ mol/l (4 mg/dL) in a newborn is often used as the diagnostic criterion for an abnormal result requiring further evaluation. Chromatography, fluorometry, enzymatic and tandem mass-spectrometry tests are now also used for screening newborns for elevated blood phenylalanine levels.

Treatment/Management

Dietary intervention aiming to restrict the intake of phenylalanine is the primary treatment for PKU. The aim of treatment is to reduce blood phenylalanine to the near normal range whilst permitting both normal growth and development. This currently requires dietary restriction of all sources of phenylalanine to approximately 10-15% of the normal intake and substitution with a phenylalanine-free formula and low protein alternative to staples such as bread and pasta (personal communication with Dorothy Francis). If treatment is begun as early as possible after birth, symptoms of PKU are preventable. The benefit of treatment is incremental – the

earlier treatment is begun, the better the results. Best results are obtained for treatment begun before two months of age. However, treatment begun up to 18 month of age has some benefits. Treatment should be continued at least until age 10 years, with continuation into adolescence or adulthood now usually recommended to prevent subsequent deterioration in cognitive functioning (particularly executive functioning). However, the diet is restrictive and deviation from it is common. In one randomised trial of 167 children with PKU, 20% of children randomised to strict dietary control and 30% randomised to moderate control had lost dietary control (blood phenylalanine level exceeding 900 μ mol/l for 3 consecutive 6 month periods) by 6 years of age. The majority of children had lost dietary control by age 10 (Azen, Koch, Friedman, Wenz, & Fishler 1996). Outcomes vary and may be determined by brain levels of phenylalanine, which cannot be correlated with blood phenylalanine.

Prevention

There is no known primary prevention for PKU. Tests to detect carriers of the PKU gene by enzymatic testing, fluorometry, or tandem mass spectrometry are not commonly performed and tend to be unreliable because levels of phenylalanine in their blood are too low. However, secondary prevention of impairment from PKU is possible. Babies born to mothers with PKU invariably suffer abnormalities including mental retardation, microcephaly, intrauterine growth delay and congenital heart disease unless the mother's PKU is treated throughout the pregnancy (personal communication with Stephen Kahler). A recent randomised controlled trial demonstrated that this risk can be reduced if women with PKU reinstate dietary treatment prior to conception and to a lesser extent if treatment is reinstated during pregnancy (Waisbren, Hanley, Levy, Shifrin, Allred, Azen, Chang, Cipic-Schmidt, de la Cruz, Hall, Matalon, Nanson, Rouse, Trefz, & Kock 2000).

How might screening reduce the burden of suffering?

Untreated PKU almost invariably results in severe, irreversible mental retardation and other neurological and behavioural deficits. Such outcomes of untreated PKU are associated with extreme cost to the individual, the family and the wider community. Individuals with severe mental retardation require ongoing support and care throughout their lives and nearly all cannot live unassisted. If detected early, these consequences can be prevented by a dietary intervention.

Individuals with PKU where treatment was initiated early in life have been found to have intelligence scores within the normal range. However their average IQ is approximately 0.5 standard deviation below the normal population (Seymour, Thomason, Chalmers, Addison, Bain, Cockburn, Littlejohns, Lord, & Wilcox 1997), and they perform less well on a variety of IQ and development tests than their non-PKU siblings (Azen et al. 1996). In comparison, less than 1% of individuals with untreated PKU have intelligence scores exceeding 70.

Screening for PKU was widely introduced in the 1960s and has been one of the great success stories of our time. Nonetheless, the final effects on the next generation of babies born to mothers with PKU have yet to be quantified but may be considerable, particularly if poor executive functioning impairs these women's ability to reinstate dietary control during pregnancy.

6.3.11.2 EVIDENCE

Tests

All screening tests for PKU detect raised levels of phenylalanine in the blood.

Guthrie bacterial inhibition assay test

This bacterial inhibition assay was the first test developed to detect PKU in newborns. It is semi-quantitative with low capital costs and is used only for the detection of PKU (Pollitt, Green, McCabe, Booth, Cooper, Leonard, Nicholl, J, Nicholson, Tunaley, & Virdi 1997). As the test is only semi-quantitative the classification of abnormal results is somewhat imprecise (open to some judgement error), therefore specificity varies between laboratories. False positive rates are approximately 5%, with approximately 0.15% of newborns requiring a second blood sample to be taken to confirm positive or borderline results (Seymour et al. 1997).

Chromatography

This screening test is semi-quantitative with low capital costs. As with the Guthrie test, this test is only semi-quantitative resulting in varying specificity between laboratories. However, it can be used for the detection of other amino acid disorders in addition to the detection of PKU (Pollitt et al. 1997).

Fluorometry

This screening test is quantitative down to the normal range of blood phenylalanine level and is used only for the detection of PKU. It requires special equipment but has low reagent costs (Pollitt et al. 1997). It can detect phenylalanine levels to an accuracy of 15-30 μ mol/l with coefficient variation of 10%, and false positive rates of 0.02-0.64% on first screen. Despite the lower reported false positive rate, a second blood sample is required for the same number of newborns as for the Guthrie test (approximately 0.15%) to confirm a positive result or borderline result (Seymour et al. 1997).

Enzymatic

This screening test is quantitative down to the normal range of blood phenylalanine levels and is specific to the detection of PKU. It can be run on standard laboratory equipment but has relatively high reagent costs (Pollitt et al. 1997).

Tandem mass-spectrometry

Mass-spectrometry involves the breaking down of molecules into ions by a process of collision. Tandem mass-spectrometry involves two mass-spectrometers working together which enables them to detect the mass of the molecules of interest. This screening test is able to detect multiple disorders including PKU with a single test (Pollitt et al. 1997). It can detect phenylalanine levels to an accuracy of 3 μ mol/l with intra- and inter-assay coefficient variation of 6-10%. Simultaneous measurement of tyrosine allows calculation of phenylalanine:tyrosine ratio therefore fundamentally eliminating false positive results. In one comparison of fluorometry and tandem mass-spectrometry methods all false positive cases generated by fluorometry were correctly identified as negative cases by tandem mass-spectrometry (Seymour et al. 1997). In a similar comparison, fluorometry elicited 91 false positive results compared to only three with tandem mass-spectrometry; this was reduced to one when the phenylalanine:tyrosine ratio was used (Chace, Sherwin, Hillman, Lorey, & Cunningham 1998). The only downside to this test is that the equipment is expensive to purchase.

In the UK in 1988, 43% of infants were tested by laboratories using bacterial inhibition assay (Guthrie), 34% by laboratories using chromatography and 23% by laboratories using fluorometry (Smith, Cook, & Beasley 1991). There is no objective evidence of any major differences in outcome between the Guthrie, chromatography and fluorometry tests. Differences in outcome using these three methods seem more due to variation in screening thresholds at different laboratories and age at which the blood sample is taken than to the tests themselves (Seymour et al. 1997). Tandem mass-spectrometry appears to have improved effectiveness over the other tests and the advantage of being able to screen for a number of conditions. More laboratories are introducing tandem mass-spectrometry as money becomes available to purchase the required equipment.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✓
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	✓

Programs

Screening programs for PKU are already established in Australia and in many other countries and are well accepted by both the medical profession and parents. No randomised controlled trials have been conducted, but the benefits are clearcut for affected individuals.

Most screening programs involve all newborns in a population having a heel prick to extract capillary blood within the first 2 weeks of life. This is usually coordinated through hospitals, midwives and maternal and child health nurses (or equivalent). The blood sample is either collected on a Guthrie card or in a small test tube and is sent to a laboratory for analysis using one of a number of methods (Guthrie test, fluorometry, enzymatic, chromatography or tandem mass-spectrometry). If the level of phenylalanine in the blood exceeds a set cut-off (usually $240\mu\text{mol/l}$ but varies between laboratories) the child is traced back through the hospital or midwife and contacted so that a second blood sample can be taken to confirm the first test. If the initial blood sample is taken prior to 48 hours of life, some programs require a second blood test a few days later, regardless of the result of the initial test, due to slightly lower sensitivity of tests on blood taken prior to 48 hours of life. Following receipt of the second sample, a second analysis is performed to confirm or refute diagnosis of PKU. For positive cases, parents are contacted (through hospitals or maternal and child health nurses) and usually invited to attend a tertiary institution to discuss the results of the test and be provided with education, counselling and medical consultation.

The heel prick test itself can cause distress to infants and parents and taking a second blood sample from "suspected" PKU infants causes parental anxiety. However, these harms are far outweighed by the benefit of detecting PKU early enough to prevent irreversible damage.

In the UK where the Guthrie test, fluorometry and chromatography are all widely used in PKU screening programs, there is no evidence of any difference in outcome between the three methodologies, with all being effective in the detection of PKU (Smith, Cook, & Beasley 1991). In Victoria where the enzymatic test is used there has not been an undetected case of PKU since the inception of the screening program in the 1960s (personal communication with Ivan Francis).

Regardless of which method is used, the sensitivity of the laboratory process for PKU screening at 6 days or later is extremely high at approximately 99.5%. Almost all missed positive cases have been found to be due to technical or human error, for example, no sample was taken or a sample was lost in transit and therefore not actually screened. Screening at 48 hours of life or later is also highly sensitive. However the sensitivity of screening infants within the first 2 days of life is lower, hence most programs advocate a second screen in the first few weeks of life if the initial blood sample was taken prior to 48 after birth.

In a 1984-1988 audit of the PKU screening program in the UK, coverage rates were estimated to approach 100% of live births. However one infant with PKU was known to have been missed during this 4 year period following a negative test result. This child was later diagnosed due to developmental delay but had only mild levels of phenylalanine accumulation ($<720\mu\text{mol/l}$) at diagnosis suggesting that the negative test result at screening may have been accurate (Smith, Cook, & Beasley 1991).

In 1994-1996 an audit of the screening program for PKU and congenital hypothyroidism in the Bath clinical area, UK, was undertaken. This program involves collection of blood samples by heel prick at 6-10 days after birth, performed by a midwife or neonatal nurse. The audit found that 100% of live born babies were screened (excluding 5 whose parents refused screening). However, only 97.9% of blood samples were taken in the prescribed 6-10 day period after birth; 0.7% were taken prior to day 6 and therefore repeated due to concerns with the sensitivity of tests taken early, and 1.3% were taken after day 10 potentially delaying identification and treatment of positive cases, but not requiring a repeat test (unless there was a positive test result). The main areas of the screening program not meeting the prescribed standards related to delays in testing of some blood samples due to delays in the samples reaching the laboratories (0.2%) and inadequate quality of samples for testing (1.9%). However, all positive cases commenced treatment within 28 days after birth as recommended (Simpson N., Randall R., Lenton S., & Walker S. 1997).

No studies have yet systematically and prospectively reported development or parenting outcomes for children born to mothers with PKU detected by screening. This is potentially a major area of concern warranting systematic study and preventive efforts. These women may need instruction in how to prepare PAH free food, may find the diet unpalatable, and some may have impaired executive functioning limiting their ability to plan reinstatement of the PKU dietary treatment before pregnancy and maintain dietary control throughout their pregnancy.

Cost effectiveness

A number of studies have considered the cost effectiveness of screening for PKU. It is difficult to directly compare these studies as each differs in methodology, incidence of PKU used for calculations, discount rate applied, the costs and benefits included in analyses, and other assumptions made. In addition, most studies have failed to consider the cost of lifetime care for individuals with untreated PKU or the costs and benefits

associated with therapy for maternal PKU during pregnancy, or of additional resources required for their children if control during pregnancy is sub-optimal. However, Pollitt and colleagues (1997) reviewed the available economic evaluations and concluded that, despite the differences in methodologies and lack of consideration for psychosocial benefits of screening, there is good evidence that screening programs for PKU are cost effective. Only one of the 11 studies considered yielded inconclusive cost-benefit ratios ranging from 0.8 to 1.4 (a ratio of 1 meaning the costs and benefits are equal). The remaining 10 studies considered yielded cost-benefit ratios in excess of 1, ranging from 1.3 to 6.6.

Quality of evidence

PAPER	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
Smith et al 1991	n/a (program audit)		n/a	n/a	
Simpson et al 1997	n/a (program audit)	Comprehensive audit of clinical standards, parent satisfaction and benefits.	n/a	n/a	UK program but generalizable to the Australian context.

Generalisability of evidence

The majority of evidence comes from the USA and Europe. However the incidence of PKU in these countries is similar to the incidence in Australia, therefore, the evidence presented here is generalisable to the Australian context.

6.3.11.3. CONCLUSIONS

Fair evidence to recommend screening

Comments:

Although never subjected to formal controlled trial, the benefits of screening for phenylketonuria are not in doubt.

However, although treatment of PKU detected by screening prevents intellectual disability, some lasting cognitive effects are common. In addition, long term disability that may be experienced by the offspring of women with PKU (due to high levels of phenylalanine during pregnancy) has not yet been well quantified at population level.

6.3.11.4 RECOMMENDATIONS

Continuation of current universal newborn screening programs for phenylketonuria is recommended.

Screening programs should incorporate population based monitoring and management programs for women of childbearing age, aiming to systematically prevent avoidable intellectual disability in their infants. Such monitoring should meet defined, stringent quality and reporting standards, and include reporting of long term outcomes for these children.

6.3.11.5 FURTHER RESEARCH

As per recommendations.

6.3.11.6 SEARCH TERMS

phenylketonuria

6.3.11.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Stephen Kahler – Director, Victorian Clinical Genetics Services

Dorothy Francis – Dietician, Victorian Clinical Genetics Services

Colin Morley – Reference Group

Bob Williamson – Reference Group

6.3.11.8. REFERENCES

Anonymous. 1999, "Phenylketonuria. The Genetic Mutations and Mode of Inheritance. Phenylketonuria - Mutations and Mode", <http://willroberts.com/pku/mutinh.html>, last updated 10-11-1999.

Azen, C., Koch, R., Friedman, E., Wenz, E., & Fishler, K. 1996, "Summary of findings from the United States Collaborative Study of children treated for phenylketonuria", *European Journal of Pediatrics* p. S29-S32.

Chace, D. H., Sherwin, J. E., Hillman, S. L., Lorey, F., & Cunningham, G. C. 1998, "Use of phenylalanine-to-tyrosine ratio determined by tandem mass spectrometry to improve newborn screening for phenylketonuria of early discharge specimens collected in the first 24 hours", *Clin.Chem.*, vol. 44,no. 12, pp. 2405-2409.

Dietitians Working Party of the Australasian Society for Inborn Errors of Metabolism 1996, *P.K.U. Handbook* Human Genetic Society of Australasia.

- Pollitt, R. J., Green, A., McCabe, C. J., Booth, A., Cooper, N. J., Leonard, J. V., Nicholl, J., Nicholson, P., Tunaley, J. R., & Viridi, N. K. 1997, "Neonatal screening for inborn errors of metabolism: cost, yield and outcome.", *Health Technology Assessment (South Hampton, NY)*, vol. 1, no. 7.
- Seymour, C. A., Thomason, M. J., Chalmers, R. A., Addison, G. M., Bain, C., Cockburn, F., Littlejohns, P., Lord, P., & Wilcox, A. H. 1997, "Newborn screening for inborn errors of metabolism: a systematic review", *Health Technology Assessment*, vol. 1, no. 11, pp. 1-97.
- Simpson N., Randall R., Lenton S., & Walker S. 1997, "Audit of neonatal screening programme for phenylketonuria and congenital hypothyroidism.", *Archives of Disease in Childhood*, vol. 77, pp. F228-234.
- Smith, I., Cook, B., & Beasley, M. 1991, "Review of neonatal screening programme for phenylketonuria", *British Medical Journal*, vol. 303, pp. 333-335.
- Waisbren, S. E., Hanley, W., Levy, H. L., Shifrin, H., Allred, E., Azen, C., Chang, P., Cipic-Schmidt, S., de la Cruz, F., Hall, R., Matalon, R., Nanson, J., Rouse, B., Trefz, F., & Kock, R. 2000, "Outcome at Age 4 Years in Offspring of Women with Maternal Phenylketonuria. The Maternal PKU Collaborative Study", *Journal of the American Medical Association*, vol. 283, no. 6, pp. 756-762.

6.3.12 SCOLIOSIS

6.3.12.1 BACKGROUND

Clinical picture

Scoliosis is a physical disorder characterised by abnormal curvature of the spine. This usually results in x-ray of the spine resembling an “S” (double spinal curve) or “C” shape (single spinal curve). In approximately 80% of cases no specific cause can be identified and these cases are labelled idiopathic (Scoliosis Research Society 1999). Idiopathic scoliosis is the focus of this report.

Traditionally cases of idiopathic scoliosis are considered in three age groups: infantile (develops before 3 years of age), juvenile (3-9 years) and adolescent (10-18 years) (Scoliosis Research Society 1999). The development of scoliosis in children aged 5 years or younger is commonly associated with serious cardiopulmonary abnormalities which are rarely seen in children who develop scoliosis at an older age. Therefore idiopathic scoliosis may be more practically considered in two age groups: early onset (0-5 years of age) and late onset (>5 years of age) (Dickson, Lawton, Archer et al 1948 as cited in Dobbs & Weinstein 1999). Less than 1% of cases of idiopathic scoliosis are infantile and such cases are distinctly different from juvenile and adolescent idiopathic scoliosis (James, Lloyd-Roberts & Pilcher 1959 as cited in Dobbs & Weinstein 1999). Infantile scoliosis will not be discussed further here, since it presents on clinical grounds that do not require screening.

Many cases of idiopathic scoliosis go undetected, particularly minor curves, as generally there is no pain or other symptoms associated with the condition. Pain, fatigue and other symptoms may be present in cases of severe spinal curvature (Connecticut Children's Medical Centre 1999).

Natural History

The natural history of idiopathic scoliosis is unclear. Curves occur along a continuum from mild (Cobb angle of ≤ 30 degrees) through moderate ($>30-45$ degrees) to severe (>45 degrees). Children who develop scoliosis at an earlier age are more likely to develop more severe curves and their curves are more likely to progress from the time of detection to the time of skeletal maturity. In addition, children with more severe curves at the time of detection are more likely to have curve progression, as are those with thoracic curves rather than thoracolumbar or lumbar curves (Scoliosis Research Society 1999). In some cases curve progression may also occur after skeletal maturity. However, it is not clear why curve progression occurs in some cases and not others (Dickson & Weinstein 1999).

Prevalence

The majority of cases (approximately 80-85%) of idiopathic scoliosis occur in individuals 10 years of age or older (Connecticut Children's Medical Centre 1999). The prevalence of adolescent idiopathic scoliosis (defined as a spinal curve of ≥ 11 degrees in 10-16 year olds) in the United States has been estimated at 1-3% (Weinstein 1994 as cited in Rowe, Bernstein, Riddick, Adler, Emans, & Gardner-Bonneau 1997).

Scoliosis is much more common in females. The female to male ratio is 3.6:1 for scoliosis overall and 10:1 for moderate to severe scoliosis (spinal curve of ≥ 30 degrees) (Connecticut Children's Medical Centre 1999). One in 952 girls (1.05%) screened for scoliosis in the Hornsby and Ku Ring Gai regions of Sydney between 1989 and 1996 were found to have significant scoliosis (defined as a spinal curve of ≥ 30 degrees) (Wilkinson 1999).

Genetics relevant to screening issues

None known

Diagnosis

The unit of measurement used for the lateral spinal curve is the Cobb angle. However, use of the Cobb angle as a measure of severity has been questioned as this measurement is not linearly proportional to the severity of the condition. Equally, using the Cobb angle to measure curve progression is questionable as a difference of 5 degrees in a mild curve (ie from 10 to 15 degrees) is not equal to a progression of 5 degrees in a severe curve (ie from 65 to 70 degrees) (Dickson & Weinstein 1999).

As most cases of idiopathic scoliosis are associated with no pain or other symptoms, diagnosis can be difficult and many cases go undetected (Connecticut Children's Medical Centre 1999). Severe curves are more easily detected than minor curves. Diagnosis is generally made through physical examination and radiological examination is the gold standard. For diagnosis of adolescent idiopathic scoliosis the Scoliosis Research Society stipulates a lateral spinal curve (Cobb angle) of more than 10 degrees in children 10 years

of age or older who have reached skeletal maturity (Scoliosis Research Society Terminology Committee 1976 as cited in Rowe et al. 1997).

Treatment/Management

Mild curves (Cobb angle of 20 degrees or less before skeletal maturity) tend not to receive treatment, but are monitored, usually six monthly, for signs of progression. Moderate curves (> 30 degrees) or curves that progress 5-10 degrees before the patient reaches skeletal maturity tend to be treated with a brace. Severe curves (> 45 degrees in growing adolescents or > 60 degrees in skeletally mature adolescents) or curves that have progressed despite brace treatment tend to be treated surgically (Rowe et al 1997).

Milwaukee brace treatment

The Milwaukee brace was initially developed as an adjunct to surgical treatment, not as a stand alone therapy. However, the brace is now used as a conservative treatment alternative to surgery, but its effectiveness as a treatment is debated. Some suggest that the effectiveness of brace treatment is a reflection of the flexibility of the individual curve, with smaller curves tending to be more flexible and therefore appearing to benefit most from brace treatment. However, all curves, even the most severe ones, have some flexibility so brace treatment can superficially appear to be beneficial. The brace increases the rigidity of the spine rather than allowing the maintenance of flexibility as would be the case in an individual who was not braced. The brace holds the curve between the standing magnitude of the curve and the supine or maximum traction or maximum side bending. The brace reduces the lateral curvature, thereby preventing expression of the primary lordosis in the frontal plane, but does not impact on the other two aspects of the deformity, the secondary lateral spinal curvature and the rotation. As a treatment to prevent curve progression, the effectiveness of the brace is also questioned as approximately three quarters of mild to moderate curves do not progress when receiving no treatment (Dickson & Weinstein 1999).

Compared to other non-operative treatments for scoliosis (observation or lateral electrical surface stimulation), a meta-analysis found bracing had greater effectiveness (Rowe et al. 1997). However, this meta-analysis does not consider the issues discussed above, namely that the 'effectiveness' of the brace treatment may be merely a reflection of the flexibility of the curves in the girls studied. In a recent study of 33 Australian girls diagnosed with scoliosis through a screening program, treatment with a brace was recommended for 20 girls. The spinal curve in only two girls was improved following brace treatment (by 7 and 11 degrees), with the curves of five girls worsening to the point of requiring surgical treatment (Wilkinson 1999).

The meta-analysis also found that the effectiveness of bracing increased with the more hours per day the brace was worn (Rowe et al. 1997). Thus compliance, which is a major problem with brace treatment, would impact on the effectiveness of this treatment. Many girls find brace treatment to be uncomfortable, cumbersome and cosmetically unacceptable. Of the 20 girls recommended brace treatment in the Australian study, two refused outright, three tried the brace but abandoned it and three abandoned the brace after wearing it for less than 75% of the time recommended. Overall the girls were generally unsatisfied with this form of treatment (Wilkinson 1999). In one study using hidden compliance meters, it was found that children generally wore their braces for only 10% of the prescribed time, despite reporting greater compliance to their doctors (Houghton, McInerney & Tew 1987 as cited in Dickson & Weinstein 1999).

Surgical treatment

Surgical treatment involves straightening of the spine by inserting rods, balancing the trunk of the pelvis and stabilising the spine by arthrodesis (Cassella & Hall 1991). It generally results in a favourable cosmetic outcome, improving the shape of the spine with a very low rate of complications (Dickson & Weinstein 1999). In the Australian study, five girls identified through screening received surgical treatment initially and a further five received surgical treatment after bracing treatment proved ineffective and their curves progressed. The girls treated surgically were generally happier with their treatment outcome than girls who received brace treatment (Wilkinson 1999).

Prevention

No information was identified on the prevention of idiopathic scoliosis.

How might screening reduce the burden of suffering?

Generally, early onset scoliosis has a poorer prognosis than late onset scoliosis. Significant, progressive early onset spinal curves can lead to cardiopulmonary conditions in later life with implications on later medical care. Significant curves can result in problems including back pain, fatigue, body image and self esteem problems (Connecticut Children's Medical Centre 1999). In such cases treatment can be helpful, particularly to improve the psychological functioning of the patient. However, the majority of such cases would be detected without screening as the severe curves which cause such problems are easily visible and/or symptomatic (Dickson & Weinstein 1999).

In most cases of late onset idiopathic scoliosis organic health is not affected regardless of the size of the curve and there is little or no reduction in the functioning of the spine (Dickson & Weinstein 1999). Therefore regardless of whether late onset scoliosis is detected or goes undetected, such cases are unlikely to burden the individual (in terms of decreased functioning or increased health problems) or the community (in terms of increased medical costs, loss of earnings or decreased life expectancy) (Dickson & Weinstein 1999).

6.3.12.2 EVIDENCE

Tests

Forward Bend Test

Also known as the Adams Forward Bend Test, this is the most common test used in screening programs and for initial clinical assessment. It involves physical examination of the back when the child is standing with feet together and bending forward from the waist with arms dangling. This test is usually preceded by visual inspection of the back for asymmetry when the child is standing in an upright position (Goldbloom 1994). Children are required to remove all clothing above the waist, although girls are usually permitted to leave their bras on. This test is not quantitative, relying on clinical judgement, and thereby provides nothing with which to compare subsequent examinations or document curve progression. However, it is quick and easy to perform, making it desirable for population-based community screening programs.

A study of 31,416 Australian children reported the Forward Bend Test to have a sensitivity of 100%, specificity of 98%, likelihood ratio for a positive result of 57 and likelihood ratio for a negative result of 0 (Wilkinson 1999). However, less than 145 girls with positive screens were x-rayed in the three year period to confirm a spinal curve ≥ 30 degrees and no negative screens were subjected to gold standard (personal communication with Helen Wilkinson).

Scoliometer

This instrument is an inclinometer that quantifiably measures trunk asymmetry or axial trunk rotation or rib prominence. It is used while the child is performing the Forward Bend Test, is small and portable making it potentially useful for community-based screening.

One study investigated the reliability and validity of the Scoliometer in a sample of 65 5-37 year old patients (mean age 14.8 years) with idiopathic scoliosis (34 with single spinal curves and mean angle of 21 degrees, and 31 with a double spinal curve and mean angle of 29 degrees). With two experienced clinicians self taught in use of the instrument there was good interrater (0.88-0.96 SD=0.97-1.35) and intrarater (0.86-0.97 SD=0.77-1.13) reliability for assessment of patients with single and double curves and upper and lower curves. The Scoliometer readings were moderately correlated with lateral curvature (Cobb angle) ($r=0.46-0.54$) and vertebral (pedicle) rotation ($r=0.32-0.46$), both assessed using radiography. The sensitivity, specificity, PPV and NPV of the Scoliometer for assessment of a curve (results for single and double curves combined) using a diagnosis threshold of five degrees was 98%, 29%, 70% and 86% respectively, using a threshold of 7.5 degrees was 78%, 75%, 84% and 67%, and using a threshold of 10 degrees was 51%, 96%, 95% and 53% respectively (Amendt, Aulsebrook, Eybers, Wadsworth, Nielsen, & Weinstein 1990).

Moire Topography

This is a biostereometric method producing a three dimensional image of the trunk which can pick up any asymmetries. It uses a portable moire screen and strong portable light source to illuminate the subject. A photograph can be taken of the image to provide a permanent record. The child stands upright with feet together behind the screen with shoulders and buttocks lightly touching the screen. As with all tests for scoliosis, children must remove all clothing above the waist, although girls are usually permitted to leave their bras on.

In a study of 182 females aged 12-16 it was found that approximately 40 children per hour could be screened using moire topography. All children also received a physical examination involving the Forward Bend Test. Both screening tests were conducted double blind. 23 girls (12.6%) were considered to have some spinal curvature by the moire topography technique, 18 (9.9%) by physical examination, and a further three children by both methods. X-ray examination confirmed a spinal curvature > 5 degrees in 15 of the children detected by moire topography, 13 detected by physical examination and two detected by both methods. Of these, nine detected by moire topography had curves of > 10 degrees, four detected by physical examination and two detected by both methods (Thompson, Walsh, & Colville 1985). However, curves of ≤ 20 degrees are considered mild and typically do not receive treatment, therefore the value of detecting curves of this size is questionable.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	✗
Accepted treatment	✗
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	?
Suitable test or examination	✓
Test acceptable to the population	✓
Natural history adequately understood	✗
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	✗

Programs

In a review of the Hornsby and Ku Ring Gai (NSW) scoliosis screening program, 34,416 girls were screened using the Forward Bend Test and 579 positive cases were identified, of which 33 were confirmed by radiology to have a moderate-severe spinal curve of ≥ 30 degrees. However, 79% of the 33 cases reported already knowing about their scoliosis or being confident that they would have noticed their condition without screening. In addition, despite the premise that screening leads to earlier identification and therefore earlier treatment of conditions, the mean time between diagnosis prompted by screening and treatment was 12.6 months (range 2-38 months), with eight of the 33 cases advised that they did not require treatment at all (Wilkinson 1999). Recommended treatments were frequently unacceptable to the recipients, and varied widely between surgeons for similar curves. While the screening program was feasible and efficient, serious questions about its efficacy were raised by the author.

A further review considered the community-based school screening program for adolescent idiopathic scoliosis in Rochester, Minnesota. This program involves public health nurses screening all children in grades 5-9 in the Rochester school districts (public and private schools) every year during physical activity classes. Screening involves visual inspection of the back in the upright position, the Forward Bend Test, and as of 1986, use of a scoliometer. Children are rescreened within 2-4 weeks if an obvious curve is detected during the physical examination or if they have a scoliometer reading of >6 units. If rescreening confirms the original screening results, a letter is sent home to parents recommending that their child see a health care professional for further evaluation. If children do not meet the criteria for rescreening within 2-4 weeks but appear to have some spinal curvature, the public health nurse screens these children every 6 months rather than every 12 months. Of 2242 children who started school between 1979 and 1981 and were therefore involved in the screening program between 1984 and 1990, 92 (4.1%; 43 boys and 49 girls) were referred one or more times for abnormal screening results. Of these, only 68 children (74% or 3% of the total population) received further evaluation for scoliosis; by 19 years of age 27 (1.2% of the total population) had received a diagnosis of scoliosis (true positives) and 41 had not been diagnosed as having scoliosis (false positives). Of the 27 children with scoliosis, 11 had curves of 11-19 degrees, 10 had curves of 20-39 degrees and six had curves ≥ 40 degrees, but only five had received treatment by the age of 19. A further 328 children (14.6%) were assessed by public health nurses as having some spinal curvature during one or more screens and therefore received six monthly screens until they left the program after grade nine. Of these children, 62 (19%) had received further evaluation for scoliosis but there were only five confirmed diagnoses (three with curves 11-19 degrees and one each with curves 20-39 degrees and ≥ 40 degrees) and none had received treatment. Three children were diagnosed with scoliosis prior to entering the screening program. Six children were diagnosed with scoliosis despite not being picked up in the screening program for referral or closer monitoring. Therefore this community-based school screening and surveillance program for adolescent idiopathic scoliosis had a sensitivity of only 0.56 and PPV of 0.05 for cases of scoliosis receiving treatment by age 19 if it is assumed that all referred children who did not attend for assessment did not receive treatment. The sensitivity and PPV are higher if the outcome is curves assessed as ≥ 20 degrees by

19 years of age (0.64 and 0.17 respectively) or curves of ≥ 40 degrees by age 19 (0.67 and 0.07 respectively) (Yawn, Yawn, Hodge, Kurladn, Shaughnessy, Ilstrup, & Jacobsen 1999).

Another study assessed a school-based scoliosis screening program in the Netherlands, where children undergo periodic health checks every two years which includes scoliosis screening. Three cohorts of children were screened during a scheduled school periodic health check and then re-screened one and two years later: 10,000 children aged 10 years at the initial screening, 10,000 children aged 12 years at the initial screening, and 10,000 children aged 14 years at the initial screening. Screening was carried out by school doctors and involved a Forward Bend Test. Children with a positive screening result were invited to attend a second stage screening session involving moire topography and measurement of rib hump height and the angle of trunk rotation. Children with positive moire topography or measurement results who had not been detected in previous health checks were referred to their GP, at whose discretion a radiographic examination and/or referral to an orthopaedic surgeon could occur. In total, 30,611 children were screened in the first year of whom 4,189 (13.7%) had a positive Forward Bend Test and were invited to attend the second stage screening sessions. Only 3,065 children attended the second stage screening sessions (73.2% response) of whom 1,931 (63.0%) received positive results and were referred to their GP; 123 children had been previously identified and previously referred. At the two year follow up (9% drop out at one year and 22% drop out at two years), of all children referred to GPs (new and previously identified cases), only 671 had received x-rays. Of these, 188 new cases had confirmed Cobb angles of 10-19 degrees and 57 had Cobb angles >19 degrees, and 34 previously identified cases had Cobb angles >19 degrees. Less than 0.3% (91 children) of the children initially screened were identified with curves >19 degrees and only ten of the 57 newly identified cases with curves >19 degrees received any treatment (a yield of one child for every 3,000 screened). Anxiety caused by the program resulted in many parents bypassing the second stage screening and taking their child directly to their GP for further assessment (Prujjs, van der Meer, Hageman, Keessen, & van Wieringen 1996). As not all children received x-rays, sensitivity and specificity values could not be calculated.

In a 1993-1994 study of scoliosis in north-western and central Greece, 82,901 school children aged 9-14 years were screened for scoliosis using the Forward Bend Test. Children who were suspected of scoliosis from the first screen were reassessed later the same day by another examiner. When a positive result was confirmed by the second screen, children were referred to a local hospital for a radiograph. Positive signs included asymmetrical shoulder levels, scapular prominence, unequal distance from the upper extremities to the flanks or inequality of the lengths of the lower limbs while the patient was standing, and lateral deviation of the spine during the Forward Bend Test. One or more positive signs were observed in 5,803 children (7%; 62% female) at the first screen and positive results were confirmed by the second screen in 4,185 cases (66% female). Radiograph results confirmed 1,436 children (1.7% of the total sample; 80% female) could be diagnosed with scoliosis (curve ≥ 10 degrees) and a further 1,347 children had a curve of 1-9 degrees, while 1402 children had no radiographic signs of spinal curvature (Souacos, Souacos, Zacharis, Beris, & Xenakis 1997).

In summary, it appears that the number of significant cases detected by scoliosis screening that would not be detected otherwise is relatively small. There does not appear to be great benefit in detecting cases earlier, particularly as many of the detected cases do not require treatment. The efficacy of the available treatments is also questionable, and is unacceptable to many.

Cost effectiveness

No studies of cost effectiveness have been reported.

Quality of evidence

PAPER	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
Wilkinson	IV	Gold standard not applied to all children and no mention of blinding.	n/a	n/a	Australian study.
Yawn et al	IV	Gold standard not applied to all children and no mention of blinding.	n/a	Sensitivity = 56% and PPV = 5% if the outcome is treatment by 19 years of age (assuming all referred children who did not attend for assessment did not receive treatment) Sensitivity = 64% and PPV = 17% if the outcome is curves assessed as ≥ 20 degrees by 19 years of age Sensitivity = 67% and PPV = 7% if outcome is curves of ≥ 40 degrees by 19 years of age.	Generalizable to the Australian context.
Prujjs et al	IV	Gold standard not applied to all children and no mention of blinding.	n/a	n/a	Generalizable to the Australian context.
Soucacos et al.	IV	Gold standard not applied to all children and no mention of blinding.	n/a	n/a	Generalizable to the Australian context.

Four studies of screening programs for scoliosis were identified. These were all of a relatively low level of evidence (level IV). Two further studies were identified which investigated screening tests for scoliosis.

Generalisability of evidence

The evidence presented here is generalizable to the Australian context.

6.3.12.3. CONCLUSIONS

Fair evidence to recommend against screening

Comments:

Although there are good, relatively inexpensive tests available to screen for idiopathic scoliosis, there is no evidence that screening programs are beneficial. Most cases of idiopathic scoliosis are detectable without screening, many do not require treatment and the treatments available do not appear to be of benefit to those with mild to moderate spinal curvature.

6.3.12.4 RECOMMENDATIONS

Implementation of new scoliosis screening programs is not recommended.

In communities where screening programs are already in place, continuation of these programs should be reassessed.

6.3.12.5. FURTHER RESEARCH

- Quantification of the proportion of children with scoliosis who will have significant progression of their spinal curves, and in which children significant progression will occur.
- Randomised controlled trials to quantify the effectiveness of current therapies on an intention-to-treat basis.

6.3.12.6. SEARCH TERMS

Scoliosis

Spinal curvatures

6.3.12.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Helen Wilkinson – Community Paediatrician, Hornsby Child Health, NSW

Ian Torode – Orthopaedic surgeon, Department of Orthopaedics, Royal Children's Hospital, Melbourne

John Aloizos – Reference Group

George Patton – Reference Group

6.3.12.8 REFERENCES

Amendt, L. E., Ause-Ellias, K. L., Eybers, J. L., Wadsworth, C. T., Nielsen, D. H., & Weinstein, S. L. 1990, "Validity and reliability testing of the Scoliometer", *Physical Therapy*, vol. 70,no. 2, pp. 108-117.

Cassella, M. C. & Hall, J. E. 1991, "Current treatment approaches in the nonoperative and operative management of adolescent idiopathic scoliosis", *Physical Therapy*, vol. 71,no. 12, pp. 897-909.

Connecticut Children's Medical Centre. Idiopathic Scoliosis. Orthopaedics / Idiopathic Scoliosis, <http://www.ccmckids.org:8081/departments/Orthopaedics/orthoed21.htm>, last updated 1-5. 1999. Connecticut Children's Medical Centre, Hartford, Connecticut.

Dickson, R. A. & Weinstein, S. L. 1999, "Bracing (and Screening) - Yes or No?", *The Journal of Bone and Joint Surgery*, vol. 1,no. B(2), pp. 193-198.

Dobbs, M. B. & Weinstein, S. L. 1999, "Infantile and Juvenile Scoliosis", *Disorders of the Pediatric and Adolescent Spine*, vol. 30 ,no. 3, pp. 331-341.

Goldbloom, R. B. 1994, "Screening for Idiopathic Adolescent Scoliosis". *The Canadian Guide to Clinical Preventive Health Care*, Canadian Taskforce of Preventive Health Care, Ottawa.

- Prujls, J. E. H., van der Meer, R., Hageman, M. A. P. E., Keessen, W., & van Wieringen, J. C. 1996, "The benefits of school screening for scoliosis in the central part of the Netherlands", *European Spine Journal*, vol. 5, pp. 374-379.
- Rowe, D. E., Bernstein, S. M., Riddick, M. F., Adler, F., Emans, J. B., & Gardner-Bonneau, D. 1997, "A meta-analysis of the efficacy of non-operative treatments for idiopathic scoliosis", *Journal of Bone & Joint Surgery American Volume.*, vol. 79, p. 10.
- Scoliosis Research Society. Indepth review of Scoliosis. Scoliosis Research Society. <http://srs.org>. Viewed December 1999.
- Souacos, P. N., Souacos, P. K., Zacharis, K. C., Beris, A. E., & Xenakis, T. A. 1997, "School-screening for Scoliosis: A prospective epidemiological study in Northwestern and Central Greece", *Journal of Bone & Joint Surgery American*, vol. 79-A,no. 10, pp. 1498-1503.
- Thompson, F., Walsh, M., & Colville, J. 1985, "Moire topography: a method of screening for adolescent idiopathic scoliosis", *Irish Medical Journal*, vol. 78, pp. 162-165.
- Wilkinson, H. F. "Scoliosis screening outcomes", Hornsby and KuRing Gai Hospital and Community Health services, Child, Adolescent and Family Health Team, Sydney.
- Yawn, B. P., Yawn, R. A., Hodge, D., Kurland, M., Shaughnessy, W. J., Ilstrup, D., & Jacobsen, S. J. 1999, "A population-based study of school scoliosis screening", *Journal of the American Medical Association*, vol. 282,no. 15, pp. 1427-1432.

6.3.13. CRYPTORCHIDISM (UNDESCENDED TESTIS)

6.3.13.1 BACKGROUND

Clinical picture

Cryptorchidism is a condition in which one or both testes do not reside or cannot be manipulated into the scrotum. Migration of the testis from its origin in the abdomen through the inguinal canal to the scrotal sac occurs over time. Failure of this process results in maldescent and the testis may be found anywhere along this path. An undescended testis (UDT) is arrested in the normal path of descent whereas an ectopic testis follows an aberrant course and can be found in the abdomen, groin, perineum or thigh.

In addition, some testes found to be in the normal position in infancy later ascend out of the scrotum. This 'acquired cryptorchidism' is thought to be due to failure of the spermatic cord to lengthen sufficiently with body growth. Boys with UDT that descend spontaneously in the first few months of life may be at greater risk of 'ascending testes'.

'Retractile testes' can be manipulated into the normal position in the scrotum, but will ascend again after a period of time.

Testicular descent occurs late in gestation, and in many infants, continues after birth. The majority of undescended testes have spontaneously descended by 3 months of age. Those with cryptorchidism at 3 months are more likely to have other genito-urinary abnormalities eg hypospadias, a small scrotum and poor scrotal rugation (John Radcliffe Hospital Cryptorchidism Study Group 1992).

Testicular biopsies have shown irreversible progressive changes in the germinal epithelium within the first 2 years of life, and as a result, most paediatric centres recommend elective orchidopexy before the child's second birthday. Growth of the testis following surgery for unilateral cryptorchidism has been demonstrated, with increasing testicular size more common among boys operated upon when younger than 18 months rather than older (Nagar & Haddad 1997).

Boys with a history of undescended testes have increased rates of infertility, malignancy, and torsion of the cryptorchid testis. Development of antisperm antibodies is also associated with cryptorchidism (Sinisi, Pasquali, Papparella, Valente, Orio, Esposito, Cobellis, Cuomo, Angelone, Martone, Fioretti, & Bellastella 1998). It is known that ectopic testes have worse outcomes and that bilateral cryptorchidism left untreated until puberty will result in infertility. However it is not clear what long term effect surgery has on infertility and malignancy, and what age is optimal for treatment.

Biopsy of undescended testes has shown that they are normal at birth, but that histological changes can be demonstrated before the child is 12 months old. These abnormalities progress if the testis remains undescended and can also be found in the descended contralateral testis over time. Untreated bilateral cryptorchidism will result in infertility; surgically corrected UDT is associated with reduced fertility; and the age at which surgery is carried out is linked with adult dyspermia – those with postpubertal correction having worse outcomes (Lenzi, Gandini, Lombardo, Dondero, Culasso, Ferro, Cambiaso, Caione, & Cappa 1997).

The risk of malignancy in a unilateral undescended testis is approximately 1 in 80 and in bilateral undescended testes the rate is 1 in 40-50. This is 4-10 times higher than the general population (Hadziselimovic, Herzog, & Emmons 1997). An increased risk of malignancy in patients who had testicular biopsy at the time of orchidopexy is noted in a recent study (Swerdlow, Higgins, & Pike 1997). The rates of malignancy are higher in abdominal testes. In addition, there is a 2-3 fold increased risk in the unaffected contralateral testis in patients with unilateral maldescent (Gordon 1995). This suggests either an underlying mechanism that predisposes the individual to both testicular maldescent and malignancy, or perhaps the effect of auto-antibodies triggered by the unilateral undescended testis. There is no evidence that orchidopexy reduces the risk of tumour development, but it permits regular examination and thus early detection of testicular cancer.

Retractile testes, previously thought to be an innocuous variant, have also been associated with infertility and abnormal histology (Caucci, Barbatelli, & Cinti 1997; Sang Won Han, Tack Lee, Jang Hwan Kim, Seung Kang Choi, Nam Hun Cho, & Ji Young Han 1999).

Prevalence

The fact that testicular descent occurs late in gestation and may continue after birth results in a higher incidence in premature babies and a decreasing incidence in older infants as spontaneous resolution occurs. Cryptorchidism affects 5% of full-term and 25% of premature male births (John Radcliffe Hospital Cryptorchidism Study Group 1986). The majority of cryptorchid testes have descended by 3 months of age, and by 6 months the prevalence has decreased to 0.8% (Elder 2000). The John Radcliffe Hospital

Cryptorchidism Study examined boys at 3 months of age and, after excluding those with congenital malformations, found the overall prevalence of undescended testes to be 1.55%. In this study, 1.92% of boys had bilateral and 3% unilateral cryptorchidism at birth (John Radcliffe Hospital Cryptorchidism Study Group 1986).

Low birth weight has been identified as a risk factor for UDT (Gill & Kogan 1997; Jones, Swerdlow, Griffith, & Goldacre 1998) and in the study by Jones et al gestational age was not independently associated with cryptorchidism after adjusting for birthweight (Jones et al. 1998).

Genetics relevant to screening issues

None known

Diagnosis

The diagnosis of UDT is made on clinical examination. The criteria used in the 1956 study by Scorer are still widely cited (Gordon 1995). Complete descent was defined by Scorer as cases in which the testis could be drawn down 4cm or more from the pubic crest. The top of the pubic crest and the middle of the testicle were taken as the points of measurement. A shorter distance of 2.5cm was used for premature babies, weighing 5lb 8oz or less, in accordance with their smaller size. In the John Radcliffe Hospital Cryptorchidism Study researchers found that position in the scrotum was as reliable as diagnosis by measurement (John Radcliffe Hospital Cryptorchidism Study Group 1992). Definitions used include normal, high scrotal, supra-scrotal, non-palpable and other.

Undescended testes are usually found in the inguinal canal, however approximately 15% will be non-palpable on clinical examination. These ectopic testes may be in the superficial inguinal pouch or perineum; may be intra-abdominal; or absent. In the child with bilateral non-palpable testes, female virilisation should be considered.

Treatment/Management

Palpable testes

Testes in the inguinal canal can be brought down into the scrotum surgically. This procedure, orchidopexy, is usually carried out as a day stay in hospital. Following an inguinal incision, the testis and spermatic cord are mobilized, the testis is brought down and fixed in the scrotum and the associated hernia is corrected.

Biopsy studies have shown that irreversible progressive changes in the germinal epithelium can first be seen at 12 months of age. In view of these degenerative changes and the infrequency of spontaneous descent after 1 year of age, it is currently recommended that elective orchidopexy be performed in the child's second year.

Non-palpable testes

Diagnostic laparoscopy is required in these infants to locate the ectopic testis. For those testes found close to the internal inguinal ring, orchidopexy is often successful. Due to the high risk of malignancy, orchidectomy (removal of the testis) may be considered for more difficult intra-abdominal testes or those that appear atrophic. Testicular prostheses are available for boys in this situation.

Hormonal treatment

Human Chorionic Gonadotrophin (hCG), Luteinising Hormone Releasing Hormone and Gonadotrophin Releasing Hormone have been used to stimulate testicular descent. Results have been mixed, but pre-operative use of hCG has been linked to an increase in germ cell death, smaller testicular size and high FSH levels in adults. (Dunkel, Taskinen, Hovatta, Jonathan, & Wikstrom 1997) (Taskinen & Wikstrom 1997)

Prevention

There are no known preventive factors for UDT.

How might screening reduce the burden of suffering?

Screening to detect and treat testicular maldescent will reduce the increased rates of infertility, malignancy, torsion of the cryptorchid testis and the possible psychological effects related to an empty scrotum.

6.3.13.2 EVIDENCE

Tests

Clinical examination is the only screening test for UDT

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	?
Repeatable	?
Sensitive	?
Specific	?

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	?
Test acceptable to the population	✓
Natural history adequately understood	?
policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	N/A

Programs

Examination of the infant at birth, at the 6 week postnatal check, and at other milestones in the pre-school years have been performed in various screening programs, as well as re-examination at school entry.

It has been proposed that the best time to examine a child for UDT is in the newborn period when the cremasteric reflex is absent, the scrotum is proportionately at its largest and there is little subcutaneous fat. (Gill & Kogan 1997;Goh & Hutson 1992). Physical examination becomes technically more difficult and less reliable as the child gets older. A follow up examination at 3-6 months for boys noted to have UDT at birth has been recommended to exclude those whose testes have descended spontaneously (Gill & Kogan 1997;Goh & Hutson 1992).

No evidence of effectiveness of screening for UDT was found during the course of this review, however several audits of screening programs provides some information about the outcomes and limitations.

A retrospective audit of the screening history and subsequent management of boys treated for UDT identified low detection rates (Rao, Wilkinson, & Benton 1991). The study by Rao et al found that medical records of neonatal examinations only documented the diagnosis in 19% of boys who subsequently underwent surgery for UDT testes, and only half of those were referred for further management. There was no record of testicular examination in 40% of patients, and 32% were recorded as having normal testicular descent. It is possible that boys in this latter group went on to develop ascending testes, and hence were not in fact 'missed'. It is difficult to interpret the information regarding Health Visitor screening of pre-schoolers as this was not a routine part of their examination; or medical examinations at school entry as many of the children undergoing the screening examination had already been diagnosed.

An audit of screening in successive birth cohorts suggests that although there is an increase in cases detected by community screening, there are also a significant number of identified cases that are not referred for further management (Tamhne, Jarvis, & Waterson 1990). The average ages of surgical referral (Tamhne, Jarvis, & Waterson 1990) and orchidopexy (Kaul & Roberts 1992) are falling, suggesting that detection by screening may be improving. However the average age at surgery is reported as 6-7 years in the Tamhne et al audit (Tamhne, Jarvis, & Waterson 1990). Other studies report only 39% of children had surgery before the age of 6 and 13% before the age of 2 (Kaul & Roberts 1992).

Problems with screening that have been identified from these studies include lack of clear diagnostic criteria, lack of appropriate training for personnel involved, inadequate guidelines and pathways for referral.

Cost effectiveness

No information regarding the cost-effectiveness of screening for UDT was found in the course of this review.

Generalisability of evidence

The studies discussed are from the UK where screening programs for UDT are conducted in the same manner as Australia.

Other issues

Parental reporting of UDT was the impetus for over 40% of cases in one study (Rao, Wilkinson, & Benton 1991). As parental participation is now recognised as an important part of child health surveillance this is an area of potential involvement.

Testicular self-examination should be recommended to all boys who have undergone orchidopexy so that developing tumours are recognised and treated early.

6.3.13.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening

Comments:

No evidence regarding the effectiveness of screening programs or screening tests for undescended testes was found.

The current practice in Australia is for all children to have a neonatal examination conducted by a health professional that would routinely include examination of the genitalia. There are several factors that would encourage continuation of this practice:

- Clinical examination is the only test available for detecting undescended testes
- Consensus is that the optimum time for this examination is at birth
- If the testis remains undescended after the first year of life, there is evidence that damage ensues
- There is evidence that treatment before the age of 2 years improves adult fertility

However, it appears that screening for undescended testes frequently falls short of *program* requirements for an effective detection system. UK data suggests poor rates of detection from neonatal screening and that, of those detected, many are not referred for surgical opinion and treatment. In addition, lack of clear diagnostic criteria, lack of appropriate training for personnel involved, inadequate guidelines and pathways for referral have been identified as problems related to this process.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

No evidence is available about the effectiveness of screening for undescended testes at later ages, nor is there evidence regarding the benefits of treatment beyond the age of two.

6.3.13.4. RECOMMENDATIONS

Although there is little firm evidence to support the value of screening, we recommend continuation of specific examination of the genitalia at the newborn and 6 week checks, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical and nursing staff
- clear pathways and/or guidelines for referral
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records
- strategies to improve attendance at the 6 week postnatal check (particularly for disadvantaged groups)
- opportunistic examination of children not examined in the newborn period
- program evaluation and quality improvement

6.3.13.5. FURTHER RESEARCH

- Accuracy of the clinical examination as a screening tool
- Assessment of effectiveness of treatment and the optimum age for treatment
- The natural history of retractile and ascending testes
- Evaluation and continuing quality improvement of programs
- Effectiveness of neonatal screening programs
- Role of parental participation in surveillance for undescended testes

6.3.13.6. SEARCH TERMS

MeSH headings: Genitalia, Urogenital abnormalities

Keywords: Genitalia, Urogenital abnormalities, Cryptorchidism

6.3.13.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Reference group: Caroline Briggs, Neil Wigg

Clinical experts: Prof John Hutson, Professor of Paediatric Surgery, University of Melbourne, Royal Children's Hospital, Melbourne

6.3.13.8. REFERENCES

Caucci, M., Barbatelli, G., & Cinti, S. 1997, "The retractile testis can be a cause of adult infertility", *Fertility and Sterility*, vol. 68,no. 6, pp. 1051-1058.

Dunkel, L., Taskinen, S., Hovatta, O., Jonathan, L., & Wikstrom, S. 1997, "Germ Cell Apoptosis after Treatment of Cryptorchidism with Human Chorionic Gonadotropin is Associated with Impaired Reproductive Function in the Adult", *The Journal of Clinical Investigation*, vol. 100,no. 9, pp. 2341-2346.

Elder, J. 2000, "Disorders and Anomalies of the Scrotal Contents," in *Nelson Textbook of Paediatrics*, 16 edn, R. Behrman, R. Kliegman, & H. Jensen, eds., WB Saunders Company, Philadelphia.

Gill, B. & Kogan, S. 1997, "Cryptorchidism. Current Concepts", *Pediatric Clinics of North America*, vol. 44,no. 5, pp. 1211-1227.

Goh, D. W. & Hutson, J. M. 1992, "Screening for undescended tests", *Archives of Disease in Childhood*, vol. 67, p. 350.

Gordon, N. 1995, "Undescended testes: screening and early operation.", *British Journal of Clinical Practice*, vol. 49,no. 6, pp. 318-320.

Hadziselimovic, F., Herzog, B., & Emmons, L. R. 1997, "The incidence of seminoma and expression of cell adhesion molecule CD44 in cryptorchid boys and infertile men.", *Journal of Urology*, vol. 157,no. 5, pp. 1895-1897.

John Radcliffe Hospital Cryptorchidism Study Group 1986, "Cryptorchidism: an apparent substantial increase since 1960", *British Medical Journal*, vol. 293,no. November, pp. 1401-1404.

John Radcliffe Hospital Cryptorchidism Study Group 1992, "Cryptorchidism: a prospective study of 7500 consecutive male births, 1984-8", *Archives of Disease in Childhood*, vol. 67, pp. 892-899.

Jones, M. E., Swerdlow, A. J., Griffith, M., & Goldacre, M. J. 1998, "Prenatal risk factors for cryptorchidism: a record linkage", *Paediatric and perinatal Epidemiology*, vol. 12,no. 4, pp. 383-386.

Kaul, S. A. & Roberts, D. P. W. 1992, "Preschool screening for cryptorchidism", *British Medical Journal*, vol. 305, p. 181.

Lenzi, A., Gandini, L., Lombardo, F., Dondero, F., Culasso, F., Ferro, F., Cambiaso, P., Caione, P., & Cappa, M. 1997, "Unilateral cryptorchidism corrected in prepubertal age: evaluation of sperm parameters, hormones, and antisperm antibodies in adult age", *Fertility and Sterility*, vol. 67,no. 5, pp. 943-948.

- Nagar, H. & Haddad, R. 1997, "Impact of early orchidopexy on testicular growth", *British Journal of Urology*, vol. 80, pp. 334-335.
- Rao, M., Wilkinson, J., & Benton, D. C. 1991, "Screening for undescended testes", *Archives of Disease in Childhood*, vol. 66, pp. 934-937.
- Sang Won, H., Tack L., Jang Hwan K., Seung Kang, C., Nam Hun, C., & Ji Young, H. 1999, "Pathological differences between retractile and cryptorchid testes", *The Journal of Urology*, vol. 162, no. 3, p. 878.
- Sinisi, A. A., Pasquali, D., Papparella, A., Valente, A., Orio, F., Esposito, D., Cobellis, G., Cuomo, A., Angelone, G., Martone, A., Fioretti, G. P., & Bellastella, A. 1998, "Antisperm antibodies in cryptorchidism before and after surgery", *Journal of Urology*, vol. 160, no. 5, pp. 1834-1837.
- Swerdlow, A. J., Higgins, C. D., & Pike, M. C. 1997, "Risk of testicular cancer in cohort of boys with cryptorchidism", *British Medical Journal*, vol. 314, no. 7093, pp. 1507-1517.
- Tamhne, R. C., Jarvis, S. N., & Waterson, A. J. R. 1990, "Auditing community screening for undescended testes", *Archives of Disease in Childhood*, vol. 65, pp. 888-890.
- Taskinen, S. & Wikstrom, S. 1997, "Effect of Age at Operation, Location of Testis and Preoperative Hormonal Treatment on Testicular Growth after Cryptorchidism", *The Journal of Urology*, vol. 158, no. 2, pp. 471-473.

6.3.14. URINARY ABNORMALITIES

6.3.14.1. BACKGROUND

Clinical picture

This section is a discussion of abnormalities of urine as detected by screening urinalysis. The many underlying conditions that result in urinary abnormalities are not dealt with specifically, as each could be the topic of a separate chapter. The general concept of 'screening urine for abnormalities' is discussed rather than screening for specific diseases by urinalysis.

Urinalysis has been suggested as a screening tool in childhood and has been included in routine health examinations in the past to detect the presence of glucose, protein, blood and bacteria. More recently, additional tests that indicate infection by detecting leucocytes and nitrites may also be included.

Glycosuria

Glycosuria is the presence of an abnormally high level of glucose in the urine. A common cause of glucose in the urine is diabetes mellitus, but it can also be due to renal glycosuria, a benign congenital disorder of the renal tubule with isolated defective reabsorption of glucose. Rare causes include Fanconi syndrome; inborn errors of metabolism; heavy metal poisoning; drugs; trauma or infection.

Proteinuria

Proteinuria is the presence of an abnormally high level of protein (usually albumin) in the urine. Proteinuria may be a finding in healthy children after exercise or during fever. Postural (orthostatic) proteinuria is a condition where otherwise healthy children excrete abnormally high levels of protein when standing, but not when supine. The pathological causes of proteinuria result from renal disorders, the most common being increased permeability of the glomerular wall. Glomerulonephritis, pyelonephritis and developmental abnormalities account for the majority of this group.

Haematuria

Haematuria is the presence of blood in the urine. The source of blood may be from anywhere in the renal tract – kidney, ureter, bladder or urethra. The causes are many and include glomerular diseases, infection, urolithiasis, anatomical abnormalities, haematological abnormalities, drugs and exercise. This discussion focuses on asymptomatic children, so macroscopic haematuria is not considered here. Criteria for the definition of microscopic haematuria varies, ranging from 5 to 20 RBC per high power field (Benbassat, Gergawi, Offringa, & Drukker 1996).

Bacteruria

Bacteruria is the presence of bacteria in the urine. This can be due to urinary tract infection (UTI); contamination of the specimen during or after collection; or 'asymptomatic bacteruria' (ABU). ABU is the presence of significant bacteria in the urine without symptoms of UTI.

Natural history

The natural history of abnormalities in the urine alone is not well understood. Positive results on a single test in childhood can be normal or transient findings, as well as possible indicators of underlying disease.

Benbassat et al state that the main finding of their review is the paucity of available data on the clinical significance of isolated haematuria in childhood (Benbassat, Gergawi, Offringa, & Drukker 1996).

Children with ABU have different bacteria isolated from their urine than those with UTIs, and it is postulated that perhaps there is a particular host-parasite interaction that results in the tendency for the bacteria to become less virulent with time (Hansson, Martinell, Stockland, & Jodal 1997). Children with ABU treated with antibiotics were found to have high relapse rates and high incidence of pyelonephritis and renal scarring, while those left untreated had no similar sequelae. Antimicrobial treatment alters the bacterial flora associated with ABU and may predispose the patient to infection with organisms related to renal pathology (Hansson et al. 1997; Kemper & Avner 1992).

Prevalence

Glycosuria

The prevalence of initial asymptomatic glycosuria has been reported to be 0.4%. (Kaplan cites the origin of this data from a 1976 paper by Gutgesell which was not sourced for this review) (Kaplan, Springate, & Feld 1997). No other information on the prevalence of glycosuria in childhood was found.

Proteinuria

In a large study of more than 12,000 children in Texas over a five year period from 1967 to 1972, Dodge et al found that 4.7% of girls and 1.4% of boys had significant proteinuria (>50 mg/dl) in 2 of 3 consecutive urine specimens (Dodge, West, Smith, & Bunce 1976). In the initial specimen, 8.5% of children had >10mg/dl of protein, however only half of these went on to have protein in a second or third specimen. The occurrence of proteinuria increased progressively with increasing age, and was significantly greater at each age for girls. This study included all children in the participating schools, including those known to have underlying disease. The prevalence of initial asymptomatic proteinuria has been reported as 6.3% (Kaplan, Springate, & Feld 1997).

Haematuria

In a review of the literature, Benbassat notes that the point prevalence of asymptomatic microscopic haematuria in school children is reported to be 0.25 – 1%. The variability is related to the criteria used for definition - >5 RBC/HPF is seen more frequently than 20 RBC/HPF (Benbassat et al. 1996). This review found some studies reported an increasing incidence with age, but not others. In the large Texan study reported above, the prevalence of haematuria varied by age, sex and definition of haematuria. As for proteinuria, only half of the children with a detected abnormality on the initial test had positive results in a second or third specimen. The occurrence of haematuria as defined by 10 or more RBC/HPF in 2 of 3 consecutive specimens was 2.2% in girls and 0.9% in boys.

The incidence of renal disease in children with asymptomatic microscopic haematuria varies with different studies (Benbassat et al. 1996). Ig A nephropathy is the most commonly reported biopsy-proven glomerular disorder found in 2.4%, 2.8% and 21.2% of patients in 3 respective studies. However, renal biopsies are rarely performed on this group of patients. Asymptomatic UTIs were the most common non-glomerular finding in 4.8% and 6.0% of patients in 2 reported studies. Benbassat et al state in their review that 'most children with microscopic isolated haematuria had no significant underlying disease. Some had disorders that may benefit from early treatment (membranoproliferative glomerulonephritis, obstructive uropathy, urolithiasis) or counselling (hereditary nephropathy, renal cystic disease). The combined prevalence of these 5 diseases was 0 - 7.2% in children detected by screening.'

Bacteruria

The prevalence of ABU varies with gender, age and race and is reported to vary from 2.5% in infancy to 0.14% at school age (Kemper & Avner 1992). In infancy, ABU generally occurs in boys in the first 6 months and in girls throughout the first year. Overall, girls had higher rates than boys (Hansson et al. 1997; Kemper & Avner 1992).

Genetics relevant to screening issues

None known

Diagnosis

Detection of abnormalities in the urine is the topic of this section and not diagnosis. No diagnosis of underlying disease can be made on the basis of urinalysis alone.

Glycosuria

No diagnosis should be made on the discovery of glycosuria on a single occasion and the test should be repeated. In childhood, diabetes usually presents acutely with the clinical symptoms of polydipsia, polyuria, weight loss, lethargy and weakness. It is therefore unlikely to be detected on screening urinalysis in asymptomatic children. However if diabetes is suspected, it can be confirmed by the presence of hyperglycaemia and, if appropriate, by glucose tolerance testing. If diabetes is not the cause of the glycosuria, further tests will be required to investigate the other possibilities.

Proteinuria

Many authors state that dipstick testing for proteinuria has a 'high' false positive rate. In addition, many children have transient proteinuria. This means that all children with an abnormality on dipstick testing must have the test repeated on a second occasion. A positive test is followed up with further urine examination, 24 hour urine collections to quantify the protein excretion and other tests considered relevant in the clinical scenario.

Haematuria

Quantification of blood on dipsticks is not accurate and should only be interpreted as negative or positive (Bergstein 2000). Microscopy is required to confirm the diagnosis, quantify the amount of blood present and identify the source eg red cell casts suggest that the blood came from the kidney. The test should be repeated. Further tests relevant to the potential underlying conditions will be required.

Bacteruria

The gold standard for diagnosis of UTI or ABU is urine culture. Due to the problem of contamination, the definition of UTI based on bacterial culture varies according to the method of collection. More than 10^5 colonies of pure growth are required for a definition of UTI from a 'clean catch' specimen, $>10^4$ for a catheter specimen, and many authors say any growth from a suprapubic aspirate is indicative of infection. Fewer colonies or mixed bacterial growth suggest that the urine may be contaminated.

Visualising bacteria and white blood cells on microscopic examination has been used in the past to indicate a UTI, however a recent systematic review has demonstrated that dipstick testing for leucocyte esterase or nitrite is superior to microscopy (Gorelick & Shaw 1999).

Treatment/Management

Treatment will depend on the specific underlying disease.

Prevention

Prevention will depend on the specific underlying disease.

How might screening reduce the burden of suffering?

This will depend on the underlying disease process

6.3.14.2 EVIDENCE

Tests

These tests are conducted using 'dip sticks' – small strips with patches of reagent attached. The sticks are dipped in the urine and the results read against a key, usually available on the container. Any colour change in the reagent patches can be compared with a scale on the key to determine whether the result is abnormal, and if so, to what degree. Confirmation of the urinary abnormality may require microscopy, bacterial culture or additional biochemical tests. Further tests will then be required to ascertain the underlying disease process.

Proteinuria

Kaplan reports the false positive/transient proteinuria dipstick rate as 89% based on Gutgesell's 1976 paper. (Kaplan, Springate, & Feld 1997)

Haematuria

Kaplan reports the false positive/transient haematuria dipstick rate as 88% based on Gutgesell's 1976 paper. (Kaplan, Springate, & Feld 1997) Examination of the urinary sediment by microscopy has low accuracy for measuring RBC. There are technical problems with consistency of microscopic examination related to the density and volume of the centrifuged specimen, the volume of the discarded supernatant fluid, the duration of centrifugation and the width of the sediment layered on the slide (Benbassat et al. 1996). These issues are less problematic in evaluation of a child with significant haematuria or other symptoms and signs of disease, but pose dilemmas for its use as a screening tool in the detection of asymptomatic haematuria.

Asymptomatic Bacteruria

Kemper reports both the sensitivity and specificity of screening for bacteruria as 80%, and using prevalence rates, calculates the positive predictive value as 0.1% in boys and 3.9% in girls, and the negative predictive value as greater than 99% (Kemper & Avner 1992). Based on the prevalence rate, 20 out of every hundred children screened would be labelled falsely with ABU and submitted to further investigation unnecessarily. Of those children with true ABU, 28% would be false negatives and reassured inappropriately.

Australian data

In their 1998 study of 9355 children, Hogg et al set out to measure the incidence of abnormalities in urinalysis and blood pressure in Australian preschoolers (Hogg, Harris, Lawrence, Henning, Wigg & Jureidini 1998). These findings are difficult to interpret in the light of screening programs for healthy children or testing for asymptomatic urinary abnormalities. All children in the group were tested and their results included in the figures although several were known to have existing pathology. For example, the 3 children with true positive tests for glycosuria were known to have diabetes, and 2 with UTI had previously diagnosed vesico-ureteric reflux.

- Glycosuria – 9 positive results, 3 children found to have clinical abnormality
- Proteinuria – 23 positive results, 7 children found to have clinical abnormality
- Haematuria – 216 positive results, 43 children found to have clinical abnormality

- Nitrites – 18 positive results, 10 children found to have clinical abnormality

Sensitivity, specificity and PPV are only available for their combined regimen of urinalysis plus BP measurement – 81%, 87%, and 20% respectively, but since the results include children known to have pre-existing disease they have limited applicability to this review.

Programs

Screening urinalysis is examination of the urine at the time of a routine ‘well child’ visit or opportunistically at the time of a clinic visit for an unrelated reason. No evidence of effectiveness for screening urinalysis was found in the course of this review.

Abnormalities on screening urinalysis have such a low prevalence, and since most diseases associated with these abnormalities present symptomatically in childhood, the benefits are very small. Conversely, the false positive and transient abnormality rates are very high, and when applied to a population with low prevalence of disease, huge numbers of children and their families will be subject to anxiety, discomfort, cost, etc.

Several studies indicate that treatment of children with ABU would seem to be harmful, with those who receive antimicrobial therapy going on to develop symptomatic infections, pyelonephritis and renal scarring (Hansson et al. 1997; Kemper & Avner 1992).

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test Cochrane & Holland 1971	heam	glyc	prot	bact
Simple, quick and easy to interpret	✓	✓	✓	✓
Acceptable to the public	✓	✓	✓	✓
Accurate	✗	✗	✗	✗
Repeatable	✗	✗	✗	✗
Sensitive	✗	✗	✗	✓
Specific	✗	✗	✗	✓

Criteria for a Screening Program Wilson & Jungner 1968	heam	glyc	prot	bact
Important health problem	✗	✗	✗	✗
Accepted treatment	Depends on diagnosis			
Facilities for diagnosis and treatment	✓	✓	✓	✓
Latent or early symptomatic stage	Depends on diagnosis			
Suitable test or examination	✗	✗	✗	✗
Test acceptable to the population	✓	✓	✓	✓
Natural history adequately understood	✗	?	✗	✗
Agreed policy on whom to treat	Depends on diagnosis			
The cost of case-finding balanced with expenditure on medical care as a whole	✗	✗	✗	✗

Cost effectiveness

Cost of an initial routine urinalysis on 100,000 children has been estimated at between \$US 251,000–324,000 (Kaplan, Springate, & Feld 1997) and \$US 625,000–1,950,000 (Kemper & Avner 1992). Follow up urine cultures on those identified as positive would bring the overall costs to \$US 1,085,991 – \$US 2,470,680 (Kemper & Avner 1992). This estimate does not include physician visits, therapy, diagnostic imaging or time missed from work and school.

Generalisability of evidence

This discussion should apply to screening urinalysis universally.

6.3.14.3. CONCLUSIONS

Fair evidence to recommend against screening

Comments:

There is no evidence that screening urinalysis prevents renal or other disease. There is information to suggest that treating asymptomatic bacteriuria is harmful. Screening urinalysis is costly to the community, may result in physical and psychological costs to the patients and their families, and has high rates of misclassification.

6.3.14.4. RECOMMENDATIONS

Screening programs of urinalysis in well children are not recommended.

6.3.14.5. FURTHER RESEARCH

None recommended.

6.3.14.6. SEARCH TERMS

MeSH headings: Urinary tract infections; Proteinuria; Glycosuria; Bacteriuria; Hematuria

Keywords: Renal abnormalities

6.3.14.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Reference group: Neil Wigg, David Hall

Clinical experts: Dr Colin Jones, Director, Dept of Nephrology, Royal Children's Hospital, Melbourne

6.3.14.8. REFERENCES

Benbassat, J., Gergawi, M., Offinga, M., & Drukker, A. 1996, "Symptomless microhaematuria in school children: causes for variable management strategies", *Quarterly Journal of Medicine*, vol. 89, no. 11, pp. 845-854.

Bergstein, J. M. Clinical evaluation of the child with hematuria. In: *Nelson Textbook of Paediatrics*, edited by Behrman RE, Kliegman RM, and Jenson HB, Philadelphia: WB Saunders Company, 2000, p. 1577-1578.

Dodge, W. F., West, E. F., Smith, E. H., & Bunce, H. 1976, "Proteinuria and hematuria in schoolchildren: Epidemiology and early natural history", *The Journal of Pediatrics*, vol. 88, no. 2, pp. 327-347.

Gorelick M. H. & Shaw K. N. 1999, "Screening tests Urinary Tract Infection in children: A meta-analysis", *Pediatrics*, vol. 104, no. 5, pp. 1-7.

Hansson, S., Martinell, J., Stockland, E., & Jodal, U. 1997, "The natural history of bacteriuria in childhood", *Infectious Disease Clinics of North America*, vol. 11, no. 3, pp. 499-512.

Hogg, R. J., Harris, S., Lawrence, D. M., Henning, P. H., Wigg, N., & Jureidini, K. F. 1998, "Renal tract abnormalities detected in Australian preschool children", *Journal of Paediatrics and Child Health*, vol. 34, no. 5, pp. 420-424.

Kaplan RE, Springate JE, & Feld LG 1997, "Screening Dipstick Urinalysis: A Time to Change", *Pediatrics*, vol. 100, no. 6, pp. 919-921.

Kemper, K. J. & Avner, E. D. 1992, "The case against screening urinalyses for asymptomatic bacteriuria in children.", *American Journal of Diseases of Children*, vol. 146, no. 3, pp. 343-346.

6.3.15. VISION DEFECTS

6.3 15.1 BACKGROUND

Clinical Picture

The visual system consists of the eye, the visual pathway (consisting of the optic nerve, optic chiasm, optic tracts, lateral geniculate nuclei, optic radiations) and the occipital cortex. Disorders of vision can be due to abnormalities in any one or more of these areas. Cortical vision impairment is poor vision due to malfunction of the visual cortex. This is now the commonest cause of blindness in childhood and is most commonly seen in the context of acquired brain damage, especially following extreme premature birth.

Even with an intact visual system, appropriate visual stimulation is needed for normal vision to develop. This requires focussed images on the retina as well as conjugate visual axes during the period of visual development. The two eyes compete with one another, especially in early infancy, to make connections on the occipital cortex, with the critical period for the development of vision in man extending from birth to about age eight years. The most rapid visual development occurs during the first three months of life.

Interference with vision during the critical period will prevent the attainment of normal visual acuity. This is known as **amblyopia** if it occurs in an eye which is ophthalmoscopically normal. Visual acuity is the resolving power of the eye, and is usually expressed as Snellen notation for each eye separately (6/6 representing normal acuity, and less than 6/60 representing legal blindness). Several reasons for poor visual acuity in childhood are relevant to screening; these include media opacities such as cataracts, high or asymmetric refractive errors, and strabismus (squint). Strabismus refers to misalignment of the eyes. The eyes may be crossed (convergent squint, or esotropia), turned out (divergent squint, or exotropia) or vertically misaligned (hyper- or hypotropia, both of which are uncommon in childhood). The squint can be intermittent or constant (manifest), and one eye may squint constantly or the squint may alternate between the two eyes (see Table 1 for terminology). These factors occurring singly or together in early childhood can produce amblyopia, so are considered risk factors for amblyopia or “amblyopiogenic factors”.

Table 1.Squint Terminology

Eso-:	a convergent deviation.
Exo-:	a divergent deviation.
Hyper-:	a vertical deviation (left or right used with this term denotes the higher eye).
-tropia:	a constant deviation of the eyes.
-phoria:	a latent deviation of the eyes, brought out when fusion is interrupted eg by the cover test.
Example	Left esotropia = a convergent squint present constantly in the left eye.
Alternating:	a squint occurring for an equal amount of time in each eye.
Manifest Squint:	a constant squint.
Comitant squint:	the angle is constant in all directions.
Noncomitant (incomitant)squint:	the angle varies in different positions of gaze eg sixth nerve palsy.

There is a great deal of evidence to support the idea of visual plasticity. Both animal and clinical experience demonstrate that vision development can be influenced negatively early in life, but if treatment occurs early enough the negative effects can be reversed (Daw 1995; Maurer, Lewis, Brent, & Levin 1999). This has led to strong beliefs that early detection by screening is needed to attain the goal of early treatment of risk factors.

A number of conditions need to be considered in vision screening, and screening activities for these currently occur across a wide range of ages. Broadly, these may be categorised (with some overlap) into:

1. Neonatal conditions (including high risk neonates/infants)
2. Risk factors for amblyopia
3. Visual acuity abnormalities
4. Colour vision abnormalities

1. Neonatal conditions

Congenital cataract (opacity in the normally clear lens), whether unilateral or bilateral, must be corrected by three months of age or irreversible amblyopia can develop. Therefore detection must occur within these first 3 months. Congenital glaucoma (raised intra-ocular pressure due to obstruction of the flow of aqueous into and out of the eye) is present at birth or often develops in the newborn period. Most cases have developed by one year of age. Congenital glaucoma leads to blindness if not treated early.

Cortical vision impairment is vision failure due to malfunction of the visual cortex. Patients with cortical vision impairment are almost always identified because of associated neurological disorder. Most patients are premature babies and as such will usually already be in a follow up program. Retinopathy of prematurity, formerly known as retrolental fibroplasia, is a disorder of the growing retinal vessels which occurs almost exclusively in premature babies born before 32 weeks gestation. It blinds 1-2% of premature infants of birth weight less than 1500 grams.

Retinoblastoma is a malignant tumour of the retina with an incidence of about 1/20,000 births (Sanders, Draper, & Kingston 1988). It may be present at birth and develops rapidly in the first two years or so of life. It is usually detected on clinical grounds before the age of two, usually presenting with a white pupil (leukocoria) due to light reflection off the tumour mass or with a marked squint. It may be sporadic, or be inherited on a dominant basis (though with only partial penetrance).

Nonparalytic (intermittent) strabismus in the newborn usually resolves spontaneously.

2. Risk factors for amblyopia

Amblyopia affects 2-6% of the population (Simons 1996). It is usually unilateral and caused by abnormal binocular interaction (eg a constantly crossed eye or a unilateral blurred image caused by anisometropia) or by stimulus deprivation. Organic lesions such as congenital cataracts and retinoblastoma may be identified by amblyopia screening. In these cases, however, not only will decreased vision be due to an ophthalmoscopically abnormal eye, but there will also be some degree of amblyopia. Whilst screening for vision disorders in older children must of necessity involve directly screening for amblyopia (by measuring visual acuity), in younger children where a direct measurement of acuity is not possible clinically, screening is for strabismus and other risk factors for amblyopia.

It is not known why some children with a refractive error or squint develop amblyopia, while others with apparently identical abnormalities do not. Amblyopia is more common in certain populations, such as premature infants and children with Down Syndrome. Because cortical plasticity ceases by about age eight, treatment of amblyopia is generally not possible beyond this period, though for most children with amblyopia identification in the preschool period still offers good chances of successful treatment.

3. Visual acuity abnormalities

With increasing age and general development, visual acuity can be assessed more accurately and reliably. Reduced visual acuity is often due to refractive errors, which are disturbances of the optical system of the eye such that a sharp image is not formed precisely on the retina. Specific types include myopia (short sightedness), hypermetropia (long sightedness), anisometropia (unequal refraction in the two eyes), and astigmatism (difficulty focusing in both planes simultaneously due to differences in the curvature of the cornea and lens in the horizontal and the vertical planes) (see Table 2)

Table 2: Refraction Terminology

Anisometropia:	A difference in the refractive error of the two eyes.
Accommodation:	The adjustment of the focus of the eye for different distances to enable a sharp image to be formed on the retina. This is carried out by altering the shape of the lens.
Astigmatism:	A refractive error which prevents light rays from coming to a single focus on the retina because the light rays are refracted differently by different meridians of the cornea. An astigmatic surface can best be conceptualised like the back surface of a spoon; curved more in one plane than another. Another way to understand astigmatism is to think that the front of an astigmatic is shaped like a football and not a soccer ball. Astigmatism is very common in the first year of life.
Emmetropia:	Light rays from infinity (ie parallel) are brought to a focus on the retina without the use of accommodation.
Hypermetropia:	(Long sightedness) is present if the focal point of light rays from infinity is behind the retina. In this refractive state accommodation is necessary to bring light rays from

infinity to a focus of the retina. Mild hypermetropia is the rule in infancy and early childhood.

Myopia: (short-sightedness) is present when light rays from infinity come to a focus in front of the retina. Light rays from near objects are divergent so they will come to a focus on the retina of a myopic eye without the need for accommodation. Light rays from distant objects can only be focussed on the retina with help of glasses or contact lenses containing concave (diverging) lenses.

Anisometropia is the commonest cause of non-strabismic amblyopia, and usually develops between the ages of two and four years. Unless severe, myopia or hypermetropia do not usually cause amblyopia provided the refraction is roughly the same in both eyes, though they may preclude or limit activities such as near or far reading.

Though isolated refractive error may have an educational impact, firm evidence is lacking to link poor visual acuity (of a degree usually only picked up by vision screening) to poor school performance.

4. Colour vision abnormalities

Persons with colour vision defects (usually affecting red-green discrimination, and of varying severity) generally do not become aware of their problem until later in childhood. There is not good evidence that impaired colour vision affects learning. There are some occupations from which people with impaired colour vision or visual acuity are excluded (for example, persons affected by red-green colour blindness cannot become electricians or train drivers.)

Natural history

The neonatal conditions for which vision screening can occur are usually associated with blindness or other significant morbidity. They have a major impact unless detected very early and intervention provided.

Determining the natural history of risk factors for amblyopia (such as squint and refractive error), amblyopia itself, and other causes of decreased visual acuity has been more difficult. The review by Snowdon only found methodologically limited studies examining this issue (Snowdon & Stewart-Brown 1997). There were reported gaps in the data, but their conclusion was that it did not support the need to treat all children with these conditions identified at three to four years of age. Some children improved spontaneously, and groups who were treated did not necessarily have lower prevalence of problems at follow up.

This review has subsequently been heavily criticised. One criticism is that that poor outcomes are often due to lack of compliance with treatment. Though one of the points being made by Snowdon is that there can be major differences between the potential effectiveness of an intervention, and its efficacy in the general clinical situation. There is also other evidence that the prevalence of amblyopia does in fact decrease as a result of screening (Kvarnstrom, Jakobsson, & Lennerstrand 1998; Vinding, Gregersen, Jensen, & Rindziunski 1991).

There is no good evidence describing the degree of functional impact due to decreased visual acuity in one eye in childhood. However, whilst amblyopia as a rule only affects one eye, the loss of binocularity is a bar to some occupations and the loss of the remaining sighted eye can be disastrous. There is some evidence that individuals with amblyopia are at greater risk of becoming blind as a result of trauma or changes associated with aging affecting the better eye, but in this particular study there was likely significant selection bias (Tomilla & Tarkkanen 1981). Surgery for senile cataracts is generally delayed for as long as possible in subjects with ocular comorbidities because of the risk of blindness associated with the procedure (Lee, Kamberg, Hillborne et al 1993).

Prevalence

Disorders of vision in childhood can be subdivided into two categories (Hall 1996):

- serious defects likely to cause a disabling impairment of vision, ranging from partial sight to complete blindness (2–5 cases per 100,000 births). Cortical vision impairment is now the commonest cause of bilateral poor vision in childhood in Western countries. Childhood cataracts occur 1 in 2,500 live births, but often do not result in blindness.
- common and usually less incapacitating defects, including refractive errors, squints, amblyopia, and abnormalities of colour vision. The prevalence of amblyopia is quoted as 2-6%, but varies depending upon the age of the population studied and on the method of diagnosis. This population overlaps with children afflicted by strabismus. Approximately 7% of the male population have congenitally impaired colour vision, and about 0.5% of girls.

Genetics related to screening issues

There are known genetic contributions to the development of refractive errors, squint, and colour vision, but these are not currently of relevance to screening. Genetic testing for carrier status of retinoblastoma will shortly be available when one case has already occurred in a family.

Diagnosis

There are numerous vision assessment and vision screening tests for use in children. Their nature is very much influenced by the child's developmental stage (and therefore the extent to which they can cooperate and follow instructions) as well as the condition to be identified.

1. Neonatal conditions

Congenital cataract is detected by examining the fundus red reflex in dim illumination using a direct ophthalmoscope. Congenital glaucoma causes an enlarged eye (buphthalmos), often though not always accompanied by a cloudy cornea. These changes are detected by external examination of the eyes. Any child with a corneal diameter greater than 12.5mm should be considered to have congenital glaucoma until proven otherwise. Retinoblastoma is suggested through viewing the retina at arms length using a direct ophthalmoscope and obtaining a white instead of red reflex (leukocoria). Leukocoria is the most common presentation of retinoblastoma, but leukocoria is most commonly due to congenital cataract.

Serious vision impairment in high risk children is usually identified through formal ophthalmological examination of at-risk children (such as very low birth weight babies, or infants with other significant disabilities affecting the nervous system).

2. Risk factors for amblyopia

Amblyopia is diagnosed through demonstrating subnormal visual acuity and the associated reason for its occurrence. Measurement of refractive errors requires expert examination. This requires considerable clinical training and experience or the use of expensive apparatus (autorefractor).

Diagnosis of strabismus is performed clinically using a corneal light reflections or Hirschberg Test (a test where the observer compares the position of a projected light on the corneal surface of the two eyes), or cover test (alternately placing an opaque occluder in front of each eye and watching whether or not the uncovered eye takes up fixation). Photoscreening is a photographic technique that can be used and is discussed below.

3. Visual acuity abnormalities

Accurate determination of visual acuity requires a cooperative and understanding child. It is commonly measured using charts of letters or other geometric forms. Snellen and logMAR are two commonly used scales in older children and adults. Children older than three and a half can be tested using HOTV letters with confusion bars or the Lea (LH) symbols (Hered, Murphy, & Clancy 1997). Potential problems relate to the responses having a major behavioural and developmental component. It is difficult to measure the visual acuity of most children younger than three and a half years, but most children older than this can be reliably tested (Friendly 1978). Behavioural observations, such as fixing/following small objects, may be used in younger children, but these are less reliable and more difficult to interpret.

4. Colour vision abnormalities

These are usually diagnosed using colour vision plate tests based on pseudoisochromatic principles. This can be reliably carried out in the preschool period, though mostly children are screened at late primary or early secondary school level.

Treatment/management

1. Neonatal conditions (including high risk neonates/infants)

Untreated, almost all retinoblastomas are fatal. There are better treatment outcomes when treating small tumours (by laser, cryotherapy and/or chemotherapy) than treating late presenting large tumours (where the sight may already be lost and the eye may need to be enucleated, and as well the tumour spread beyond the eye and producing a fatal outcome).

There is effective treatment currently available for congenital cataracts (ie lensectomy/vitrectomy), congenital glaucoma (ie surgery) and retinopathy of prematurity (ie laser photocoagulation). There is currently no treatment for cortical vision impairment.

2. Risk factors for amblyopia

Refractive errors without amblyopia can be corrected with the use of optical lenses. In their 1997 systematic review of preschool vision screening (children between 3 and 4 years of age) Snowdon & Stewart-Brown

questioned the significance of isolated nonsymptomatic refractive errors that cause reduced visual acuity in preschool children (Snowdon & Stewart-Brown 1997). They concluded that correction of refractive errors resulted in a definite improvement in visual acuity, but the evidence did not warrant the identification and treatment of minor errors before school entry. There is disagreement over detail in recommending which degree of a particular refractive error to treat.

Early treatment of strabismus will produce better outcomes as far as binocularity is concerned. Strabismus surgery per se does not correct amblyopia except in a small percentage of cases. It is generally considered better to treat amblyopia before undertaking surgery for strabismus.

3. *Visual acuity abnormalities*

Snowdon & Stewart-Brown reviewed the effectiveness of therapy for established amblyopia, usually by occlusion of the non-affected eye in some manner (Snowdon & Stewart-Brown 1997). They noted that the natural history of the condition was not well understood and that most of the studies reviewed did not have untreated controls. There is a concern in the ophthalmology community, however, that it would be unethical to deny or significantly delay treatment (Simons & Preslan 1999). Many of the children with amblyopia in the studies reviewed did improve with treatment, but the review authors could not say whether the treatment itself made a difference. The extent to which any benefits might be maintained was not consistent in the small number of studies that offered follow up (usually short term for 1 to 3 months). Once again, there is a concern that the authors of this review excluded important studies (eg Scott & Dickey 1988; Lithander & Sjostrand 1991) and that there is more recent evidence which supports the efficacy of amblyopia treatment (Bowman, Williamson, Andrews, Aitchison, & Dutton 1998). This other evidence, however, has not been subjected to critical systematic review.

4. *Colour vision abnormalities*

There is no treatment for colour vision abnormalities. There is not good evidence that the identification of impaired colour vision improves learning in any way.

Prevention

Early detection and treatment of strabismus and refractive errors in some children may prevent development of amblyopia. Prevention of the development of amblyopia may further prevent blindness due to damage to the one good eye later in life.

Blindness from congenital cataract, congenital glaucoma and retinopathy of prematurity is largely preventable with early detection and treatment.

How might screening reduce the burden of suffering?

Congenital cataract, congenital glaucoma and retinopathy of prematurity are blinding disorders if undetected in early infancy, and each is able to be treated. The treatment of retinoblastoma is more effective and with better outcomes if carried out early (enucleation can be avoided and some sight retained).

There is no evidence that detection of asymptomatic refractive errors not associated with amblyopia improves outcomes.

Amblyopia screening in the preschool period should detect most children with significant refractive errors. There is no good evidence that decreased visual acuity in one eye in childhood is of great functional impact. Effective screening and treatment would mean fewer older people with only one good eye, for whom loss of binocularity may be a problem and loss of the remaining sighted eye a disaster.

Strabismus, as well as causing amblyopia, may also be a significant cause of psychosocial disability. However, squints of this degree would usually be detected without screening so would not be expected to alter this outcome greatly (Burke, Leach, & Davis 1997; Olitsky et al 1999; Satterfield, Keltner, & Morrison 1993). Early correction of infantile esotropia produces better functional outcomes (Birch, Fawcett, & Stager 2000).

There are some occupations from which people with impaired colour vision or visual acuity are excluded. It is possible that early career counselling might be of benefit to a small number of people.

6.3.15.2 EVIDENCE

Tests

NEONATAL CLINICAL SCREENING

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	X
Acceptable to the public	√
Accurate	?
Repeatable	?
Sensitive	?
Specific	?

There are a number of serious ocular conditions that can be detected in the neonatal period. A large proportion are identified to be of high risk (eg premature) and consequently are referred for formal ophthalmological examination. Neonatal ocular examination can be difficult, and there is little evidence enabling determination of how well it is carried out in the context of screening all newborn children by non eye specialists.

SCREENING FOR RISK FACTORS FOR AMBLYOPIA (PHOTOSCREENING)

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	√
Acceptable to the public	√
Accurate	?
Repeatable	?
Sensitive	X
Specific	X

Photoscreening is a photographic technique used to detect risk factors for amblyopia such as strabismus, media opacities, high refractive errors or anisometropia. The refractive state of each eye and its alignment is semi-quantitatively assessed by photographing the pattern of light returned from the pupil, mostly using eccentric flash photography.

Detection of strabismus and refractive errors is possible by photoscreening but there is as yet no photoscreener with sufficient sensitivity and specificity to recommend the adoption of this technology for community based screening, even though it is in widespread use in the United States. A recent Australian study (Cooper, Gole, Hall, Colville, Carden, & Bowling 1999) evaluating two photoscreeners found that they performed unsatisfactorily in a study population aged 1-4 years with a high prevalence of amblyogenic risk factors, mainly because of low sensitivities (<68%). The NEI Task Force on Preschool Vision Screening does not support photoscreening because of a lack of validation studies of most instruments (Hartmann, Dobson, Hainline, Marsh-Tootle, Quinn, Ruttum, Schmidt & Simons 2000). There are almost no published studies on performance of photoscreeners not carried out by the developer of the particular device (with attendant bias consequent on a declared or undeclared commercial interest). There are no studies of treatment outcomes consequent on the detection using photoscreening of children with amblyopogenic risk factors.

The possibility of targeted screening of higher risk populations (such as those identified through parental concerns questionnaires) has not been explored, but may be expected to improve test performance (Cooper et al. 1999).

VISUAL ACUITY SCREENING

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	√*
Acceptable to the public	√
Accurate	√*
Repeatable	√*
Sensitive	√*
Specific	√*

*The nature and performance of visual acuity screening tests are very much influenced by the developmental age of the child being assessed. Visual acuity screening tests in the preschool child have variable test characteristics and reliability, though these do improve with the use of more highly skilled and trained screeners (such as orthoptists compared with community nurses). They also improve with increasing age of the child (Friendly 1978).

COLOUR VISION SCREENING

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	√
Acceptable to the public	√
Accurate	√*
Repeatable	√*
Sensitive	√*
Specific	√*

*Colour vision screening is more accurate in the older child. There are many colour vision screening tests. They vary in their characteristics quite considerably, with only some being acceptable.

Programs

Vision screening programs have a multitude of forms. Different ages of children being screened, different tests used, and different levels of skill and expertise in the screeners are key areas of difference. Most countries have some form of vision screening program.

Formal consideration of the evidence in support of vision screening before embarking on community programs would appear not yet to have made a major impact on vision screening.

There is a reasonable amount of evidence about how well screening tools function in specific groups, and some information about how they function in community programs. There is good evidence that screening leads to the detection of refractive errors in children who would not otherwise have been detected. Unfortunately, there are gaps in the information about the benefits of such detection. Absence of evidence for a benefit from vision screening, by way of 'gold standard' randomized controlled trials, however, does not constitute proof that screening and early treatment has no value. Two studies from Scandinavia show reduced prevalence of amblyopia in population groups following screening (Kvamstrom, Jakobsson, & Lennerstrand 1998; Vinding et al. 1991)

There is reasonable evidence in some countries that separate orthoptic screening programs in the preschool child usually have uptake levels lower (average 64.8%) than vision screening programs provided by health professionals as part of ongoing health monitoring (average 76.2%) (Snowdon & Stewart-Brown 1997). However, there is considerable variability in the yield of vision abnormalities detected (2.4 to 6.1% for all conditions screened combined). Positive predictive values for orthoptist screening tend to be higher than those for primary care givers.

We identified only one 1997 systematic review of the effectiveness of preschool (3-4 year old children) vision screening (Snowdon & Stewart-Brown 1997). On the basis of a general lack of good quality evidence in many different areas, the authors did not feel that new vision screening programs should be implemented. They suggested that there should be consideration to discontinuing existing programs unless they were part of a controlled trial of treatment. The validity of this review has been hotly debated in the literature and its

recommendations found support from neither practising ophthalmologists nor from an expert panel convened by the US National Eye Institute to review preschool vision screening (Hartmann 2000). Many of the criticisms relate to the fact the clinicians feel very strongly that certain diagnostic modalities and therapies are beneficial and that to withhold them would be harmful, despite their lack of demonstrated efficacy.

Potential specific harms of vision screening programs include children being treated for visual defects (patching for amblyopia, spectacles for myopia etc) unnecessarily, with attendant costs and possible psychological sequelae.

NEONATAL SCREENING

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	?*
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?

*Though information relating to any screener is lacking, general belief is that the test or examination would only be suitable if carried out by well trained and skilled personnel.

SCREENING FOR RISK FACTORS FOR AMBLYOPIA (PHOTOSCREENING)

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓/✗
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✗
Test acceptable to the population	✓
Natural history adequately understood	✗
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	?

VISUAL ACUITY SCREENING

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	√/X
Accepted treatment	√
Facilities for diagnosis and treatment	?
Latent or early symptomatic stage	√
Suitable test or examination	√*
Test acceptable to the population	√
Natural history adequately understood	X
Agreed policy on whom to treat	X
The cost of case-finding balanced with expenditure on medical care as a whole	?

*Age dependant

COLOUR VISION SCREENING

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	X
Accepted treatment	X
Facilities for diagnosis and treatment	X
Latent or early symptomatic stage	X
Suitable test or examination	√
Test acceptable to the population	√
Natural history adequately understood	X
Agreed policy on whom to treat	X
The cost of case-finding balanced with expenditure on medical care as a whole	X

Cost effectiveness

There is no evidence regarding the cost effectiveness of early detection and treatment of amblyopia.

Quality of evidence

Quality issues for two systematic reviews are discussed. No randomised control trials of vision screening and therapy were identified. No post 1995 studies (not included in the Snowdon et al review), pseudo RCTs or cohort studies considering vision screening and therapy were identified.

PAPER	STRENGTH OF EVIDENCE		SIZE OF EFFECT		RELEVANCE OF EVIDENCE
	Level	Quality		Statistical precision	
Kemper AR, Margolis PA, Downs SM, Bordley WC. A systematic review of vision screening tests for the detection of amblyopia. <i>Pediatrics</i> 1999;104:122-1222.	I	<p><i>1. Clearly defined question and inclusion criteria</i></p> <p>B (some limitations) This review was published in 1999 and aimed to inform whether there were vision screening tests that could be recommended for primary care practitioners for use in children 2 to 5 years of age. The question was stated, but did not include clear description of the target population, which screening tests were to be considered (there were three different test reviewed), and the specific conditions to be screened for. Primary care practitioner is not defined. Inclusion criteria were noted, but it was not stated how it was determined that the vision screening test was available to the practicing physician and that it did not require procedures uncommon in the primary care setting.</p> <p><i>2. Comprehensive search</i></p> <p>C (unsatisfactory quality) The search strategy used only the Medline database and excluded non-English papers. The</p>	n/a	n/a	<p>The reviewers concluded that none of the papers considered took place in primary care settings with tools commonly available to primary care practitioners, and all used different ophthalmological standards.</p> <p>The reviewers also identified that all the studies appeared to have been performed in high risk populations, and the prevalence of failure of the ophthalmological examination was high.</p> <p>The reviewers concluded that inadequate information was provided to determine whether any of the screens evaluated would be worth recommending for the primary care practitioner.</p>

bibliographies of the identified papers were checked, but there was no other tracking down of studies or authors.

3. Critical appraisal of the validity of studies reviewed

B (some limitations) Validity was determined by looking for ‘gold standard’ examination of all or a random sample of all screened, blinding of the persons doing the screening and doing the examinations, and clearly defined criteria for abnormal ophthalmological examination. There were two independent reviewers and differences were resolved by consensus. No rating scale was used.

4. Consistency of results

B (some limitations) There was moderate homogeneity of results.

5. Number of studies included in review

Four studies were included.

Snowdon, S. & Stewart-Brown, S. 1997, "Preschool vision screening", *Health Technology Assessment (South Hampton, New York)*, vol. 1, no. 8, pp. 1-85..

Level III-2
Evidence obtained from comparative studies with concurrent controls and allocation not randomised

1. Clearly defined question and inclusion criteria

n/a

N/a

A (high quality) This review was conducted in 1996 with the aim of providing evidence-based recommendations on the future provision of preschool (3 to 4 year old) vision screening. There were a number of questions considered:

The following conclusions were reached in relation to each of the broad research questions used:

- No studies were found with the primary aim of establishing the prevalence of visual defects in preschool children.
- No studies were found with the primary aim of documenting the natural history of

(cohort studies), case control studies, or interrupted time series with a control group

the prevalence of vision disorders, their natural history, the resultant disability, effectiveness of treatment, and how well screening programs work. The questions were clearly stated and inclusion criteria provided.

2. Comprehensive search

A (high quality) The search strategy was comprehensive.

3. Critical appraisal of the validity of studies reviewed

B (some limitations) Critical appraisal of the studies was undertaken, but the criteria were not explicitly stated.

4. Consistency of results

B (some limitations) There was moderate homogeneity of results.

5. Number of studies included in review

Eighty five studies were included in the main analysis.

the target conditions in children 3 to 4 years old.

- A number of studies were found examining disability, but none were sufficient to enable firm conclusions about their impact on quality of life to be determined.
- Treatment studies were identified, though none compared treatment with no treatment. Most were methodologically flawed.
- There were a number of studies of treatment programs, of varying effectiveness. There was increased yield and positive predictive value when orthoptists were screening compared to health visitors or general practitioners.

The authors recommended against implementing new preschool vision screening programs unless they have been rigorously evaluated. They suggested that there should be consideration to discontinuing existing programs unless they were part of a controlled trial of treatment.

6.3.15.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against neonatal screening

Fair evidence to recommend against screening for risk factors for amblyopia

Insufficient evidence to make a recommendation for or against preschool visual acuity screening

Fair evidence to recommend against colour vision screening

Comments:

There is some evidence about the impacts of different disorders of vision, and that some interventions are effective. There is some evidence as to how vision screening tests function in selected samples, but less as to how they function in community samples. There is a dearth of high quality evidence about the effectiveness of vision screening programs coupled with intervention services.

A number of peak bodies currently support vision screening in children. Recommendations generally combine current understanding about the natural history and effect of different vision disorders, “common sense”, and lower levels of evidence.

There is a lack of good evidence that clinical examination of the neonatal eye is effective. The seriousness of the conditions being screened for, the agreement on interventions, and the improved effectiveness of sufficiently early interventions support the currently recommended practice of eye examination in the newborn. It may not be ethical to carry out an RCT in such a situation. However, examination requires considerable skill. There is little evidence as to how well this examination is currently carried out at program level and therefore its effectiveness.

Early discharge (<72 hours post partum) is now commonplace in Australia and issues related to this need to be addressed.

6.3.15.4. RECOMMENDATIONS

1. Although there is little firm evidence to support the value of screening, we recommend continuation of specific examination of the eye at the newborn check, *provided* it is in the context of an adequate early detection program or system.

Such a system should include:

- clear, written examination protocols
- appropriate training of examining medical staff
- clear pathways and/or guidelines for referral
- clear pathways for communication between health professionals in maternity settings and those in the community (particularly GPs and child health nurses)
- clear documentation that examination has occurred, its findings, and course of action taken in hospital and parent held health records

Because neonatal eye examination is difficult, minimum program standards of coverage and accuracy should be set and evaluated.

The examination should include external examination of the eye with particular attention paid to corneal size and clarity. The fundus red reflex should be examined in dim illumination with a direct ophthalmoscope to exclude media opacities. Examination should be carried out by well trained and skilled personnel as part of the neonatal examination.

Examination of high risk neonates by ophthalmologists should continue.

2. Screening for risk factors for amblyopia should be reviewed as photoscreening improves. Without further evidence, such programs are not recommended.

3. Programs of preschool or school entry visual acuity screening should not be instituted. What to do when such programs are already in existence is more problematic. Further research into visual acuity screening is required (see below). On the basis of current evidence, we suggest that in such cases there be community discussion and education. The aim should be to replace visual acuity screening programs with greater support for visual acuity evaluation as part of a comprehensive assessment when children are identified to have a learning or behavioural problem.

4. Without further evidence, screening for abnormalities of colour vision is not recommended.

6.3.15.5. FURTHER RESEARCH

- Assessment of the extent of disability and burden of disease attributed to amblyopia, and the benefits and harms of amblyopia treatment.
- A clearer understanding of the relative risks associated with different levels of refractive abnormalities (and the potential use of different cutoff levels in screening) such as hypermetropia, astigmatism and myopia.
- The potential role of family history and parental observation questionnaire instruments in targeted screening in the Australian context.
- Understanding about the current level of 'usual care' in order to be able to evaluate the added effects of vision screening programs.
- High quality evaluations of selected screening measures (such as the HOTV chart with confusion bars), and of the most appropriate age(s) for screening.
- If such evidence suggested that vision screening were warranted, and that there were vision screening tests with adequate characteristics, then further controlled research (preferably by RCT) in community samples would be required.

6.3.15.6. SEARCH TERMS

Vision

Eye diseases

Diagnostic techniques, ophthalmological

Visual

6.3.15.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Glen Gole, who has had significant input

James Elder

Deborah Colville

6.3.15.8. REFERENCES

Birch, E. E., Fawcett, S., & Stager, D. R. 2000, "Why does early surgical alignment improve stereoacuity outcomes in infantile esotropia.", *JAAPOS*, vol. 4, pp. 10-14.

Bowman, R. J., Williamson, T. H., Andrews, R. G., Aitchison, T. C., & Dutton, G. N. 1998, "An inner city preschool visual screening programme: long-term visual results", *Br.J Ophthalmol.*, vol. 82,no. 5 , pp. 543-548.

Burke, J. P., Leach, C. M., & Davis, H. 1997, "Psychosocial implications of strabismus surgery in adults.", *Journal of Pediatric Ophthalmology Strabismus*, vol. 34, pp. 159-164.

Cooper, C. D., Gole, G. A., Hall, J. E., Colville, D. J., Carden, S. M., & Bowling, F. G. 1999, "Evaluating photoscreeners II: MTI and Fortune", *Australian and New Zealand Journal of Ophthalmology*, vol. 27, pp. 387-398.

Daw, N. 1995, "Modifications to the Visual Input That Leads to Nervous System Changes.," in *Visual Development*, Plenum Press, New York, pp. 95-107.

Friendly, D. S. 1978, "Preschool visual acuity screening tests", *Trans.Am.Ophthalmol.Soc.*, vol. 76, pp. 383-480.

Hall, D.M.B. 1996, *Health for All Children*, 3rd edn, Oxford University Press, New York.

- Hartmann, E. E., Dobson, V., Hainline, L., Marsh-Tootle, W., Quinn, G. E., Ruttum, M. S., Schmidt, P. P., Simons, K. 2000, "Preschool Vision Screening: Summary of a Task Force Report", *Pediatrics*, vol. 106, pp. 1105-1116.
- Hered, R. W., Murphy, S., & Clancy, M. 1997, "Comparison of the HOTV and Lea Symbols charts for preschool vision screening", *J Pediatr Ophthalmol.Strabismus*, vol. 34,no. 1, pp. 24-28.
- Kemper, A. R., Margolis, P. A., Downs, S. M., Bordley, W. C. 1999, "A systematic review of vision screening tests for the detection of amblyopia", *Pediatrics* vol. 104, pp. 1220-1222.
- Kvarnstrom, G., Jakobsson, P., & Lennerstrand, G. 1998, "Screening for visual and ocular disorders in children, evaluation of the system in Sweden", *Acta Paediatr*, vol. 87,no. 11, pp. 1173-1179.
- Lee, P. P., Kamberg, C. J., & Hillborne, L. H., 1993, *Cataract surgery: a literature review and ratings of appropriateness and cruciality*. Rand, Santa Monica.
- Lithander, J. & Sjostrand, J. 1991, "Anisometric and strabismic amblyopia in the age group 2 years and above: a prospective study of the results of treatment", *British Journal of Ophthalmology*, vol. 75,no. 2, pp. :111-1116.
- Maurer, D., Lewis, T. L., Brent, H. P., & Levin, A. V. 1999, "Rapid Improvement in the Acuity of Infants after Visual Input.", *Science*, vol. 286,no. 10, pp. 108-110.
- Olitsky, S. E., Sudesh, S., Graziano, A., Hamblen, J., Brooks, S. E., & Shaha, S. H. 1999, "The negative psychosocial impact of strabismus in adults.", *JAAPOS*, vol. 3, pp. 209-211.
- Sanders, B. M., Draper, G. J., & Kingston, J. E. 1988, "Retinoblastoma in Great Britain 1969-80: incidence, treatment, and survival.", *British Journal of Ophthalmology*, vol. 72,no. 8, pp. 576-583.
- Satterfield, D., Keltner, J. L., & Morrison, T. L. 1993, "Psychosocial aspects of strabismus study.", *Archives of Ophthalmology*, vol. 111, pp. 1100-1105.
- Scott. W. E. & Dickey, C. F. 1988, "Stability of visual acuity in amblyopic patients after visual maturity", *Graefes Archive for Clinical and Experimental Ophthalmology*, vol. 226, pp. 154-157.
- Simons, K. 1996, "Preschool vision screening: rationale, methodology and outcome.", *Survey of Ophthalmology*, vol. 41,no. 1, pp. 3-30.
- Simons, K. & Preslan, M. 1999, "Natural history of amblyopia untreated owing to lack of compliance", *Br.J Ophthalmol.*, vol. 83,no. 5, pp. 582-587.
- Snowdon, S. & Stewart-Brown, S. 1997, "Preschool vision screening", *Health Technology Assessment (South Hampton, New York)*, vol. 1,no. 8, pp. 1-85.
- Tomilla, V. & Tarkkanen, A. 1981, "Incidence of loss of vision in the healthy eye in amblyopia.", *British Journal of Ophthalmology*, vol. 65, pp. 575-577.
- Vinding, T., Gregersen, E., Jensen, A., & Rindziunski, E. 1991, "Prevalence of amblyopia in old people without previous screening and treatment. An evaluation of the present prophylactic procedures among children in Denmark", *Acta Ophthalmol.(Copenh)*, vol. 69,no. 6, pp. 796-798.

6.3.16. DENTAL HEALTH

6.3.16.1 BACKGROUND

Clinical picture

Oral disease is a term used to encompass the dental health conditions of dental caries, periodontal disease, dental erosion, oral cancers and oral trauma.

Dental caries: The focus of this review, dental caries (also known as dental decay) is the main dental health condition that affects children and is one of the most common infectious diseases known in humans. Dental caries are caused by bacteria present in the oral cavity converting fermentable carbohydrates (mainly sucrose) into acid which then demineralises the dental hard tissues (both enamel and dentine) (US Preventive Services Task Force 1996). There is also evidence that the bacteria which induces dental decay can be passed from mother to child (personal communication with Nicky Kilpatrick).

Periodontal disease: Periodontal or gum disease is common in the community and its prevalence increases with age. It is rarely seen in young children, but almost 50% of adolescents have some signs of periodontal disease (Barnard 1993 as cited in Health Development Section 1999).

Dental erosion: Dental erosion is irreversible loss of tooth substance brought about by a chemical process that does not involve bacterial action. Essentially this disease is caused by acids introduced into the mouth either from external sources such as carbonated drinks and fruit juices, or from the contents of the stomach such as in children who reflux or those with eating disorders. There are no prevalence figures for Australian children. In the 1993 National Survey in the UK 52% of 5 year olds and 27% of 12 year olds had evidence of erosive wear on their upper incisors (O'Brien 1994). Current opinion suggests that dental erosion is on the increase and is missed by many dental health professionals (personal communication with Nicky Kilpatrick).

Oral cancers: Oral cancers are rarely seen in children or adolescents. They mainly affect adults with smoking and alcohol consumption both thought to be risk factors for their development (Brugene, Gueneni, Leder & Rodriguez 1986 as cited in Health Development Section 1999).

Oral trauma: More than 40% of oral trauma or dental injury presenting to Victorian public hospitals between July 1995 and June 1997 occurred in children under the age of 10 years, with a further 12% occurring in 10-14 year olds (Monash University Accident Research Centre, unpublished data as cited in Health Development Section 1999).

Natural History

The development of dental caries begins when bacteria contained in plaque (particularly streptococcus mutans and lactobacilli) or cariogenic foods (particularly fermentable carbohydrates) are in contact with the tooth surface and begin to demineralise the tooth enamel, the protective layer of the tooth. The bacteria then penetrates the dentin under the tooth enamel and progresses through to the soft pulp tissue (Centre for Disease Control and Prevention 2000). The amount of time the plaque can remain on the tooth surface before demineralisation begins is not known, and likely differs between individuals and different tooth surfaces. Certain individuals appear more susceptible to dental caries than others, although the exact reasons for this remain unclear. Caries on occlusal surfaces tend to be smaller in size and to progress more slowly than caries on proximal surfaces of the teeth (Widmer & Mekertichian 1996).

Prevalence

Just over half (53%) of all Victorian primary school children examined by the Victorian School Dental Service have some sign of dental caries. Close to 80% of the caries identified in 5 year old children in Victoria were untreated (Dental Health Services Victoria 1999 as cited in Health Development Section 1999) and nearly 40% of caries in 5 year old South Australian children were untreated (Slade, Spencer, Davies, Burrow 1996). Forty-two percent (n=650) of 4-6 year old school children in the Wallasey area of the UK were found to have untreated caries during a routine dental screening examination (Zarod & Lennon 1992). In the US, although the prevalence of caries in children has decreased, the 1986-7 national survey found that on average by the age of nine, children have at least one dental carie in their permanent teeth and an average of three caries in permanent teeth by the age of 12. However approximately 50% of 5-17 year olds have no dental caries (Public Health Service 1989 as cited in US Preventive Services Task Force 1996).

Genetics relevant to screening issues

None known

Diagnosis

Oral dental examination is the main form of diagnosis for dental caries. Oral x-ray may also be used, particularly where caries are suspected between the teeth where visibility is limited. Oral examination rarely identifies more than 50% of the approximal lesions in children which can be detected by the additional use of bitewings (oral x-ray) (Kidd & Pitts 1990). Many caries in the primary dentition go clinically undetected as they mainly occur on the large flat contact areas between the teeth. In contrast, most caries in the permanent dentition occur on the pitted occlusal (biting) surfaces of the molar teeth (Dummer, Addy, Olevier & Shaw 1988).

Treatment/Management

The main treatment for dental caries is the restoring or filling of the tooth with dental amalgam. However, there is debate over whether any treatment should be applied to the deciduous teeth (Baird 1999). If it is considered necessary to only treat caries in the permanent teeth, the case for screening young children (those with only deciduous teeth) seems unwarranted.

Prevention

Most dental health conditions are preventable. However the cost varies between different prevention strategies. The most common strategies to prevent dental caries and periodontal disease are (Health Development Section 1999):

Fluoridation of community water supplies: There is Level I evidence (good systematic review) indicating that water fluoridation is a beneficial strategy for the prevention and control of dental caries (Green & Wycoff 1989 as cited in Health Development Section 1999; NHMRC 1991 as cited in Health Development Section 1999).

The prevalence of dental caries is greatly influenced by water fluoridation. Six year old Victorian children living in fluoridated areas have been shown to have a 42% lower incidence of dental caries than those living in non-fluoridated areas (Dental Health Services Victoria 1999 as cited in Health Development Section 1999). A review of 113 studies world-wide found the modal reduction in tooth decay from fluoridated water was 50-59% in deciduous teeth and 40-49% in permanent teeth (Murray 1991 as cited in Health Development Section 2000).

Regular brushing of teeth using a fluoride toothpaste and flossing: There is Level I evidence (good systematic review) that regularly (at least daily) brushing teeth with fluoride toothpastes is beneficial for the prevention and control of dental caries, reducing incidence by approximately 30% (Green & Wycoff 1989 as cited in Health Development Section 1999; NHMRC 1991 as cited in Health Development Section 1999).

The use of dental floss has been found to be beneficial in adults (Ismail & Lewis 1993) and there is some evidence that it may also be beneficial for children. One study showed that regular flossing by a trained dental auxiliary reduced caries by 44% over 8 months in first grade school children (Wright, Banting & Feasby 1977), but no information was available for its value when used by children themselves.

Reducing consumption of refined sugars and adherent carbohydrates: There is some evidence that the risk of dental caries in infants may be decreased by avoiding putting infants to bed with a bottle and by encouraging breast feeding, however this is not definitive (US Preventive Services Task Force 1996).

There is Level II evidence (randomised controlled trials) demonstrating that replacing sugar with sugar substitutes is effective for the prevention and control of dental caries. However, sugar substitutes also have adverse effects including some evidence of a carcinogenic effect in mice consuming large doses (Scheinin & Makinen 1975 as cited in Health Development Section 1999; Nordblan, Suominen-Taipale, Murto, Vartiainen & Koskela 1995 as cited in Health Development Section 1999). Reducing the amount of sweeteners in the diet by avoiding refined foods seems preferable to using sugar substitutes.

Regular dental examinations: A number of well designed observational studies (Level IV evidence) have indicated that access to clinical examination (regular check-ups) prior to the obvious manifestation of dental decay has potential benefit for the prevention and control of dental caries, and for positive oral health behaviour change, particularly for school aged children (Jendresen, Allen, Bayne, Donovan, Hansson, Klooster & John 1994 as cited in Health Development Section 1999).

How might screening reduce the burden of suffering

Those suffering oral disease may experience pain or discomfort, which in severe cases can lead to difficulty eating with the consequential effects of poor diet and reduced general health and can also lead to infection. If not detected and treated early, dental caries in the permanent teeth can progress to a stage where the teeth require extraction in later adulthood. It is possible for gross caries in a primary tooth to cause a dental abscess that in turn may affect the underlying developing dentition resulting in defects in the enamel of the underlying permanent tooth and making it more susceptible to dental caries (personal communication with

Nicky Kilpatrick). However, such cases are rare and apart from pain and discomfort, there are usually no long term problems associated with untreated dental caries in the primary dentition (as these teeth fall out and are replaced by permanent teeth).

6.3.16.2 EVIDENCE

Tests

Oral examination

The predominant test involved in dental screening is examination of the teeth and mouth to identify dental caries, periodontal disease or any other oral conditions requiring further monitoring or treatment by a dentist.

A number of well designed observational studies have indicated that access to clinical examination prior to the obvious manifestation of disease has potential benefit for the prevention and control of dental caries, and for positive oral health behaviour change.

Another benefit of regular dental examinations is that it provides the opportunity for the dentist-initiated preventive oral health strategies. In addition to education of patients, dentists can apply fluoride to the teeth and/or dental sealants (acid-resistant adhesive coatings) to tooth surfaces thought likely to be susceptible to tooth decay. Both of these strategies have been shown in systematic reviews to be beneficial to the prevention and control of dental caries (NHMRC 1991 as cited in Health Development Section 1999; Kay & Locker 1996 as cited in Health Development Section 1999; Ripa 1993 as cited in Health Development Section 1999).

One study was identified which evaluated the use of a visual-only oral examination by a registered nurse or a dental hygienist against the gold standard visual-tactile oral examination by a dentist. The study involved 632 school children (75% of those invited to participate) aged 5-12 years. Data from 434 children was available to assess the two measures for the identification of untreated caries. The visual-only examinations produced a sensitivity of 94%, specificity of 97%, PPV of 95% and NPV of 96% with an untreated caries prevalence of 40% as determined by the visual-tactile examination. The results for examinations by the registered nurse and dental hygienist were very similar (Beltran, Dolores, & Eklund 1997).

Parent questionnaire

In the study reported by Beltran and colleagues (Beltran, Dolores, & Eklund 1997), they also asked parents to complete a multiple response questionnaire asking about conditions present in the child's mouth, including a question on untreated dental caries. Data were available for 305 children for the question asking parents whether their child had any untreated dental decay. Compared with the gold standard visual-tactile examination, this question produced a sensitivity of 69%, specificity of 88%, PPV of 80% and NPV of 80% with an untreated caries prevalence of 41% as determined by the visual-tactile examination.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	✓
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✓
Test acceptable to the population	✓ ¹
Natural history adequately understood	✓
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	?

¹ but treatment may not be acceptable to the population

Programs

Programs for dental screening generally aim to detect children in need of treatment for dental caries or other conditions classified under poor dental health. Programs are typically school-based and usually employ a dental nurse or dental therapist to conduct oral examinations. If problems are detected, children will usually be referred to a dentist for further examination and treatment where appropriate.

We identified one randomised controlled trial of a school-based dental screening program. The study sampled all 4-6 year old children attending schools in Wallasey, UK. All children received baseline dental screening. Children who had oral sepsis or extensive cavitation were immediately referred to a dentist and, in addition to those with evidence of recent dental treatment, were excluded from the study. The target study population consisted of all non-excluded children who were detected as having untreated caries (n=528). Children were stratified within schools by age, sex and number of teeth detected as having caries and randomised within each stratum. Those in the test group had letters sent home to their parents advising them that they should take their child to a dentist for clinical examination. Those in the control group received no further action. Four months after the baseline dental screening, all parents in the study population were asked to complete a questionnaire (97% response from test group and 94% response from control group). Results indicated a significant difference between the two groups (31% difference; $p < .001$); 73% of the test group and 42% of the control group had attended a dental examination in the four month period after baseline screening. Nonetheless, this still leaves 27% who did not act on the invitation. The proportion of children attending a dental examination in the control group was inversely related to area unemployment levels, and the difference in dental attendance between the test and control groups increased as area unemployment levels increased (Zarod & Lennon 1992).

However, we identified no information as to the longer term benefits of dental health screening programs.

Cost effectiveness

Dental health services account for approximately 6% (\$1.8 billion) of Australia's total health care costs and poor dental health has additional economic costs in terms of decreased productivity and days lost at work and school (Mathers, Penm, Carter & Stevenson 1998 as cited in Health Development Section 1999). While it is likely that the majority of these costs are incurred by adults, poor child dental health does contribute. Effective preventive strategies are appealing to reduce these costs, but information on the cost effectiveness of such strategies is limited.

Quality of evidence

PAPER	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
Zarod & Lennon, 1992	Level II	<p>1. Randomisation and concealment of allocation B (some limitations) the trial is stated as randomised but the method of randomisation and/or allocation concealment is not clearly described</p> <p>2. <i>Blinding of outcomes</i> n/a</p> <p>3. Withdrawals and dropouts B (some limitations) some description of withdrawals and dropouts</p> <p>4. Number of subjects included in trial 528</p>	<p>The difference between the test and control group on percentage of children attending a dentist following screening was p=0.002 at 1 month and p<0.001 at 2, 3 and 4 months.</p>	<p>The difference between the test and control group on percentage of children attending a dentist following screening was χ^2 (1 df) = 9.8 at 1 month, 34.6 at 2 months, 35.1 at 3 months, and 48.2 at 4 months.</p>	<p>The outcome measure (visit to a dentist) was appropriate for the aims of a dental caries screening program.</p> <p>The percentage of children receiving treatment for dental caries was not assessed. This may have provided information on the accuracy of the screening program to detect true cases of dental caries.</p>

One randomised controlled trial of a caries screening program for school children was identified. This study was of fair design and produced strong results in favour of the test group for the outcome measure (visit to a dentist). However the study did not consider whether the untreated caries that were identified by the screening received treatment.

One study of tests to identify untreated dental caries was identified. This study found visual-only oral examinations by a dental hygienist or registered nurse to be effective in identifying untreated dental caries (with visual-tactile oral examination by a dentist as the gold standard), but asking parents whether their child had any untreated caries was not a sensitive test.

Generalisability of evidence

The evidence identified was based on US and UK studies, but should generalise to the Australian context.

6.3.16.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening

Comments:

There is no clear evidence or consensus opinion on whether caries in the primary dentition should receive treatment.

6.3.16.4. RECOMMENDATIONS

Screening programs for dental caries in the deciduous teeth are not recommended.

We recommend regular surveys to document the prevalence and severity of caries in preschool and school aged child populations.

Public health efforts should focus on preventive dental health in preschool and school aged children. This should be given higher priority than screening efforts.

6.3.16.5. FURTHER RESEARCH

- Outcomes for treated vs untreated dental caries in the deciduous teeth
- The effectiveness of preventive strategies, and how to maximise uptake of those that are effective

6.3.16.6. SEARCH TERMS

Dental care

Dental care for children

Health education, dental

Tooth abnormalities

Tooth injuries

Tooth erosion

Tooth disease

Dental plaque

Dental caries

Dental caries susceptibility

Oral hygiene

Dentistry

Caries

6.3.16.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Nicky Kilpatrick – Director, Department of Dentistry, Royal Children’s Hospital, Melbourne.

Chris del Mar – Reference Group

George Patton – Reference Group

6.3.16.8 REFERENCES

- Baird, D. 1999, "Deciduous caries and screening", *British Dental Journal*, vol. 186, no. 6, p. 263.
- Beltran, E. D., Dolores, M. M., & Eklund, S. A. 1997, "Validity of Two Methods for Assessing Oral Health Status of Populations", *Journal of Public Health Dentistry*, vol. 57, no. 4, pp. 206-214.
- Centre for Disease Control and Prevention 2000, "Achievements in Public Health, 1900-1999: Fluoridation of Drinking Water to Prevent Dental Caries", *MMWR - Morbidity & Mortality Weekly Report in Journal of the American Medical Association*, vol. 283, no. 10, pp. 1283-1286.
- Dummer, P. M. H., Addy, M., Oliver, S. J., & Shaw, W. C. 1988, "Changes in distribution of decayed and filled tooth surfaces and the progression of approximal caries in children between the ages of 11-12 and 15-16 years", *British Dental Journal*, vol. 164, pp. 277-282.
- Health Development Section, Public Health Division, Department of Human Services. 1999, *Promoting Oral Health 2000 - 2004. Strategic Directions and Framework for Action*, Department of Human Services, Victoria.
- Health Development Section, Public Health Division, Department of Human Services. 2000, *Evidence-Based Health Promotion. Resources for Planning. No. 1 - Oral Health 1*.
- Ismail, A. I. & Lewis, D. W. 1993, "Periodic health examination, 1993 update: 3. Periodontal diseases: classification, diagnosis, risk factors and prevention. Canadian Task Force on the Periodic Health Examination.", *Canadian Medical Association Journal*, vol. 149, no. 10, pp. 1409-1422.
- Kidd, E. A. M. & Pitts, N. M. 1990, "The reappraisal of the value of the bitewing radiograph in the diagnosis of posterior approximal caries", *British Dental Journal*, vol. 169, pp. 195-200.
- O'Brien, M. 1994, "Children's Dental Health in the United Kingdom", HMSO, London.
- Slade, G. D., Spencer, A. J., Davies, M. J., & Burrow, D. 1996, "Intra-oral distribution and impact of caries experience among South Australian school children", *Australian Dental Journal*, vol. 41, pp. 343-350.
- US Preventive Services Task Force *Guide to Clinical Preventive Services*, 2 edn, US Department of Health and Human Services, Office of Public Health and Science, Office of Disease Prevention and Health Promotion.
- Widmer, R. P. & Mekertichian, K. 1996, "A contemporary approach to the art and science of caries risk assessment", *Annals of the Royal College of Dental Surgeons*, vol. 13, pp. 119-126.
- Wright, G. Z., Banting, D. W. & Feasby, W. H. 1977, "Effect of interdental flossing on the incidence of proximal caries in children", *Journal of Dental Research*, vol. 56, no. 6, pp. 574-578.
- Zarod, B. K. & Lennon, M. A. 1992, "The effect of school dental screening on dental attendance. The results of a randomised controlled trial", *Community Dental Health*, vol. 9, no. 4, pp. 361-368.

6.3.17. DEVELOPMENT

6.3.17.1. BACKGROUND

Clinical picture

Developmental delay or disability refers to the situation where there is either a delay in acquisition of developmental tasks or milestones, or (less commonly) a disorder of development in which milestones are achieved but qualitatively different (such as with disordered word combinations, or some personal-social milestones). Because a delay invariably leads to some form of disability, the terms developmental delay and developmental disability will be used interchangeably.

Child development is a complex, nonlinear process affected by multiple factors. There is a large degree of individual variation. Developmental delay is present when a child does not reach developmental milestones at the expected age (with adequate leeway for the broad variation among normal children). Although delay may result primarily from a biological factor such as a chromosomal disorder, or an environmental factor such as maternal depression, the principal model for the pathogenesis of developmental delay is a 'transactional' one. The process of development is viewed as a transaction between the child and the environment, in which each can have profound effects on the other.

About 17% of children have developmental delay (Yeargin-Allsopp, Murphy, Oakley, & Sikes 1992). Many, however, are not detected before school entrance (Drillien, Pickering, & Drummond 1988; Glascoe & Shapiro 1999), mainly because the disabilities are subtle or because they relate to tasks only then attempted by the child. Of the 17%, a much smaller proportion has more severe disability. Most young children with severe developmental delay are identified by methods other than screening (Dworkin 1989). They may be identified because they attend follow-up clinics for children who are biologically at risk, or they may be identified by clinical observation in the course of routine health care.

Severe and relatively obvious impairments are not usually the target of developmental screening programs – rather developmental screening is focused on identifying less severe problems that are not immediately obvious (Dworkin 1989). It has been demonstrated that clinicians are not accurate in identifying children with developmental delay, especially when it is less severe (Dearlove & Kearney 1990; Bierman Connor & Vaage 1964; Korsch, Cobb & Ashe 1961). There are a number of reasons suggested for this (Glascoe & Dworkin 1993) including insufficient time, a lack of training and expertise, together with a natural propensity to focus on more acute or pressing concerns. Mainly, however, this relates to the inherent difficulties in rapidly identifying whether a child has a developmental delay or not.

Measurement of child development centres around the measurement of abilities and aptitude, comparing the individual child to those of other children of the same chronological age (or corrected age in the first 2 years for premature infants). A child's development is usually described in terms of the developmental tasks they can or cannot perform, and the way in which these tasks are carried out. There are innumerable tasks of development, though in general a smaller number are considered regularly and form the basis of commonly used developmental milestones. Developmental milestones are often considered within a number of smaller categories, such as language, gross motor, fine motor, and personal/social. Often the measurement and description of development is presented in these categories. This section examines global measures of development, and not development in a specific area, such as of language or for autism.

In some screening areas, such as hypothyroidism, only one parameter need be measured and it is possible to develop clear cut-off levels. In contrast, child development has multiple parameters within each of which there is considerable individual variation. The development of quick and simple identification measures is therefore problematic. Difficulties may be compounded at a community level if, for example, many parents and health practitioners are reluctant to confront the reality that a developmental delay may be present, and there is a lack of adequate resources for assessment and intervention.

Natural History

Child development is nonlinear. It occurs in spurts and lulls, with peaks and plateaus. These will be influenced by child-related factors such as illness, and environmental factors such as family dysfunction. Development in different areas will proceed at different rates. Normal development will vary significantly from one individual to the next.

The age of presentation of developmental delay in different areas is influenced by the nature of the child's delay. Delay in walking, for example, is usually not considered before an age by which the majority of children have already started walking (eg at 18 months). Identification of developmental problems often occurs when normally there is rapid progress, so that the difference between normal and delayed development becomes more obvious. For example, language delay often presents towards the end of the second year of life following an obvious rapid burst in vocabulary in the great majority of children.

Genetics relevant to screening issues

There are a number of genetic conditions associated with intellectual disability such as Fragile X syndrome and Trisomy 21. Genetic conditions can also be associated with specific delays in areas such as language (often as a manifestation of hearing impairment) and motor function. At present known genetic aetiology explains only a fraction of disabilities.

Diagnosis

Diagnosis of developmental delay is made by the administration of a standardised developmental assessment tool or tools by an experienced clinician or clinicians. These measures may be of a global nature, covering multiple areas of development, or specific to one area of development, such as language function. There are many to choose from. These are used to provide a developmental age that can be compared with the child's chronological age, as well as a more detailed understanding of a child's specific strengths and weaknesses.

Contention surrounds issues such as the degree of deviation from the average that should be used to determine delay (eg 1 to 2 standard deviations), the validity of individual developmental assessment tools, and the developmental theories on which they are based.

Treatment/Management

There is evidence that early identification and intervention improves outcomes both for the child and for the family, and in general intervention is more likely to be successful when it focuses as much on supporting and training parents as it does on directly working with the child.

A meta-analysis in 1986 examined the effects of early intervention services on a broad range of disabled children younger than 3 years of age, and their families (Shonkoff & Hauser-Cram 1987). Results indicated that early intervention is effective in promoting developmental progress in infants and toddlers with biologically based disabilities. Programs oriented towards less severely affected children, which enrolled children before 6 months of age and encouraged high levels of parent involvement, achieved the best outcomes.

The evidence relating to interventions for children with specific disabilities, such as cerebral palsy, Down Syndrome, and autism, is very variable. Many specific forms of intervention, such as the neurodevelopmental treatment approach of Bobath for children with cerebral palsy, have not been shown to offer benefit (Guralnick 1997). Some others, particularly multifaceted intensive programs, have been shown to be beneficial for conditions such as autism.

There is good evidence that early, prolonged intervention with groups of children "at risk" on the grounds of poverty and social disadvantage, results in better short and long term outcomes educational and social outcomes (Gomby, Laner, Stevenson, Lewit, & Behrman 1995; Guralnick 1997; Weikart 1998). The degree of benefit however, is very variable, and is influenced by the nature and duration of the intervention and the level of disadvantage in the community. Many of the children who benefit would not be classified as having defined development delays.

In summary, early intervention has been shown to result in improved developmental, educational and social outcomes. There is some evidence that the earlier the intervention takes place, the better these outcomes. Finally, intervention doesn't just benefit those with criterion-based developmental delay. Intervention also benefits children at risk for developmental delay, including those children who do not meet diagnostic criteria, but are at the lower range of normal development.

Prevention

Developmental surveillance refers to a process of eliciting and attending to parents' concerns, making accurate and informative longitudinal observations of children, obtaining a relevant developmental history, and promoting development. Developmental surveillance may include the use of developmental screening tests. Developmental surveillance has not been demonstrated to actually prevent developmental delay and disability

Many children with high risk of developmental delay (such as through extreme prematurity) participate in follow up programs, often with the provision of early intervention services. There is reasonable evidence that these children can benefit, though this would appear to be more in the form of decreasing the degree of disability rather than prevention of all disability (Guralnick 1997).

Most children with developmental delay and disability have a problem before they are identified. For some children, in particularly high risk groups, it may be possible to prevent the formation of or decrease the extent of the disability. For children at risk due to poverty or other social disadvantage there is the potential for preventing developmental delay and disability.

How might screening reduce the burden of suffering?

Screening might reduce the burden of suffering if the children so identified could be provided with appropriate effective intervention. There is the potential for increasing the burden of suffering through false positives and through participating in unsuccessful and inadequately resourced interventions.

6.3.17.2 EVIDENCE

Tests

There are numerous developmental screening tests. They can be broadly grouped into those that measure global development, and those that measure specific areas of development, such as language. They can also be grouped into those that are based solely on parent report, and those that are all or partially based on observation and direct elicitation of tasks. This review considers global screens only. Some commonly used examples of global developmental screening tests follow:

Name	Age range	Sensitivity/specificity ¹	Time frame	Elicitation
Ages and Stages Questionnaire	0 to 60 months	Sensitivity 70-90% Specificity 76-91%	5 minutes	Parent report
Bayley Infant Neurodevelopmental Screen	3 to 24 months	Sensitivity 63-80%	10-15 minutes	Direct elicitation
Brigance Screens	21 to 90 months	Sensitivity \cong 75% Specificity \cong 80%	10 minutes	Parent report or direct elicitation
Child Development Inventories screening versions ²	3 to 72 months	Sensitivity > 75% Specificity 70%	10 minutes	Parent report
Denver II ³	0 to 6 years	Sensitivity 43-80% Specificity 56-83%	20-30	Parent report and direct elicitation
Parents' Evaluation of Developmental Status	0 to 8 years	Sensitivity 74-79% Specificity 70-80%	2 minutes	Parent report

¹ The figures quoted are based on information from a number of sources/studies. The 'gold standards' with which screens were compared and the figures obtained will vary and therefore values may not be directly comparable. There needs to be caution in utilising this information for test selection.

² There are three versions: Child Development Chart [Birth - 21 months]; Early Child Development Inventory [15m - 3 years]; Pre-School Development Inventory [3 years - Kindergarten]

³ Sensitivity and specificity are from (Glascoe, Byrne, Ashford, Johnson, Changs, & Stricklands 1992)

The Ages and Stages Questionnaire (Bricker & Squires) covers five key developmental areas: communication, gross motor, fine motor, problem solving, and personal-social. It provides clear drawings and simple directions to help parents indicate their child's skills, with 10-15 items for each age range. There are 19 different questionnaires, each for a different age. It is specifically designed to be part of a child health monitoring program.

The Bayley Infant Neurodevelopmental Screen (Aylward 1995) uses 10 - 13 directly elicited items per 3 - 6 month age range to assess neurological processes (reflexes, and tone), neurodevelopmental skills (movement, and symmetry) and developmental accomplishments (object permanence, imitation, and language). Cut off scores determine low, moderate, and high risk of developmental delays or neurological impairment.

Child Development Inventories screening versions consist of three separate instruments each with 60 yes-no descriptions. It can be mailed to families, completed in waiting rooms, administered by interview or by direct elicitation. There is a 300 item assessment level version. A single cutoff for failure is tied to 1.5 standard deviations below the mean.

The Denver II is possibly the most commonly used developmental screening test. It considers development in domains of gross motor, fine motor, personal-social, and language (Frankenburg, Dodds, Archer, Shapiro, & Bresnick 1992) A combination of parent report and observation is used. There are very good data about the milestone ranges from the normative population. However, though originally the Denver Developmental

Screening Test was proposed to be used with pass/fail scoring, this is not the case with the Denver II which is expected to be used as an aid in the evaluation of a child's development. Despite this, it is often used as pass/fail in screening programs. The sensitivity and specificity quoted in the table above reflect this mode of use. In 1992 Dworkin concluded that "Results should not serve, in isolation, as a basis for referral, diagnosis, or prediction of future functioning. Rather the Denver II should serve as an effective aid to developmental surveillance, better enabling the pediatric provider to perform skilled observations of children." (Dworkin 1992).

The Brigance Screens (Brigance 1985) consists of seven separate forms, one for each 12 month age range. It covers speech-language, motor, readiness and general knowledge at younger ages, and also reading and math at older ages. It uses direct elicitation and observation. Compared to the other screening tests mentioned here, the Brigance Screens focuses more on academic performance.

The Parents' Evaluation of Developmental Status (PEDS) (Glascoe 1997) is a recently developed screening test that relies on parent report of concern in response to 10 specific questions. It can be completed at interview or as a written questionnaire. A suggested scoring and response algorithm offers a number of alternatives: referral for assessment; further screening if available; counselling (with appropriate written information) and review; or reassurance depending on the types of concerns and the child's age. This questionnaire has been well validated and parent concern been shown to work satisfactorily except when there is a language barrier or a mental health problem in the parent. The PEDS has been designed to be part of a regular monitoring process.

One Australian developmental screening test for children 6 to 60 months has been identified, the Australian Developmental Screening Test (ADST) (Burdon 1993). There was no reference to this identified other than in the ADST Examiners Manual. This provides some information about test development and test characteristics. The item selection and norming process are not well described though. There are no specific data on sensitivity or specificity. Reliability is quoted as 77 to 86%.

When considering psychometric properties of developmental screening tests, it is important to be aware that even those children who are false positives on screening (in that with formal assessment they are subsequently shown not to meet criteria for developmental delay) perform substantially lower than children with true-negative scores on measures of intelligence, language, and academic achievement, as well as having more psychosocial risk factors (Glascoe 2001). They are thus an at-risk group who would benefit from some form of intervention. As such, it is not appropriate or necessary to expect specificity figures to be as high as with some other screening tests.

Studies in the USA (Shonkoff & Dworkin 1979) and UK (Bain 1977) have shown that practitioners use a variety of methods in attempts to detect problems early. These include checklists, milestones, taking developmental histories from parents, clinical observations, and informal testing using a variety of ad hoc items. A similar diversity of methods is likely to be the norm amongst health practitioners in Australia. Unfortunately there is only rarely information relating to the characteristics and performance of these methodologies.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓ ¹
Acceptable to the public	✓
Accurate	✓/✗ ¹
Repeatable	✓ ²
Sensitive	✓/✗ ^{1,3}
Specific	✓/✗ ^{1,3}

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓/✗ ⁴
Facilities for diagnosis and treatment	✗
Latent or early symptomatic stage	✓
Suitable test or examination	✗✓
Test acceptable to the population	✓
Natural history adequately understood	✗
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	?

¹ Varies depending on test.

² Repeatability is usually only evaluated in the short term (few weeks).

³ Developmental screening tests suffer particularly from the problem of rapidly decreasing sensitivity when specificity is decreased through alterations to cutoff points or content, and vice versa. Attempts to improve this usually require the addition of more items, which moves the screening test closer to a developmental assessment tool.

⁴ Whilst there is evidence that in certain situations specific treatments are beneficial, there is not general acceptance about which treatments should be used and when.

Programs

Screening aims only to detect those children with subtle delays, or at risk for delay, who would not otherwise be detected. For developmental screening to be effective, at risk children must proceed to more detailed assessment, parents must be provided with appropriate counselling and information, and effective early intervention programs must be offered. However, as discussed above, significant benefits of early intervention are often only able to be identified after long term follow up, and after a relatively resource-intensive intervention. Because the outcomes of screening are so closely linked to the intervention, which has so many variables and is heavily influenced by resource availability, evaluation of the effectiveness of a developmental screening program ideally requires that outcomes of the intervention be included also.

Programs may be universal or may focus on high risk populations such as graduates of a neonatal intensive care units. In such situations, the developmental screen is usually only one part of the follow up and evaluation of development. Most proposed and functioning developmental screening programs are universal in that they focus on whole communities (though some of these communities may be “high risk”, for socioeconomic reasons for example). Some universal developmental screening programs are very specific and take place as freestanding exercises. Most are offered by practitioners as part of a primary care service.

Only one randomised control trial of developmental screening and intervention was identified and this is appraised below (Cadman, Chambers, Walter, Ferguson, Johnston, & McNamee 1987). This used a less satisfactory screening test, and the intervention was counselling and referral. Subsequently the child received the usual services available in that community for those with suspected developmental delay. This study did not demonstrate a benefit from screening, counselling and referral. In part, this failure is likely due to inadequacy of the intervention.

The authors compared a group of children registering to commence school who were randomly allocated to receive either (a) developmental screening with the Denver Developmental Screening Test, counselling and referral (receiving services as any other similar referral in the region), (b) developmental screening with DDST, but no counselling of results or referral, or (c) no developmental screening. Outcomes were measured at 1 and 3 years after screening and included referrals to specialised services, teacher assessment of performance, parent-completed child health status, and formal academic and cognitive testing.

The study was of moderate quality. One criticism relates to the use of the DDST, which has been identified to have major problems in concurrent and predictive validity and is not recommended for use as a pass/fail screen. Another area of potential concern was in the follow up of the DDST screening normal group, of which a random sample was taken of the total pool. Follow up was of between 54% and 68% depending on the outcome measure. Standardised outcome measures for these groups were in the expected range suggesting the loss to follow up did not influence the major results of the study relating to children who failed the DDST. Follow up of the DDST failing group was 88%.

A further area of potential concern was the blinding of participants. Parents were not blind to whether their child received treatment or not, though the group who failed the DDST but were not given treatment were blinded to the fact that their child had failed. The community nurses and most teachers were not blind to treatment status. The psychologist doing individual child testing was blind to treatment.

Finally, it would appear that many children did not receive an ideal level of intervention. This would have had impact on the apparent effectiveness of the screening/intervention arm. Whilst it was the author's intention to test the process in ‘real life’ rather than in an ideal situation, it is quite possible that a more effective assessment and intervention efficient system (or not unreasonable cost and complexity) may have improved the outcomes.

Whilst overall those children who had failed the DDST did worse than those who had a normal DDST, there was no difference in most outcomes for those who failed the DDST and who received intervention compared to those who failed but had no intervention. In fact parents in the group who received the counselling and intervention reported much higher rates of worry about their child's school performance in the preceding 3 months. The author's conclusion was that they did not find good evidence for the effectiveness of developmental screening followed by a program of community nurse counselling and referrals to specialised services of children with positive screening tests results. The potential for harm as a result of the process needs to be considered.

Cost effectiveness

There are some differences in the cost of screening depending on the measure used and by whom it is administered (Glascoe 1997).

Cost effectiveness studies of early intervention have generally considered highly intensive strategies, which are most likely to result in benefit. The economic benefits of early intervention have been demonstrated for disadvantaged children and their families, but are not so clear when instigated for developmentally disabled children (Shonkoff & Hauser-Cram 1987).

No studies described the cost effectiveness of a community screening program for developmental delay.

There may be detailed cost effectiveness information in government and administrative data sets and reports which were not accessed for this review.

Quality of evidence

<u>PAPER</u>	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
Cadman et al 1987	Level II	<p><i>Randomisation and concealment of allocation</i> B (some limitations) the method of randomisation is not stated, but concealment was through sealed opaque envelopes.</p> <p><i>Blinding of outcomes</i> B (some limitations) not stated the extent to which parents were aware of the allocation, but other key reporters of outcomes and providers of interventions were.</p> <p><i>Withdrawals and dropouts</i> B (some limitations) follow up rates for the three groups were 54%, 68% and 88.5% (for the abnormal DDST, counselling, and referral group)</p>	<p>No statistically significant differences ($p < .05$) between intervention and non intervention groups, except a greater proportion of intervention group parents (33%; 95% CI 6% to 64%) were worried about their child's school work</p>	<p>DDST used as screen (Has poor test characteristics.)</p> <p>This study examined a screening program in a 'real life' as compared to ideal system. As such there is the room for different outcomes if there had been greater compliance and resources available.</p> <p>Only 2/3 positive screened children were seen by their primary care physician, and a proportion who were did not get referred for intervention. It is not clear on what grounds this occurred.</p>	

No systematic reviews (Level 1 evidence) of developmental screening, either of communities or of individuals as seen by health care providers, were been identified. One randomised controlled trial of a screening program for developmental delay (Level II evidence) was identified.

6.3.17.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against developmental screening

Comments:

Currently developmental delay does not appear to meet criteria for screening programs. However, there is evidence that the identification of developmental delay/disability (or significant risk factors) and subsequent intervention can improve developmental and other social outcomes. There is some evidence that the earlier the intervention, the better the outcome. In addition, if it can be argued that in the case of child development there should be less stringent expectations in terms of test sensitivity and specificity (given the intrinsic difficulties in measurement of development, the fact that even false positives from developmental screening may benefit from some form of intervention, and the high prevalence in the community), there are some developmental screening tests that are acceptable. There still will be, however, a large number of children who will not be identified by developmental screening tests.

There is no high quality evidence putting together all the links in the chain and reporting on the effectiveness of developmental screening programs on child developmental outcomes.

6.3.17.4. RECOMMENDATIONS

The identification of children who would benefit from early intervention should not be based solely on the use of developmental screening tests, or limited to inquiry at one point of time.

It is not recommended that screening programs for developmental delay be implemented at this stage.

Until there is evidence that alternatives to developmental screening programs function better, existing programs should be reviewed to ensure that adequate tools and processes are used

Community programs of developmental screening should use tools that have been demonstrated to have adequate psychometric properties (sensitivity and specificity greater than 70%).

Community programs of developmental screening should allow for intervention or assistance of some degree for children who fail screening, but who pass diagnostic tests.

Individualised checklists of milestones should not be used as developmental screening tests.

6.3.17.5. FURTHER RESEARCH

- Research to demonstrate the link between participation in developmental screening, with subsequent intervention where indicated, or not and subsequent development, educational and social outcomes of children.
- Further research into better developmental screening tests is encouraged. However, this should take second priority to research at program level. Given the nature of child development it would appear unlikely that there will be significant improvements in test characteristics in the near future from those already achieved.
- Research into screening programs in high risk or targeted groups.
- Consideration of new models of early identification of developmental delay and disability, combined with intervention that is responsive to the child's needs. This might involve aspects of what is in some places called surveillance or monitoring, as well as the use of screening tests. There could be less reliance on screening tests; they would be used more as adjuncts when a course of action is not already clear.

6.3.17.6. SEARCH TERMS

Child development

Child development disorders

Motor skills

Motor skills disorders

6.3.17.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Dinah Reddihough – Director, Child Development and Rehabilitation, Royal Children’s Hospital, Melbourne

Doug Bryan – Divisional Director, Community Orientated Paediatric and Adolescent Services, Royal Children’s Hospital, Melbourne

Frances Page Glascoe – Reference Group

Paul Dworkin – Reference Group

6.3.17.8 REFERENCES

Aylward, G. P. 1995, *Bailey Infant Neurodevelopmental screener*.

Bain, D. G. B. 1977, "Methods used by general practitioners in developmental screening of preschool children", *British Medical Journal*, vol. 2, pp. 363-365.

Bierman, J. M., Connor, A. & Vaage, M. 1964, "Pediatrician’s assessment of the intelligence of two year olds and their mental test scores", *Pediatrics*, vol. 43, pp. 680-690.

Bricker, D. & Squires, J. *Ages & Stages Questionnaires (ASQ). A Parent-Completed, Child-Monitoring System*, 2 edn, Paul H. Brookes, Publishers, Baltimore, Maryland.

Brigance, A. N. 1985, *Curriculum Associates, Inc.* North Billerica, MA..

Burdon, B. 1993, *The Australian Developmental Screening Test (ADST)* Harcourt Brace and Co., Sydney.

Cadman, D., Chambers, L. W., Walter, S. D., Ferguson, R., Johnston, N., & McNamee, J. 1987, "Evaluation of public health preschool child developmental screening: the process and outcomes of a community program", *Am.J Public Health*, vol. 77,no. 1, pp. 45-51.

Dearlove, J. & Kearney, D. 1990, "How good is general practice developmental screening?", *British Medical Journal*, vol. 300, pp. 1177-1180.

Drillien, C. M., Pickering, R. M., & Drummond, M. B. 1988, "Predictive value of screening for different areas of development", *Dev.Med Child Neurol.*, vol. 30,no. 3, pp. 294-305.

Dworkin, P. H. 1989, "Developmental screening: Expecting the impossible? ", *Pediatrics*, vol. 83, pp. 619-622.

Dworkin, P. H. 1992, "Developmental screening: (Still) expecting the impossible.", *Pediatrics*, vol. 89, pp. 1253-1255.

Frankenburg, W. K., Dodds, J., Archer, P., Shapiro, H., & Bresnick, B. 1992, "The Denver II: a major revision and standardization of the Denver Developmental Screening Test.", *Pediatrics*, vol. 89, pp. 91-97.

Glascoe, F. P. 1997, "Parents' concerns about children's development: Prescreening technique or screening test.", *Pediatrics*, vol. 99, pp. 522-528.

Glascoe F. P. 2001, "Are overreferrals on developmental screening tests really a problem? ", *Arch Ped Adol Med*, vol. 155, pp. 54-59.

Glascoe, F. P., Byrne, K. E., Ashford, L. G., Johnson, K. L., Changs, B., & Stricklands, B. 1992, "Accuracy of the Denver II in Developmental Screening.", *Pediatrics*, vol. 89, pp. 1221-1225.

Glascoe, F. P. & Dworkin, P. H. 1993, "Obstacles to effective developmental surveillance: Errors in clinical reasoning", *J Dev Behav Pediatr*, vol. 14, no. 5, pp. 344-349.

Glascoe, F. P. & Shapiro, H. L. Developmental and Behavioural Screening. Developmental Behaviour and Pediatrics Online Community . <http://www.dbpeds.org/articles/dbtesting/overview.html> 1999. 23-9-1999.

Gomby, D. S., Lamer, M. B., Stevenson, C. S., Lewit, E. M., & Behrman, R. E. 1995, "Long-term outcomes of early childhood programs: analysis and recommendations", *Future.Child*, vol. 5,no. 3, pp. 6-24.

Guralnick, M. J. 1997, *The effectiveness of early intervention*. Paul Brookes Publishing, Baltimore.

Korsch, B., Cobb, K. & Ashe, B. 1961, "Pediatricians’ appraisal of patients’ intelligence", *Pediatrics*, vol. 27, pp. 990-1003.

Shonkoff, J. P., Dworkin, P. H., Leviton, A. & Levine, M. D. 1979, "Primary care approaches to developmental disabilities", *Pediatrics*, vol. 64, pp. 506-514.

- Shonkoff, J. P. & Hauser-Cram, P. 1987, "Early intervention for disabled infants and their families: a quantitative analysis", *Pediatrics*, vol. 80,no. 5, pp. 650-658.
- Weikart, D. P. 1998, "Changing early childhood development through educational intervention", *Preventive Medicine*, vol. 27,no. 2, pp. 233-237.
- Yeargin-Allsopp, M., Murphy, C. C., Oakley, G. P., & Sikes, R. K. 1992, "A multiple-source method for studying the prevalence of developmental disabilities in children: the Metropolitan Atlanta Developmental Disabilities Study", *Pediatrics*, vol. 89,no. 4 Pt 1, pp. 624-630.

6.3.18. LANGUAGE

6.3.18.1 BACKGROUND

Clinical picture

Children with language problems may have language which is disordered/deviant, or delayed in acquisition but apparently normal in form. The distinction between delay and disorder/deviance is not always clear and often a child's language impairment will have delay and disorder. Because this distinction is often not clear, in this review both will be referred to as "language delay", as in the recent major systematic review by Law and colleagues (Law, Boyle, Harris, Harkness, & Nye 1998). Other terms used in the literature are language impairment and language disability.

Language delay may be classified as primary or secondary delay. In primary delay, speech and/or language skills are delayed relative to other developmental skills; there is usually no clear cause for such delay. In secondary delay, speech and/or language skills are delayed in the context of other issues or other developmental delays; usually there is an identified cause, eg hearing impairment, autism, cleft palate, or intellectual disability, although single or multiple factors may be involved (Law et al. 1998).

Primary delay is the focus of screening programs for language delay ie delay for no apparent reason other than speech/language development. The level of delay and the specific area(s) of language affected vary widely (Law et al. 1998). Language delay can be pervasive, or may affect one or more domains of language (expressive, receptive) and one or more linguistic area (pragmatics, phonology, articulation, syntax, semantics, prosody, voice, fluency) solely or dominantly. Comorbidities are also common (eg behaviour, social and other neurodevelopmental problems).

There are many issues that may limit the value of screening for language delay:

- It is not possible to characterise children along a single axis. Speech and language are multidimensional, as are their disorders.
- There is a wide range of normal variation in the speed and quality of language acquisition in the early years. Speed of acquisition may vary within the same child over time, resulting in children moving in and out of language delay categories. In many cases it is not possible to distinguish those with apparent early delays (particularly in expressive language) who will go on to suffer long term language delay from those who will not.
- The true prevalence of language disorders is not known. Tests used to define language disorders are usually psychometrically standardised, which means that the percentage of children defined as having a disorder is determined by the shape of the normal curve and the cutpoints used. Cutpoints are usually arbitrary in the absence of good predictive studies of adverse outcome, so that prevalence of disorders is also arbitrary and manipulable.
- Although observation of children's language skills is relatively easy, it is difficult to reliably test language skills in young children, and good predictive definitions of language delay in young children are lacking.
- Many children who present with language problems at the age of two appear to "catch up", displaying age-appropriate language by the time they reach school (Klee, Carson, Gavin, Hall, Kent, & Reece 1998).

Natural History

The natural history of language delay varies between children and areas of language affected. Different children develop language skills at different rates. Some children who appear to have language problems at two years of age have "caught up" to their peers by the time they reach school (Klee et al. 1998).

It is difficult to predict which children with early language delay at age 0-3 years will continue to have problems later in childhood and which children will have problems that resolve. Young children who do not use gesture or symbols to indicate wants, and those with poor receptive vocabularies, may be more at risk of persistent problems (Thal & Bates 1991). Studies of children with isolated expressive language delays have reported that up to 60% of these children experience a resolution of their language delay without treatment between 2-3 years of age. However, many of the children included in these studies displayed no evidence of serious difficulties due to their language delay in the first instance. Other studies have demonstrated that 41-75% of children with early expressive language delay exhibit problems with reading at 8 years of age. It is unclear what proportion of children with other types of language delay will experience persistent problems. However, it is believed that children identified with combined expressive and receptive delays are more likely to have persistent problems than children identified with expressive delays alone (Law et al. 1998) although this may be an oversimplification of the patterns and domains of language delay.

Prevalence

Language delay is a common condition, but estimates on its exact prevalence vary between studies, partly due to differences in definitions, cut-points, populations studied and methodologies. A systematic review by Law and colleagues (1998) of studies investigating primary language delay in 0-7 year olds reported a median overall prevalence of 6% with a range of 1%-33%. Median prevalences across studies by age were 16% for expressive delay at age 2 years; 3.0%, 2.3% and 2.6% at 3 years for combined delay, expressive delay and receptive delay; 2.1%, 4.3% and 3.9% at 5 years; and 2.0%, 2.8%, and 3.6% at age 7 years (Law et al. 1998). Some studies have found the prevalence to rise with age (Reep-van den Bergh, de Koning, de Ridder-Sluiters, van der Lem & van der Maas 1998) while others report a fall with age (Bax, Hart & Jenkins 1980). Disorders diagnosed at older ages tend to be more stable than those diagnosed in the preschool years.

Genetics related to screening

No genetic cause has been identified for primary language delay because the definition of primary delay excludes such causes. However language delays frequently occur in families containing other affected individuals, with both environmental and heritable mechanisms assumed.

Diagnosis

Diagnosis of language delay in the preschool years is often prompted by parental concern. Social problems and poor academic progress may be prompts for assessment in the school years. A number of approaches can be used in diagnosis. These include detailed parent/carer report, analysis of a child's free speech, and formal testing (including a range of play activities to elicit particular verbal and/or non-verbal responses and thereby indicate the child's understanding and use of speech and language). There is a battery of formal tests for children at various developmental stages focusing on different aspects of language development: expressive language, receptive language, visual language, word comprehension, sentence formation, social use of language etc. Clinical judgement is almost always incorporated into diagnosis of language delay. If confirmed, other aspects of the child's functioning should also be assessed to determine whether delay is primary or secondary.

Treatment/Management

Treatment for language delay can be 'indirect' where a parent and/or teacher administers the treatment, or 'direct' where a health professional administers treatment. Direct treatment may involve the child individually or groups of children and may also incorporate the teaching of skills to parents. Treatment is tailored to the specific areas of deficit and usually individualised for each child. Treatment may vary from formal speech pathology to general encouragement and stimulation of communication, usually involving a combination of formal and informal management.

A systematic review of studies reporting treatment interventions for primary language delay in children under seven years of age reported statistically significant ($p < .05$) results in favour of treatment outcomes for 9 of 10 randomised controlled trials and 10 of 12 quasi-randomised trials. Results were in the magnitude of one standard deviation improvement which the authors reported as corresponding to advancement from the 5th to the 25th percentile on a standardised language test. The review also reported that 'indirect' treatment could be as effective as 'direct' treatment (Law et al. 1998).

Treatment appears to result in more favourable outcomes for children identified with language delay in terms of social and academic achievement. However, the issue remains that some children identified with language delay (particularly delay in expressive language development) will have spontaneous resolution of normal language functioning ("catch up") without treatment. Such children, along with false positives identified through screening, may experience no ill effects from the treatment itself and may benefit from treatment, but may suffer harm through lowered parent and teacher expectations of the child and increased anxiety (Law et al. 1998).

Prevention

Early language delay could potentially be most effectively prevented through the early and sustained provision of language enrichment (either through centre-based early intervention programs or through parent education, training and support programs). A review of studies on the effectiveness of anticipatory guidance reported advanced language development in infants exposed to this health promotion activity (Dworkin 1998).

How might screening reduce the burden of suffering

Children with language delay may experience continuing problems with verbal language as well as general difficulties with reading, spelling and educational achievement. In addition, behaviour problems, poor self esteem and difficulties with other aspects of psychosocial adjustment may develop (Law et al. 1998) and persist into adulthood. Such difficulties create problems not only for the child, but for their family, school

and the community. However, because of difficulties with definition and prevalence estimates, it is difficult to quantify these relationships. In older children, the educational and behavioural problems associated with language delay may be more apparent than the language problem itself.

Children with language problems identified at age 5 or later have poorer prognosis with remediation than children identified earlier (Law et al. 1998). Therefore, early identification and intervention may help to relieve the burden of suffering for affected children and their families.

6.3.18.2 EVIDENCE

Tests

There are many tests available to screen children for language delay. Covered here are some of the available tests that were used in studies included in the systematic review by Law and colleagues (Law et al. 1998).

Fluharty Preschool Language Screening Test

Three US studies (136, 378 and 533 subjects respectively) compared the use of this monophasic test on different populations (different geographical areas) using the Test of Language Development, Test of Auditory Comprehension of Language or the Carrow Elicited Language Inventory as the reference for language delay. Ages of children involved ranged from 4 years 5 months to 6 years 5 months. Sensitivity ranged from 0.17 to 0.65 and specificity from 0.85 to 0.97. Likelihood ratios (sensitivity/(1-specificity)) ranged between 2.53 and 10.77 (Illerbrun, Haines & Greenough 1985 as cited in (Law et al. 1998; Stumer, Heller, Funk & Layton 1993 as cited in Law et al. 1998; Stumer, Kunze, Funk & Green 1993 as cited in Law et al. 1998).

Another study compared this test with another language screening test in a normal population of children aged 3 years to 6 years 2 months, using the Sequenced Inventory of Communication Development as the reference test. This study reported the Fluharty to have a sensitivity of 0.60, specificity of 0.80, likelihood ratio of 3.15 and PPV of 0.33 (Allen & Bliss 1987 as cited in Law et al. 1998).

Based on these results, this test appears to be reasonable at excluding children who do not have language delay, but may misclassify up to 83% of 4.5-6.5 year old children with language delay as not delayed.

Sentence Repetition Screening Test

Two US studies (343 and 382 subjects respectively) compared the use of this monophasic test on different populations (different age ranges, same geographical area) using the Arizona Articulation Proficiency Scale or Illinois Test of Psycholinguistic Abilities and Bankson Language Screening Test as reference tests. The Sentence Repetition Screening Test comprises a speech scale and a language scale. Ages of children involved ranged from 4 years 6 months to 5 years 6 months (younger group) and from 5 years 3 months to 8 years (older group). Sensitivity was better in the older age group than the younger age group for the speech scale (0.74 and 0.57 respectively) and for the language scale (0.76 and 0.62 respectively). Specificity was similar in both age groups and for both scales: 0.95 (younger group) and 0.92 (older group) for the speech scale and 0.91 (younger group) versus 0.92 (older group) for the language scale. Likelihood ratios were higher in the younger age group for the speech scale (11.40 versus 9.69), but lower in the younger age group for the language scale (6.90 versus 9.41) (Stumer, Kunze, Funk & Green 1993 as cited in Law et al. 1998; Stumer, Funk & Green 1996 as cited in Law et al. 1998).

In addition, a study comparing this test with another using 2 different references for language delay (Illinois Test of Psycholinguistic Abilities and Bankson Language Screening Test or Arizona Articulation Proficiency Scale) in a population of normal children aged 4 years 6 months to 5 years 6 months reported sensitivities of 0.62 and 0.57, specificities of 0.91 and 0.95, likelihood ratios of 6.90 and 11.40 and PPV of 0.44 and 0.75 (Stumer, Funk & Green 1996 as cited in Law et al. 1998).

This test appears to be good at excluding children who do not have language delay and is relatively good at correctly classifying children aged 5.5-8 years with language delay but may misclassify up to 43% of children aged 4.5-5.5 years with language delay.

Northwestern Syntax Screening Test

One study compared this monophasic test with another language screening test in a normal population of children aged between 3 years and 6 years 2 months, using the Sequenced Inventory of Communication Development as the reference test. This study reported the Northwestern Syntax Screening Test to have sensitivity of 0.92, specificity of 0.48, likelihood ratio of 1.79 and PPV of 0.22 (Allen & Bliss 1997 as cited in Law et al. 1998).

Revised Denver Developmental Screening Test/ Denver Developmental Screening Test II

One study compared the expressive and receptive language scale on this multiphasic test with another language screening test using the Sequenced Inventory of Communication Development as the reference test in a clinical population of children with a mean age of 3 years 6 months. They reported sensitivities ranging between 0.92 and 0.98, specificities between 0.14 and 0.49, likelihood ratios between 1.12 and 1.79 and PPV ranging from 0.53 to 0.71 (German, Williams, Herzfeld & Marshall 1991 as cited in Law et al. 1998).

Another study comparing the Denver with another test in a mixed population of normal and clinical children aged seven months to 5 years 10 months, using the Fluharty Preschool Speech and Language Screening Test or Vineland Adaptive Behaviour Scale as the reference, reported sensitivity of 0.73, specificity of 0.76, likelihood ratio of 3.02 and PPV of 0.43 (Glascoe & Byrne 1993 as cited in Law et al. 1998).

Developmental Profile II

A study compared the expressive and receptive language scales on the multiphasic Developmental Profile II with another screening test in a clinical population of children with a mean age of 3 years 6 months using the Sequenced Inventory of Communication Development as the reference test. They reported sensitivities ranging from 0.92 to 1.00, specificities ranging between 0.36 and 0.72, likelihood ratios between 1.56 and 3.30 and PPV between 0.61 and 0.81 (German, Williams, Herzfeld & Marshall 1991 as cited in Law et al. 1998).

Battelle Development Inventory Screening Test

This multiphasic test was compared with another screening test using the Fluharty Preschool Speech and Language Screening Test and Vineland Adaptive Behavior Scale as the reference, in a study of a mixed population of normal and clinical children aged seven months to 5 years 10 months. Sensitivity of 0.78, specificity of 0.70, likelihood ratio of 2.63 and PPV of 0.40 were reported (Glascoe & Byrne 1993 as cited in Law et al. 1998).

Speech and Language Screening Questionnaire

A study comparing this monophasic test with another using 2 different criteria for language delay (Illinois Test of Psycholinguistic Abilities and Bankson Language Screening Test or Arizona Articulation Proficiency Scale) in a normal population of children aged from 4 years 6 months to 5 years 6 months reported sensitivities of 0.59 and 0.68, specificities of 0.43 and 0.89, likelihood ratios of 1.04 and 6.18 and PPV of 0.12 and 0.66 (Sturmer, Funk & Green 1996 as cited in Law et al. 1998).

Nurses developmental screen

This monophasic test was compared with a parent questionnaire in a normal population of children aged 2 years 10 months to 3 years 4 months using the Reynell Developmental Language Scales as the reference test for language delay. This study reported sensitivity of 0.77, specificity of 0.97, likelihood ratio of 28.17 and PPV of 0.78 (Stokes 1997 as cited in Law et al. 1998).

MacArthur Communicative Development Inventories (CDI)

The CDI is a parent report measure of children's language development. The CDI is available in two forms, suitable for children of different ages. The CDI Infants or CDI Words and Gestures is suitable for children aged 8-16 months and assesses vocabulary comprehension, vocabulary production and use of gestures. The CDI Toddlers or CDI Words and Sentences assesses language production and is suitable for children aged 16-30 months or children older than 30 months who are thought to be functioning below the age of 30 months (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick & Reilly 1993 as cited in Klee et al. 1998).

Language Development Survey (LDS)

The LDS is a parent report measure that was designed specifically for screening toddlers for language delay. It was originally developed for parents to complete in a clinical setting with a clinician available to answer questions, but has been used successfully for postal surveys where parents received detailed written instructions. The LDS assesses children's expressive vocabulary and extent to which the child has begun combining words. It also includes a question enquiring if the parent has any concerns about their child's language development. The LDS has shown excellent concurrent validity and temporal reliability when used with children aged 18-33 months. Results on the LDS have been shown to correlate highly with vocabulary tests on the Reynell Developmental Language Scales, the Bayley Mental Development Scale and the Preschool Language Scale (Rescorla 1989 as cited in Klee et al. 1998).

Parents' Evaluations of Developmental Status (PEDS)

The PEDS is a parent report measure consisting of 10 questions enquiring if a parent has any concerns on 10 areas of their child's development. The entire instrument takes approximately 2-5 minutes to complete. Two questions relate to speech and language: "Do you have any concerns about how your child talks and makes

speech sounds?” which detects parent concerns about expressive language and articulation, and “Do you have any concerns about how your child understands what you say?” which detects parent concerns about receptive language. The PEDS has been shown to have sensitivity of 74-79% and specificity of 70-80% for detecting developmental problems including language delay (Glascoc 1997). More details are given about this instrument in the section on development.

Law and colleagues report that of the studies they reviewed, the lower quality studies tended to generate higher likelihood ratios ($r=-0.23$, $p<.05$) and higher sensitivity ratings ($r=-0.35$, $p<.005$), particularly where normal populations were involved ($r=-0.34$, $p<.05$ and $r=-0.48$, $p<.001$ respectively) (Law et al. 1998). Therefore the ratings reported for different screening tests where different study techniques were employed are not directly comparable.

In addition, higher quality studies tended to report significantly higher specificity than sensitivity ($t=4.41$, $p<.001$) indicating that it is simpler to accurately identify children who do not have language delay than it is to correctly identify children who do have language delay (Law et al. 1998). However, high specificity with only moderate sensitivity is of questionable value.

Another issue with the studies reviewed by Law and colleagues was that different tests were used as the gold standard for different studies. In fact a test that was being evaluated in one study was used as the gold standard in another study.

Does this condition meet criteria for a screening program?

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	X
Acceptable to the public	✓
Accurate	?
Repeatable	X
Sensitive	?
Specific	?

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	X
Facilities for diagnosis and treatment	?
Latent or early symptomatic stage	?
Suitable test or examination	?
Test acceptable to the population	✓
Natural history adequately understood	X
Agreed policy on whom to treat	X
The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

One controlled study examining the effectiveness of a home-based intervention to reduce language delay in very young infants identified by screening was identified (Ward 1999). The author developed and validated a screening questionnaire comprising six questions relating to receptive behaviours and a seventh relating to the onset of mixed babble. Health visitors screened 619 children attending 8-month distraction hearing testing in Manchester of whom 437 (71%) passed and were not assessed and 182 (29%) did not pass. Of these, 5 were lost to follow up, 37 were false positives on the Receptive Expressive Emergent Language Scales (REEL), 21 were excluded on the basis of NESB, hearing impairment, and/or developmental delay, and 119 (65%, or 19% of the original sample) were included in the study. Infants were alternately allocated to three experimental and three control groups based on their specific difficulties, each of which was stratified by severity of language delay on the REEL. Parents in the experimental group were visited at home by a speech pathologist at four weekly intervals for a variable duration (averaging four months of intervention involving an average of four home visits by the two workers). Parents were taught how to carry out specific play sessions comprising specific and suggested play activities, as well as suggestions for activities to be carried out in naturally occurring situations, such as showing the infant the sources of a sound, or naming objects of interest to the child. Attrition was 17% over the first year and 16% over the second year. Experimental and control groups were equivalent at study entry on the REEL. At three years of age, children were assessed using the Reynell Developmental Language Scales. 85% of controls and 5% of cases showed language delay. Mean receptive and expressive language quotients were 107 and 105 (case) vs 79 and 77 (control) respectively for Group 1 and 104 and 106 (case) vs 77 and 75 (control) respectively for Group 2; Group 3 was too small to allow analysis. The number of infants who did not attend distraction testing or who refused screening was not recorded. Assessments and interventions were carried out by the

same individuals, and no blinding occurred. These results are very encouraging, but the study needs to be replicated using a stronger evidence-based methodology, and the prevalence of language delay in the control group at 3 years of age appears to be unusually high. The resource implications of four home visits to 19% of the population are daunting at a population level, and study of similar interventions delivered in group-based parent training sessions may be worthwhile.

One report evaluated the effectiveness of a population screening program in detecting early language delay in 2 year old children in England. This program was conducted for 25 months between 1992 and 1993 and involved posting a parent report questionnaire (the Language Development Survey) and instruction sheet to the home of children in 2 rural communities within one month of their 2nd birthday. Parents were required to return their completed questionnaire using the reply paid envelope provided. Parents who had not returned their questionnaire after 2 weeks received a telephone reminder. All children who returned a positive test screen and a random sample of those who returned a negative test screen were asked to present for two 90 minute clinical evaluation sessions approximately one week apart. The clinical evaluation sessions consisted of a general history, an audiological screen (otoscopic examination and visual reinforcement audiometry), the Infant Mullen Scales of Early Learning and a conversational sample of the child's language being recorded and assessed for expressive vocabulary, average length of utterances and volubility. Parents and clinicians were blind to the child's screening test outcome until after the clinical examination (double-blind design). Following the clinical examination, two trained examiners made independent assessments of the child's language abilities (ie age appropriate or not age appropriate) blind to the screening test result and recommended either no further action, reassessment, or enrolment in an intervention program. Of the 650 questionnaires posted 582 were successfully delivered, of which only 306 (53%) were completed and returned. The prevalence of delay identified by the Language Development Survey using the criteria of <50 words or no word combinations was 15%. 17 children with positive screens (38% of those invited to attend) and 47 children with negative screens (no data on response rate) presented for a clinical evaluation. The Language Development Survey had a sensitivity of 91% and specificity of 87%, using the clinical evaluation as the gold standard. False positive results accounted for 11% of cases and false negative results for 2% of cases. It is estimated that the PPV of the Language Development Survey used as a population screening tool for 2 year old children would range between 18-37% (Klee et al. 1998).

Cost-effectiveness

Law and colleagues (Law et al. 1998) estimated the yield for 17 high quality studies of screening tests where normal populations were studied. They found a mean yield of 9.3% (range 3.6-12.6%). Yields exceeding 5% for screening tests or programs are considered good, therefore most of the screening tests included in the studies considered by Law and colleagues demonstrated good detection rates and in these terms could be considered cost-effective (Law et al. 1998). However, it is important to bear in mind the influence of prevalence on the yield. Language delay is a relatively prevalent condition, therefore the ability of tests to detect language delay are likely to be higher than for tests to detect a less prevalent condition.

Evidence suggests that 'indirect' treatment by a parent or teacher may be as effective as 'direct' treatment by a health professional (Law et al. 1998). This has implications for cost-effectiveness as 'indirect' treatment is likely to be considerably less expensive to administer than 'direct' treatment. Similarly, screening programs using parent and/or teacher questionnaire reports are generally a more cost-efficient method of population screening than tests requiring administration by a health professional (Klee et al. 1998).

Quality of evidence

PAPER	STRENGTH OF EVIDENCE			SIZE OF EFFECT	RELEVANCE OF EVIDENCE
	Level	Quality	Statistical precision		
Klee and colleagues		Independent blind comparison but not all subjects received gold standard. Poor response rate	n/a	n/a	

Generalisability of evidence

There is little evidence that the prevalence or aetiology of language delay varies from population to population, therefore the evidence should be generalizable. However, in a society such as Australia where there are large numbers of multi-lingual children and children from non-English speaking homes, there are likely to be issues surrounding the suitability and generalisability of English-based screening tests for these children and the suitability of standard screening thresholds. Screening questionnaires based on parent report will be dependent on literacy in the language in which they are written.

Further, the context in which screening takes place can have an impact on the results. Some children may not display their full repertoire of language skills in unfamiliar or less familiar contexts. For example children (particularly from less advantaged backgrounds) may demonstrate better language skills at home than at school or kindergarten (Law et al. 1998). This is not an issue with parent-reported screening questionnaires, which may however be influenced to a greater extent by parent literacy in English.

6.3.18.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening

Comments:

Difficulties with accurate definition of language delay and its natural history argue against the introduction of formal screening programs. In addition, it may be difficult to implement screening programs in the community to a sufficiently high standard to make it effective.

On the other hand, treatment appears to be beneficial for children with established language delay, and one (flawed) RCT has suggested a distinct benefit for very young children detected through screening.

6.3.18.4. RECOMMENDATIONS

Implementation of formal screening programs for language delay is not recommended at this stage.

Further research is urgently needed to better quantify early predictors of later language delay, and to confirm promising results from early, community-based secondary prevention programs.

6.3.18.5. FURTHER RESEARCH

- High quality cohort studies to better delineate the natural history of early language delay and its outcomes (language, social, academic) for at least the first seven years of life in Australian children. A major aim would be to study tools to reliably detect children with language delay that does not resolve and/or is associated with adverse outcomes. Until such tools exist, screening programs cannot be recommended.
- Community based RCTs of early language intervention via parents for very young children (8-24 months) at risk of language delay.

6.3.18.6. SEARCH TERMS

Language

Verbal learning

Communication disorders

Language development disorders

Language disorders

Speech disorders

Verbal behavior

Speech

Child language

Language development

6.3.18.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

Sheena Reilly – Professor of Paediatric Speech Pathology, Royal Children’s Hospital & La Trobe University, Melbourne.

John Fisher – Deputy Chief Speech Pathologist, Royal Children’s Hospital, Melbourne.

Frances Page Glascoe – Reference Group

6.3.18.8. REFERENCES

Bax, M., Hart, H. & Jenkins, S. 1980, "Assessment of speech and language development in the young child", *Pediatrics*, vol. 66, no. 3, pp. 350-354.

Dworkin, P. H. 1998, "Preventative Health Care and Anticipatory Guidance", *Handbook of Early Intervention 2nd edition*.

Glascoe, F. P. 1997, "Parents' concerns about children's development: Prescreening technique or screening test", *Pediatrics*, vol. 99, pp. 522-528.

Klee, T., Carson, D. K., Gavin, W. J., Hall, L., Kent, A., & Reece, S. 1998, "Concurrent and predictive validity of an early language screening program", *Journal of Speech, Language, & Hearing Research*, vol. 41, no. 3, pp. 627-641.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. 1998, "Screening for speech and language delay: a systematic review of the literature", *Health Technology Assessment*, vol. 2, no. 9.

Reep-van den Bergh, C. M. M., de Koning, H. J., de Ridder-Sluis, J. G., van der Lem, G. J. & van der Maas, P. J. 1998, "Prevalentie van taalontwikkelingsstoornissen bij kinderen (Prevalence of language disorders in children)", *Tijdschrift voor gezondheidswetenschappen*, vol. 76, pp. 311-317.

Thal, D. J. & Bates, E. 1991, "Language and gesture in late talkers; a one year follow up", *J Speech Hear Res*, vol. 34, pp. 604-612.

Ward, S. 1999, "An investigation into the effectiveness of an early intervention method for delayed language development in young children", *International Journal of Language & Communication Disorders*, vol. 34, no. 3, pp. 243-264.

6.3.19. HEIGHT

6.3.19.1. BACKGROUND

Clinical picture

Assessments of a child's growth usually involves serial measurements throughout infancy and childhood, which may be termed growth monitoring. Growth monitoring aims to detect either reduced or accelerated states of growth. Abnormal growth may occur from early infancy, or a child may initially grow normally and then deviate from normal growth (Hall 2000).

Tall and short stature both occur as part of normal variation of human growth. However, reduced growth can be a consequence of poor health, illness, social deprivation, feeding difficulties, and a range of specific conditions (isolated growth deficiency, multiple pituitary hormone deficiency, Turner's syndrome, Russell-Silver syndrome, skeletal dysplasias and bone disorders, Noonan's syndrome, neurofibromatosis, hypothyroidism, inflammatory bowel disease, coeliac disease, and chronic renal disease (Hall 2000). Accelerated growth can indicate thyrotoxicosis, pituitary gigantism, congenital adrenal hyperplasia, Marfan's syndrome, Sotos syndrome, Klinefelter syndrome, XXY syndrome, and premature sexual maturation (Hall 2000). Probably the most common cause of accelerated growth today is child obesity (Freedman, Khan, Serdula, Srinivasan, & Berenson 2000).

Growth screening or monitoring may be the clue to diagnosis of some individual cases of each of these conditions. However, almost all these conditions have characteristic signs and symptoms which more accurately lead to diagnosis. If any of these disorders justified specific screening activities aimed at their detection, one would choose more specific characteristics than the non-specific growth failure or faltering that may accompany them. Isolated growth hormone deficiency and Turner's syndrome are the only two conditions in which many of the cases exhibit substantial deviation from normal growth in the absence of other signs or symptoms (Hall 2000). In developing countries, growth monitoring can detect infants and children experiencing prolonged nutritional deprivation (Panpanich & Garner 1999).

Growth hormone deficiency or insufficiency (GHD)

GHD most commonly occurs without a clear cause (Hall 2000) and is due to insufficient production of or insensitivity to human growth hormone (Anonymous 2000a). Onset of signs can occur in infancy, early or mid-childhood or in adolescence. GHD almost invariably results in short stature, though some children are within the normal range of height into primary school years. After infancy, in which failure to thrive may be a problem, children with GHD are often relatively adipose and have facial features typical of younger children, but these features are insufficient to make a diagnosis (Hall 1996). GHD secondary to other conditions such as multiple pituitary failure or CNS conditions are usually detected clinically.

Turner's syndrome

Turner's syndrome occurs in females with only one X chromosome present in some or all cell lines. It results in short stature, delayed sexual maturation, and infertility. Learning difficulties, skeletal, heart and kidney abnormalities, and thyroid dysfunction are also common problems (Anonymous 2000b). It may be identifiable at birth from characteristic features such as lymphoedema of the hands, neck webbing, and pterygium colli. More than 50% of cases are identified in later childhood or adolescence when the presenting signs are short stature or primary amenorrhoea (Hall 1996). Adults not treated with growth hormone are usually well below the normal range of height, but much of this relative deficit develops in adolescence because of failure of the pubertal growth spurt. Comparing UK growth charts for normal children with UK charts for girls with Turner's syndrome, only 50% are below the 0.4th centile and only two thirds below the 2nd centile at 5 years (ie would be referred from screening programs at these cutpoints). Growth rate is only marginally slower than that of other children through the primary school years. Many women now have routine screening ultrasounds at about 11-12 weeks of pregnancy, principally aimed at detecting Down syndrome. Characteristic changes in the posterior neck may lead to chorionic villous sampling and chromosomal diagnosis prenatally. It is not yet known whether this will also substantially lower the number of girls born with undetected Turner's syndrome.

Natural History

Children grow at different rates, with periods of accelerated growth around particular ages. Abnormal growth may occur early in a child's life, or a child may begin growing normally and later deviate from normal growth (Hall 2000).

Much concern has been felt that very short or tall children may be subject to teasing and bullying by peers and suffer poor self esteem and other psychological and relationship problems. Traditionally, height has varied across social classes, with tall stature apparently tending to confer economic and other social benefits

for adults, particularly males. This may partly be due to complex factors contributing to shorter stature such as low birth weight, preterm birth and long term nutritional and environmental status. Recent work has shown that extremely short (<3rd centile) 11-13 year old children had a mean IQ approximately six points lower than matched normal-height controls, with reading attainment and basic number skills similarly reduced. However, self-esteem, self-perception, and parent perceptions of behaviour were similar for the two groups suggesting that even the shortest children do not show adverse effects in these areas (Downie, Mulligan, Stratford, Betts & Voss 1997). However, short children may be more prone to being bullied, even though they report having as many friends as taller controls (Voss & Mulligan 2000). Adverse effects of tall stature have not been well quantified in controlled community-based studies.

Prevalence

The number of children who fall above or below defined cut-points for height and weight varies according to the centile charts and cut-points used.

The prevalence of GHD has been estimated at 1 in 3000-5000 children (Hall 1996) with more boys than girls afflicted (Anonymous 2000a).

Turner's Syndrome occurs in at least 1 in 2500 females (Hall 1996).

Genetics relevant to screening issues

Although there has been considerable secular change in *actual* height over the last century, height *potential* is largely determined by genetic make-up. Therefore a child with shorter than average parents has a greater likelihood of being short than a child with parents of average height. This does not usually constitute a genetic defect, except when child and parents share a specific pathological condition causing growth disorder.

Some of the individual conditions that can be detected by growth monitoring involve a genetic defect or have a genetic component to their expression.

Turner's syndrome is a non-transmissible chromosomal disorder.

Diagnosis

There is much individual variation in normal growth both within and between children. While extreme abnormalities in growth are usually readily apparent, this makes the diagnosis of borderline "abnormal" or "non-optimal" growth very difficult. Diagnosis of abnormal growth is usually initially based on a child's measurement falling above or below a recommended percentile point on a growth chart.

Children with extremely short or tall stature may be referred for a specialist opinion, usually from an endocrinologist. This usually involves expert clinical assessment and auxologic assessment of height and weight, which may be followed by a range of diagnostic tests to determine whether there is a pathologic cause and if so what it is.

Treatment/Management

In Australia, human growth hormone therapy is currently available as a pharmaceutical benefit for children with short stature considered to be due to growth hormone deficiency (a combination of height less than the first centile and a growth rate below the 25th centile measured over a minimum period of one year), children with biochemical growth hormone deficiency, and children with Turner syndrome. It is also available to children with growth failure due to a range of specific pathologies which would not require screening (Commonwealth Department of Health and Aged Care 1999).

Growth hormone deficiency (GHD)

Effective treatment for GHD is available in the form of injection of a synthetic hormone which replicates the function of human growth hormone. Injections are usually required frequently (often on a daily basis) for a period of several years to be effective. Treatment started earlier leads to taller final height. Treatment may be more effective in secondary than primary GHD (Anonymous 2000a).

Turner's syndrome

There is no cure for Turner's syndrome, and management is symptomatic across a wide range of possible problems of which short stature is just one. Although not deficient in growth hormone, girls with Turner's syndrome can show improvements in growth if treated with synthetic growth hormone. Treatment with female sex hormones and thyroid hormone where appropriate can benefit height (Anonymous 2000b).

Short/tall normal individuals

Tall children (especially girls) are not infrequently referred for specialist assessment because of feared psychosocial repercussions of tall stature. Short normal children may seek specialist assessment for similar

reasons. In Australia, current Growth Hormone Advisory Board recommendations make GH therapy available to individuals without disease who have at least one year's auxologic data demonstrating height <1st centile and growth velocity/bone age <10th centile. This is in recognition of unreliability of GH testing and that growth per se is the most important indicator.

Prevention

Worldwide, the most effective method of preventing short stature is through improved maternal and child nutrition and reduction in infectious disease.

How might screening reduce the burden of suffering

Screening might identify children with short or tall stature due to a treatable cause, and thus normalise final height and avoid the postulated disadvantages of lifelong short or tall stature.

Treatment of normal short or tall children exposes them to medications or injections over a prolonged period of time, with risks of their specific side effects, discomforts and costs. It may also "pathologise" normal variation.

6.3.19.2 EVIDENCE

Tests

Height

Measured height is subject to variability. This variability is probably unimportant when describing populations and secular trends, but alters the accuracy of screening "pass/fail" criteria since even a small inaccuracy can result in many more or fewer healthy children being referred for specialist assessment. Voss has quantified the imprecision attributable to different sources of error in height measurement (Voss, Bailey, Cumming, Wilkin, & Betts 1990; Voss, Wilkin, Bailey, & Betts 1991). Some imprecision is due to differences in measurement techniques and inter-rater differences, which can be reduced through good training (eg always removing shoes, not using a "stretching" technique). An estimated 90% of variation is due to "real" changes due to children not having a "correct" height as they are not rigid objects (for example, diurnal variation can result in up to 2cm of height being "lost" over a day) (Hall 2000; Voss et al. 1990).

The greatest source of inaccuracy is through the use of inaccurate measuring equipment such as tape measures or uncalibrated wall-mounted measures. Measurement of height is best done using a rigid stadiometer. Several cheap portable measures are now available suitable for situations where the screener moves from site to site such as school-based screening, while well-calibrated wall mounted stadiometers are ideal for centre-based screening. Using such stadiometers with good technique, the standard deviation of a single height measure for a school aged child has been shown to be about 0.2-0.3cm, with a 95% confidence interval of ± 0.5 cm. A child observed to be on the 3rd centile for height therefore almost certainly lies somewhere between the 2nd and 4th centile. The interval width is similar for 3 years olds, but for 2 year olds it is double in size (± 1 cm).

Interpreting growth charts for children born prematurely (prior to 37 weeks of estimated gestational age), low birthweight (less than 2.5 kilograms) and/or small for gestational age can be problematic, particularly in the first few months of life. Once a child is two years of age standard growth charts for chronologic age are generally used as many infants experience "catch-up" growth during the first two years of life (Trachtenberg & Golemon 1998).

Growth charts

Growth charts are used to plot a child's height or weight or BMI measurement relative to a population dataset. Plotting and interpreting an individual's growth based on growth charts are open to error and therefore it is important that professionals are trained in the use of such charts. The number of children referred should directly reflect the cutpoints chosen in the screening program. Anecdotally in Australia many primary health care professionals also use "clinical judgment" as to who to refer, informally "taking into account" ethnicity, parent and professional concern about the child, how far below the cutpoint the child falls, and individual feelings about the value of such screening.

There is no standard "cut-off" for defining short stature or tall stature. Diagnosis is usually based on a child's measurement falling above or below a recommended percentile point on a growth chart. There are different recommended percentile cut-offs and different growth charts, all taking into account the child's age and gender. In Australia the National Centre for Health Statistics Growth Curves for Children are widely used (Hamill, Drizd, Johnson, Reed, Roche, & Moore 1979). These charts are based on US data from the early 1970s. Traditionally, it has been recommended that children falling below the 3rd centile on these charts be referred for further assessment. More recent nine centile charts developed from the UK 1990 Growth

Reference data are also available. Using the nine centile charts, it has been recommended that the 0.4th centile or 2nd centile can be used in a single screen to determine short stature requiring further investigation (Hall 2000). Similarly the 99.6th or 98th centile are most commonly used to determine tall stature.

By definition 0.4%, 2% or 3% of children will have height/weight measurements that fall below the true 0.4th, 2nd or 3rd centile for that population. However, the true centiles are often not known. Comparing a representative Victorian sample of 2862 primary school children from the 1997 Health of Young Victorians Study with the UK nine centile charts, 0.3% of children fell below the 0.4th centile and 1.3% below the 2nd centile, whereas 3.6% of children fell above the 98th centile and 1.2% above the 99.6th centile (unpublished data). Thus the recent general upward shift in children's height and weight leading to significant changes in prevalence at the tall end may not greatly have altered prevalences at the short end of centile charts.

It should be possible to determine Receiver Operating Characteristic (ROC) curves for height centile at different ages, and also to calculate false positive and false negative rates to determine optimal percentile cutpoints for referral. This has never been systematically attempted, although Hall goes part way toward this. He argues for screening taking each single height measurement as a separate screening test, using the 0.4th centile as a cut off point for short stature. In a hypothetical birth cohort of 100,000 children (of whom approximately 30 might have previously undetected GHD and 12 with Turner's syndrome) this would result in about 400 referrals, yet not all cases would be detected. If the cutpoint were raised to the 2nd centile, this would increase detection rate but would also increase referrals to about 2000 (Hall 2000).

Height velocity

Measures of height velocity in specialist endocrine clinics are an integral part of good tertiary practice. In community screening settings, in which height velocity of most children is normal but varies with height and over time, calculation of velocity from two time points may simply magnify the inaccuracies inherent in two screening measurements. For instance, a child on the 3rd centile for height on two successive occasions might be estimated to be growing at a marginal 25th centile for velocity (the rate of growth required to keep him on the 3rd height centile). Voss estimates, from 95% confidence intervals using a height SD of 0.25cm for each of the two single measures, that his growth velocity might in reality fall anywhere between the 8th and the 50th centiles – ie from clearly abnormal to clearly normal (Voss 1995). The Wessex study has clearly shown that, while successive heights are highly correlated, successive 12 month velocities are not. Furthermore, what is satisfactory varies by height and by age.

It has been suggested that a change of more than plus or minus one centile band might be appropriate grounds for referral. However, this criterion appears to be both insensitive (eg many prepubertal children of borderline height with Turner's syndrome would not be detected over a 12 month period; normal height children developing illness in the Wessex study did not cross centile bands) (Voss 1999) and nonspecific (eg one would expect 2% of all children to cross one centile band downward between 5 and 8 years of age, while 2% would cross one band upwards) (Hall 2000). Time between measurements would also influence the implications, eg a fall over two years would represent a greater faltering in growth than a fall of the same magnitude occurring over three years.

Parent height

It has been suggested that parent height should be taken into consideration when differentiating children with true growth problems from those with a heritable normal stature. New UK screening charts are under preparation which adjust for parent and sibling height. These have not yet been trialed (Hall 2000). Parent reporting of their own and partner's height is reasonably accurate but may not be sufficiently accurate for these purposes.

Does this condition meet criteria for a screening program?

Growth screening to detect growth hormone deficiency and Turner syndrome

Characteristics of a Screening Test	
Cochrane & Holland 1971	
Simple, quick and easy to interpret	?
Acceptable to the public	✓
Accurate	?
Repeatable	?
Sensitive	Depends on cutpoint
Specific	✗

Criteria for a Screening Program	
Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✓
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✗
Test acceptable to the population	✓
Natural history adequately understood	✓
Agreed policy on whom to treat	✓
The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

The vast majority of children with short or tall stature are normal, healthy children. This has implications in terms of defining who to refer, use of resources, and minimisation of needless anxiety and discomfort/distress in the investigation of those detected.

If one is considering a screening activity, the question becomes whether measurement of height at single specified time or age is a useful exercise. Growth screening aims to detect either absolute tall or short stature. Growth monitoring programs involve the regular measurement of height and weight by health care professionals, often during routine visits. Measurements are plotted on appropriate growth charts which it is hoped will allow identification of abnormal growth. If abnormalities in the pattern of growth are detected, further investigations are conducted to identify the source of the problem and the appropriate management. However, growth monitoring (particularly if conducted frequently) can lead to parent anxiety that their child is not growing fast enough or is growing too fast (Panpanich & Garner 1999), particularly if parents do not know that child growth often occurs in “sprints”.

We identified no randomised controlled trials or controlled clinical trials of height screening or growth monitoring in developed countries, nor did a recent systematic review of growth monitoring (Panpanich & Garner 1999).

Yield has been calculated from several large scale population-based studies, none of which have identified false negative rates or specificity.

In the Utah Growth Study, trained volunteers measured 114,881 children aged kindergarten-5th grade. A further 4.9% of the population was absent for illness and 7.3% of measurements were incomplete. Of the children measured, 1.2% had values more than 2SD below the mean for height. 79,495 children were remeasured 12 months later. Prevalence of GHD was 1:3480, of whom 17 had been previously diagnosed and 16 were detected through the program. Six girls with Turner syndrome were detected. Unfortunately, it was impossible to tell from the study which children were identified through absolute height <-2SDs at the initial screen and which from growth velocity <5cm/year and height <3rd centile (Lindsay, Feldkamp, Harris, Robertson, & Rallison 1994).

The Oxford district growth screening program was established in 1988 to screen the heights of all children aged 3 and 4.5 years. In the first three years 20,338 children (an uptake of only 61-74%) were screened for a yield of two children with previously undiagnosed GHD and two girls with Turner’s syndrome (Ahmed, Allen, Sharma, Macfarlane, & Dunger 1993).

Since 1986, the Wessex Growth Study has monitored 147 children who were very short (< 3rd centile) at school entry, along with matched controls (between 10th and 90th centiles). This study has not shown that repeated height measurements reliably identifies new disease, found height velocity to be unreliable, and concluded that height screening at school entry is the best means of identifying silent disease in school age children. Eight (5.4%) had previously undetected disease, remediable in four. Two of these would have been missed by dropping the referral criterion to the 0.4th centile (Voss 1999).

In Australia, many areas ceased universal height screening at school entry in the 1990s. While there continues to be concern among health professionals and parents about this change, there is (as yet) no evidence that age at detection has risen in the years since cessation of school entry screening.

Cost effectiveness

We identified no formal cost-effectiveness studies. Although often part of routine health care visits, growth monitoring incurs costs in terms of health professional and parent time, equipment and training costs and the costs of following up children whose growth appears abnormal (Panpanich & Garner 1999). At present evidence as to the cost-savings involved with the early detection of problems is lacking. Growth hormone treatment is very expensive (Anonymous 2000a), an issue if many short normal children identified only by screening seek to use it.

Generalisability of evidence

The evidence is generalizable to the Australian context.

6.3.19.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening

Comments:

We could identify little evidence demonstrating the effectiveness of this activity. Height screening invariably leads to large numbers of referrals of “short-normal” children; the cost, acceptability and uptake of referrals has not been quantified; and fundamental requisites of the screening test (sensitivity, specificity, positive and negative predictive value, and NNT) are yet to be quantified for different centile cutpoints.

Nonetheless, screening for short stature due to growth hormone deficiency is felt by many to be of benefit. With many areas now without routine height screening programs, formal randomised controlled trials incorporating economic analysis could be mounted to answer what is seen by many to be an important question.

Regular population-based surveys of child height are important at a population level to understand secular trends and to better understand long term nutritional status of the child population and of disadvantaged subgroups.

7.3.19.4. RECOMMENDATIONS

New growth screening programs outside the research context are not recommended.

Should retrospective review suggest that age of diagnosis of growth hormone deficiency and Turner syndrome are rising following recent cessation of height monitoring, further study of the usefulness of height measurement as a screening tool should be undertaken.

Regular, systematic population-based surveys of height are recommended to monitor secular trends in height for the whole population and for subgroups at particular risk.

7.3.19.5. FURTHER RESEARCH

Retrospective study in areas that have discontinued school entry height screening of secular trends in age of diagnosis of growth hormone deficiency and Turner syndrome, with and without screening

Properties of height measurement as a screening tool at various centile cutpoints for various ages (sensitivity, specificity, PPV, NPV, NNT)

RCT to determine whether school entry height screening is effective in lowering of age of diagnosis of growth hormone deficiency and Turner syndrome in school age children; benefits and harms; uptake of referrals; costs

Regular, systematic population-based surveys of height to monitor secular trends in height for the whole population and for subgroups at particular risk.

7.3.19.6 SEARCH TERMS

Growth
Growth disorders
Growth monitoring
Obesity
Body height
Body weight
Body constitution
Failure to thrive
Anthropometry

7.3.19.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

John Alzoizos – Reference Group

David Hall – Reference Group

Fergus Cameron – Dept of Endocrinology and Diabetes, Royal Children's Hospital, Melbourne

7.3.19.8 REFERENCES

Ahmed, M. L., Allen, A. D., Sharma, A., Macfarlane, J. A., & Dunger, D. B. 1993, "Evaluation of a district growth screening programme: the Oxford Growth Study", *Arch Dis Child*, vol. 69,no. 3, pp. 361-365.

Anonymous. 2000a, "What is GHD? Diagnosis. Treatment. Problems with Treatment", <http://www.mhamer.freeserve.co.uk>, last updated Jan 2000.

Anonymous. 2000b, Endocrinology and Turner's Syndrome. "Endocrinology and short stature", Endocrinology Society, <http://www.endo-society.org/pubaffai/factshee/shrstat.htm> last updated Jan 2000.

Commonwealth Department of Health and Aged Care. 1999, "Guidelines for the Availability of Human Growth Hormone (hGH) as a Pharmaceutical Benefit".

Downie, A. B., Mulligan, J., Stratford, R. J., Betts, P. R., & Voss, L. D. 1997, "Are short normal children at a disadvantage? The Wessex growth study", *BMJ*, vol. 314,no. 7074, pp. 97-100.

Freedman, D. S., Khan, L. K., Serdula, M. K., Srinivasan, S. R., & Berenson, G. S. 2000, "Secular trends in height among children during 2 decades: The Bogalusa Heart Study", *Arch Pediatr Adolesc.Med*, vol. 154,no. 2, pp. 155-161.

Hall, D.M.B. 1996, *Health for All Children*, 3rd edn, Oxford University Press, New York.

Hall, D.M.B. 2000, "Growth monitoring", *Arch Dis Child*, vol. 82, pp. 10-15.

Hamill, P. V., Drizd, T. A., Johnson, C. L., Reed, R. B., Roche, A. F., & Moore, W. M. 1979, "Physical growth: National Centre for Health Statistics percentiles.", *American Journal of Clinical Nutrition*, vol. 32,no. 3, pp. 607-629.

Lindsay, R., Feldkamp, M., Harris, D., Robertson, J., & Rallison, M. 1994, "Utah Growth Study: growth standards and the prevalence of growth hormone deficiency", *J Pediatr*, vol. 125,no. 1, pp. 29-35.

Panpanich, R. & Garner, P. 1999, "Growth monitoring in children", *The Cochrane Library (Oxford)*, vol. Issue 3 Oxford: Update Software.

Trachtenbarg, D. E. & Golemon, T. B. 1998, "Care of the premature infant: part I. Monitoring growth and development", *American Family Physician*, vol. 57,no. 9, pp. 2123-2130.

Voss, L. D. 1995, "Can we measure growth?", *J Med Screen.*, vol. 2,no. 3, pp. 164-167.

Voss, L. D. 1999, "Changing practice in growth monitoring [editorial]", *BMJ*, vol. 318,no. 7180, pp. 344-345.

Voss, L. D., Bailey, B. J., Cumming, K., Wilkin, T. J., & Betts, P. R. 1990, "The reliability of height measurement (the Wessex Growth Study)", *Arch Dis Child*, vol. 65,no. 12, pp. 1340-1344.

Voss, L. D. & Mulligan, J. 2000, "Bullying in school: are short pupils at risk? Questionnaire study in a cohort", *BMJ*, vol. 320,no. 7235, pp. 612-613.

Voss, L. D., Wilkin, T. J., Bailey, B. J., & Betts, P. R. 1991, "The reliability of height and height velocity in the assessment of growth (the Wessex Growth Study)", *Arch Dis Child*, vol. 66,no. 7, pp. 833-837.

6.3.20. WEIGHT

6.3.20.1. BACKGROUND

Clinical picture

The two main conditions detected by weight monitoring are failure to thrive and obesity (Panpanich & Garner 1999).

Failure to thrive (FTT)

Failure to thrive is a condition characterised by a failure of expected growth (usually weight), usually identified in infancy or toddlerhood. Children with severe FTT are usually readily identified clinically, but there is no universally accepted definition of FTT. Many texts provide no definition at all.

Traditionally, FTT has been subdivided into organic or non-organic in nature. This distinction rarely appears to be helpful. Wright reports that three separate population based studies have found $\leq 5\%$ to have major organic disease, mostly diagnosable from other symptoms or signs, while even in the presence of organic disease the underlying cause still tends to be undernutrition. While one case control study found that children with FTT were four times more likely to be abused than controls, population studies suggest that only 5-10% of children with FTT have been suspected of abuse or neglect. Indices of social deprivation do not appear to be strongly linked with FTT. Conversely, undernutrition is strongly implicated in most cases of FTT. In Wright's clinical study, moderate to strong evidence of undernutrition was eventually found in two thirds of the children, based on a combination of history, weight gain patterns, and food diaries (Wright 2000).

Obesity

Obesity is a surplus of body weight resulting from excess accumulation of body fat. Obesity increases the risk for developing other disease including diabetes, hypertension, stroke, heart disease, certain types of cancer, gall bladder disease, arthritis and sleep apnoea (Anonymous 1999). The World Health Organisation has named obesity as one of four major epidemics to be faced in the next century (World Health Organization, Division of Noncommunicable Diseases 1998).

Childhood overweight/obesity is associated with a range of adverse health effects that can affect children in the short and longer term. Psychosocial effects of social isolation, discrimination and peer problems in childhood and self-esteem problems in adolescence are more common in fatter children, though the causal direction for these links is not entirely clear (for example, there is some evidence to suggest that high self-esteem is protective against obesity) (Power, Lake, & Cole 1997). Obesity is directly related to cardiovascular risk factors (higher insulin, cholesterol, lipid and blood pressure levels in children and adults) which fall with resolution of obesity. Physical health problems with long term implications include poorer pulmonary function, advanced growth, hyperlipidaemia, glucose intolerance, hepatic steatosis and cholelithiasis, and a wide range of less common pathologic conditions. Overweight/obesity appear to be associated with reduction in level and amount of physical activity (Davies, Gregory, & White 1995; Eck, Hackett-Renner, & Klesges 1992), physical functioning, and participation in active sporting activities.

However, in the 1997 Health of Young Victorians Study of over 2,800 primary school children in Victoria, Australia, most parents of overweight and obese children saw their children as enjoying good health across a wide range of physical and psychosocial health domains, and many did not report concerns about their child's weight. Overweight parents differed little from nonoverweight parents in their perceptions of their overweight or obese children's health (Wake, Salmon, & Waters 2000).

Natural history

Failure to thrive (FTT)

Onset often occurs within weeks of birth and with hindsight growth faltering is clearly evident on growth charts by six months. However, in the large Newcastle study the majority of cases did not present till after 12 months of age (Wright & Talbot 1996). Failure to thrive often persists up to the age of 5 years (Cole 1997).

Over the preschool years FTT tends to improve without formal intervention strategies, with mean weight in two studies rising from about the 2nd centile to about the 9th centile by 4 years (Kristiansson & Fallstrom 1987 as cited in Wright 2000; Wright, Callum, Birks, & Jarvis 1998a). In a case control follow up of the Newcastle birth cohort of term babies screened for failure to thrive in infancy, cases were on average 4.4cm shorter at 7-9 years of age after adjusting for parental heights. Cases had smaller head circumferences (mean 51.9 vs 52.8 cm), were significantly lighter (medians 23.8 vs 27.9 kg) and had lower body mass index (medians 14.9 vs 16.3 kg/m²). Despite these large growth differences, failure to thrive in infancy was not associated with cognitive or educational disadvantages at school age (Drewett, Corbett, & Wright 1999). These findings seem more robust than earlier suggestions that poorer cognitive development may be a long

term effect of FTT (Dowdney et al 1987 & Skuse et al 1994 as cited in Wright & Talbot 1996) with the degree of deficit linked to the length and severity of the poor weight gain (Wright & Talbot 1996) but reversible with treatment (Kristiansson & Fallstrom 1997 as cited in Wright & Talbot 1996). Relatively greater weight than height deficits suggested that poor appetite, and therefore poor intake, was a lasting outcome of FTT, (Drewett, Corbett, & Wright 1999) but Wright's RCT, discussed below, suggested that this was modifiable at least in the preschool years (Wright et al. 1998a).

Little is known about the health effects of being underweight after infancy and early childhood. Meaningful BMI cutpoints for child underweight have yet to be defined and their short and long term relationships with child health status studied. The only large epidemiological sample studying adolescents (not children) concluded that the great majority of thin 17 year old boys were healthy, though asthma, scoliosis, intestinal conditions and emotional disorders were reported to be more common in the underweight boys (Lusky, Barel, Lubin, Kaplan, Layani, Shohat, Lev, & Wiener 1996). Once serious illness has been excluded, there is no evidence that thinness per se leads to morbidity in adults (Willett, Dietz, & Colditz 1999) or children. In the 1997 Health of Young Victorians Study, parents reported underweight girls to have poorer health than normal weight girls on the physical functioning, role physical, general health and parent impact emotional scales, and underweight boys were reported to have poorer health on the bodily pain scale (Wake, Salmon, & Waters 2000).

Obesity

There is no particular age of onset for obesity; children can become obese at any time during their life. Childhood obesity is a risk factor for adult obesity, which is compounded if one or both parents are also obese. Whitaker found that the odds of an obese child being an obese young adult rises with age of the child. Odds ratios for obese children becoming obese young adults were 1.3 at 1-2 years, 4.1 at 3-5 years, 10.3 at 6-9 years, 28.3 at 10-14 years, and 20.3 at 15-17 years (Whitaker, Wright, Pepe, Seidel, & Dietz 1997).

Prevalence

The number of children who fall above or below defined cut-points for height, weight and BMI varies enormously according to the centile charts and cut-points used.

Failure to thrive

The prevalence of FTT is artificial in that most definitions use centile cutpoints (for example, 1st, 3rd or 5th centile) which determine that a set proportion of the population will always have FTT, whether using absolute weight or weight change criteria. Cutpoints based on meaningful differences in outcomes have not been determined in prospective studies.

Obesity

Recently, the International Obesity Taskforce has developed an international reference standard based on a combination of six large (>10,000) population-representative samples of children using transformed BMI curves to allow parametric data analysis. Cutpoints for overweight and obesity are based on the points at which a child population's BMI centile curves intersect with BMI values of 25 and 30 respectively (the adult cutpoints) at age 18 years. These points have been chosen because they are fixed and therefore referable over time, are widely accepted and have a quantifiable relationship with morbidity and mortality in adults (Cole, Bellizzi, Flegal & Dietz 2000).

Based on these new cutpoints, approximately 5% of NSW and Victorian children were obese in 1997, with a further 17% overweight (Booth, Wake, Armstrong, Chey, Hesketh, Mathur 2001). This represents a significant increase in prevalence in Australian children over the last 15-20 years (Lazarus, Wake, Hesketh, & Waters 2000) as elsewhere (Troiano & Flegal 1998).

Genetics relevant to screening issues

Both height and weight are largely determined by genetic make-up. Results from twin studies suggest that genetic factors explain 50-90% of the variance in BMI, while family studies of parent-offspring and sibling BMI correlations suggest heritabilities of 20-80% (Maes, Neale, & Eaves 1997). Heritability includes the way in which individuals respond to their environment, ie certain families are much more likely than others to become obese given an "obesogenic" environment. In the coming decade it is likely that genes associated with this environmental susceptibility will be identified, which may enable development of targeted preventive and management interventions.

Some of the individual conditions leading to failure to thrive or obesity involve a genetic defect or have a genetic component to their expression, eg Prader Willi syndrome, cystic fibrosis.

Diagnosis

Diagnosis of FTT and overweight/obesity is by weighing, and usually measuring, children and then comparing to a reference standard.

Failure to thrive

Both height for age and weight for age can be affected by inadequate nutrition in childhood. Conceptually therefore length/height for age and a measure of weight for height such as BMI could be considered in diagnosis as well as weight for age. In practice the definition of FTT is usually made on weight alone in infants and young children in developed countries.

FTT is commonly defined by an absolute weight criterion, for example, weight below the 5th or the 3rd centile on a given growth chart. Such a definition will identify many children with FTT, but is likely to identify false positives (naturally small children, particularly when minor weight variations occur) while missing naturally tall children with significant FTT (Wright, Matthews, Waterston, & Aynsley-Green 1994).

In clinical practice, the diagnosis is more likely to be based on a fall on a centile chart. What constitutes a fall has not been well defined and is often left to clinical judgement. Recently, attempts have been made to quantify this more exactly. Based on large-scale prospective population data, 5% of British infants have been reported to shift up or down two intercentile spaces on the UK 1990 9-centile charts between birth and six weeks of age (ie a change of ± 1.34 SD). After this age there is less variability and only 5% of children cross through two intercentile spaces between 6 weeks and 12 months, and only 1% cross three centile spaces (Wright et al. 1994). The Parkin Service in Newcastle, UK, has incorporated these figures into a definition of FTT of a fall of 1.26 SDs in weight (equivalent to a baby on the 50th centile falling to between the 3rd and 10th centiles) between these ages which identifies the 5% of children with slowest gain. Both a computer algorithm and a weight monitoring chart for children with slow weight gain have been developed, both adjusting for the regression to the mean which makes large babies more likely to fall across weight centiles than small babies (Wright 2000). As far as we know, this diagnostic approach has not been widely used elsewhere.

Obesity

The most commonly used expression of weight relative to height is the body mass index (BMI) calculated as weight (kg) divided by height (m^2). Similar cut-offs are used for body mass index (BMI) as for height to determine overweight and obesity.

Diagnosis is usually based on a child's measurement falling above a recommended percentile point on a weight or BMI chart. There are different recommended percentile cut-offs and different growth charts, all taking into account the child's age and gender. In Australia the National Centre for Health Statistics Growth Curves for Children are widely used (Hamill, Drizd, Johnson, Reed, Roche, & Moore 1979), based on US data from the early 1970s. More recent nine centile charts developed from the UK 1990 Growth Reference data are also available. Both now have corresponding BMI curves available. New International Obesity Taskforce cutpoints are now available by 6-month age brackets, but have not yet been translated into BMI charts.

There are currently no agreed definitions of overweight and obesity for Australian children. The Commonwealth Department of Health & Aged Care recently awarded a tender to develop Australian Standard Definitions for child and adolescent overweight and obesity, expected to be complete by end 2001.

Treatment/Management

Failure to thrive (FTT)

For the minority of cases of FTT in which an organic cause, abuse or neglect are clearly implicated, these issues need to be dealt with first.

Since almost all children with FTT show evidence of undernutrition, management should be directed toward increasing the child's nutritional intake, educating the parents about the child's nutritional needs, and providing behavioural advice and family support (Wright & Talbot 1996). Improvement in growth should be evident approximately 1-3 months following initiation of treatment (Wright & Talbot 1996). One randomised controlled trial examined the effectiveness of a standardised community-based health visitor intervention backed up by dietetic, paediatric and social work support where necessary (Wright et al. 1998a). By 3-4.5 years of age, FTT had resolved in 76% of cases compared to 55% of controls, with cases on average approximately 0.3 SDs taller and heavier than controls, though still shorter and lighter than means for the national reference population. However, only 58% were followed to this endpoint. An RCT of regular lay home visiting over the first year of life provided to poor children with weight <5th centile (as opposed to a

drop across centiles) did not show an intervention effect on weight, though both case and control children improved (Black, Dubowitz, Hutcheson, Berenson-Howard, & Starr, Jr. 1995). Similar findings were published from an RCT evaluating the effectiveness of home visiting over a one year period for children referred with FTT to a specialist clinic in the UK (Raynor, Rudolf, Cooper, Marchant, & Cottrell 1999).

Obesity

So far, strategies for the prevention and treatment of obesity in children have had little impact on current upward trends in prevalence of childhood obesity. School-based prevention and family-based treatment interventions have shown some short-term effectiveness (Robinson 1999) and longer-term outcomes are awaited. For the prevention of childhood obesity family therapy has shown some promise (Glenny, O'Meara, Melville, Sheldon, & Wilson 1997). Treatment interventions involving lifestyle modification such as the reduction of sedentary behaviour may be effective (Glenny et al. 1997; Edmunds & Waters 2000). Two major Cochrane systematic reviews of the effectiveness of strategies to prevent and manage childhood obesity will be published late in 2000/2001.

Prevention

Cochrane systematic reviews examining prevention and treatment of obesity in children are currently being prepared. These may provide more information on the effectiveness of prevention (and treatment) options.

How might screening reduce the burden of suffering?

Failure to thrive

Screening might be beneficial if systematic early detection of failure to thrive led to management that avoided subsequent adverse outcomes. At present this is problematic since definitions, outcomes, and effectiveness of management of FTT are not clear.

Weight monitoring (particularly if conducted frequently) can lead to parent anxiety if a baby is seen not to be gaining weight fast enough or too fast (Panpanich & Garner 1999), particularly if parents do not know that child growth often occurs in "spurts". This can result in parents overfeeding or force-feeding their child if they believe they are not growing fast enough (Hall 1996). In addition, it has been estimated that health visitors are already aware that infants are growing poorly in approximately 80% cases identified by screening (Wright & Talbot 1996).

Obesity

Screening might be beneficial if systematic early detection of obesity led to management that avoided subsequent adverse outcomes. At present this also is problematic, since it is far from clear who will suffer adverse outcomes, or how to manage and advise children with detected obesity. It is also likely that children who are obese would be obvious to a health professional without the need to formally weigh and measure the child.

6.3.20.2 EVIDENCE

Tests

Weight

Measurement of weight can be done using any good-quality scales. Digital bathroom scales, balance scales and chair scales are most commonly used. Less imprecision is involved with weight than height measurement, provided clothing does not vary much and scales are regularly calibrated.

BMI

Relative weight is more commonly used than absolute weight, usually body mass index calculated as weight divided by height squared. The maths involved in calculation of BMI is a potential source of error in screening programs and nomograms may be a safer way of arriving at an individual's BMI. Issues in accurate measurement of height have been discussed in section 6.3.20 "Height".

BMI varies by height and age. For example, data from the Wessex growth study show that 10% of short normal children – compared with 27% of controls – had a BMI $\geq 75^{\text{th}}$ centile at the age of 9 years. To have a BMI on the 99.6th centile the discrepancy between weight and height for a child of average height is around three centile bands, regardless of age. Short children require weight to be at least four centiles above height, while for tall children it is correspondingly less. Mulligan and Voss have suggested that this is likely to lead to the underidentification of short obese children and the overidentification of tall children (Mulligan & Voss 1999). There is also concern that BMI may systematically over- or under-estimate adiposity in children in various subgroups, for example particular body builds or ethnic groups. Nonetheless, it is generally accepted

that for ease of use and repeatability BMI is the current measure of choice for large scale community based measurement.

Skin fold thickness, waist:hip ratio

Skin fold thicknesses are slightly more accurate than BMI in predicting body fatness, but are less appropriate for use on a large scale in the community than BMI measurements and may be unacceptable to younger children. Distribution of body fat is related to cardiovascular risk. However, neither child skin fold thickness nor waist:hip ratio has yet been linked quantifiably to risk.

The Thrive Index and the Weight Monitoring Chart

The Thrive Index, developed by Wright and colleagues attempts to define FTT more accurately recognising that, because of regression to the mean, large babies tend to move down percentiles after birth while small babies tend to move up percentiles. The index uses a weight or weights in the early weeks of life as a baseline for comparison with subsequent weights after adjustment for regression to the mean. Based on population data, it uses transformed weight expressed as a standard deviation score and is calculated as *Current Weight SD – (weight at 6-8 weeks of age SD x 0.4)*. It requires the transformation of weights into SD scores using a computer program (Wright, Avery, Epstein, Birks, & Croft 1998b).

Recently, the Weight Monitoring Chart has been developed to assist professionals to determine whether an infant has FTT or not. Based on the Thrive Index and developed by the same authors, it looks like a standard centile chart and is used similarly. A report by the authors indicated that its use resulted in health professionals classifying hypothetical children as having FTT more accurately (51% accurate vs 28%) (Wright et al. 1998b). However, the chart has not yet undergone further scrutiny and the concepts involved in its use may make it unsuitable for large scale screening.

Growth charts

Growth charts are used to plot a child's height or weight or BMI measurement relative to a population dataset. These give a reasonably clear picture of where a child's height or weight falls relative to the large populations comprising the reference or standard for that chart. Frequently one must interpret findings for a child in one's own population in comparison to populations of children removed ethnically, geographically, and in time. Further, growth charts separate normal from abnormal only by empirical percentile cutpoints, rather than by known risk of adverse outcomes.

Plotting and interpretation an individual's growth based on growth charts is also open to error. It is important that professionals are trained in the use of such charts.

Comparisons of different methods in the diagnosis of FTT

At least two recent attempts have been made to compare methods of categorising undernutrition by different methods (Raynor & Rudolf 2000; Wright, Ashenburg, & Whitaker 1994). Both suffered from being trialed only on children referred for evaluation for FTT in specialist clinics and from lack of a gold standard. Proportions of referred children assessed as having FTT varied widely between the various methods used. Both groups concluded that existing methods to categorise children were likely to be unreliable.

Does this condition meet criteria for a screening program?

(i) Failure to thrive

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✗
Acceptable to the public	✓
Accurate	✓
Repeatable	✓
Sensitive	✗
Specific	✗

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	?
Facilities for diagnosis and treatment	✓
Latent or early symptomatic stage	✓
Suitable test or examination	✗
Test acceptable to the population	✓
Natural history adequately understood	✗
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	?

(ii) Obesity

Characteristics of a Screening Test Cochrane & Holland 1971	
Simple, quick and easy to interpret	✓
Acceptable to the public	?
Accurate	✓
Repeatable	✓
Sensitive	✓
Specific	✓

Criteria for a Screening Program Wilson & Jungner 1968	
Important health problem	✓
Accepted treatment	✗
Facilities for diagnosis and treatment	✗
Latent or early symptomatic stage	?
Suitable test or examination	✓
Test acceptable to the population	?
Natural history adequately understood	✗
Agreed policy on whom to treat	✗
The cost of case-finding balanced with expenditure on medical care as a whole	?

Programs

Failure to thrive

A single randomised controlled trial of screening followed by an early intervention was identified (Wright et al. 1998a). In one health district in the UK, children were identified for inclusion into the study based on a screening program involving each child born in the district being weighed by a health visitor at least twice and their weights entered into the district child health computer. Children were first weighed at their 6-8 week check and again when they were between 9 and 18 months of age. Children were identified by the computer as failing to thrive if the standard deviation score of their second weight indicated a fall from the baseline weight of ≥ 1.26 standard deviations after adjustment for regression to the mean. During the two years this screening program was undertaken, 229 children (3%) were identified who fitted the criteria of failure to thrive. By 3-4.5 years of age, FTT had resolved in 76% of cases compared to 55% of controls, with cases on average approximately 0.3 SDs taller and heavier than controls, though still shorter and lighter than means for the national reference population.

A recent systematic review of growth monitoring (Panpanich & Garner 1999) identified one randomised controlled trial of a weight monitoring program in rural India and one quasi-randomised trial in Lesotho of a program teaching mothers how to interpret weight charts. No difference was seen in weight gains though mothers in the latter trial had slightly higher knowledge scores. These programs are unlikely to be relevant to the Australian setting.

Obesity

Although obesity may have immediate physical and psychosocial ramifications in childhood, the major costs are incurred through the resultant adult excess mortality and morbidity. Therefore, ideally screening would identify (a) those most at risk of adverse effects and/or (b) those least likely to remit spontaneously by adulthood, while (c) minimising false positives. It is not yet possible to accurately determine membership of these groups.

Many community-based epidemiologic studies have reported prevalences of childhood overweight and obesity for different populations (using many different definitions and standards), and some have also quantified risk factors associated with various cutpoints of BMI. However, our literature search revealed no reports of service-oriented screening programs for obesity. There is considerable debate as to the possible harms and the possible benefits such programs might engender.

It is usually very clear without screening activities that a child is obese. It seems likely that population approaches may be more successful than strategies to improve detection of individual obese children, particularly since there is current evidence that many children and their parents are not concerned about their weight. For individual counselling and advice in the clinical setting, practitioners are likely to need training in the best strategies based on current evidence translated into methodologies suited to primary care.

Cost effectiveness

We identified no formal cost-effectiveness studies.

Generalisability of evidence

This evidence is generalisable to the Australian context.

6.3.20.3. CONCLUSIONS

Insufficient evidence to make a recommendation for or against screening for failure to thrive

Comment: while there is some evidence that systematic detection and management of failure to thrive in the first year of life can improve outcomes, this has yet to be operationalised into a practicable screening system.

Fair evidence to recommend against screening for obesity

Comment: childhood overweight and obesity are emerging as a public health problem of a massive dimension, with incalculable future costs. If effective and accessible therapy becomes available, this recommendation should be urgently reviewed.

6.3.20.4. RECOMMENDATIONS

Routine weight monitoring at birth, at 6-8 weeks, and at 8-12 months is recommended as part of routine clinical care. This does not however constitute a screening program.

6.3.20.5. FURTHER RESEARCH

Failure to thrive: further research into how to accurately define and how to manage FTT should be a prerequisite to any screening activities.

Obesity: regular, systematic population-based surveys of height, weight and BMI are recommended to monitor secular trends in overweight and obesity for the whole population and for subgroups at particular risk.

Prevention appears to be the most promising approach to childhood obesity. For established childhood obesity, further research into better definitions and effective management strategies should be a prerequisite to any screening activities. Research is urgently needed to define:

- a) which children are most at risk of obesity persisting into adulthood
- b) which children are most at risk of adverse outcomes of obesity in adulthood

- c) effective management strategies suitable for large scale use in the primary care setting
- d) the accept ability, benefits, harms and uptake of such activities

6.3.20.6. SEARCH TERMS

Growth

Growth disorders

Growth monitoring

Obesity

Body height

Body weight

Body constitution

Failure to thrive

Anthropometry

6.3.20.7. CIRCULATION OF DRAFT FOR COMMENT

Drafts of this document were circulated to the following people with an invitation to comment:

John Alzoizos – Reference Group

David Hall – Reference Group

Fergus Cameron – Dept of Endocrinology and Diabetes, Royal Children’s Hospital, Melbourne

6.3.20.8. REFERENCES

Anonymous. 1999, "On-line Medical Dictionary", <http://www.graylab.ac.uk/omd/index.html>, information accessed December 1999.

Black, M. M., Dubowitz, H., Hutcheson, J., Berenson-Howard, J. & Starr, R. H. 1995, "A randomized clinical trial of home intervention for children with failure to thrive", *Pediatrics*, vol. 95,no. 6, pp. 807-814.

Booth, M. L., Wake, M., Armstrong, T., Chey, T., Hesketh, K., & Mathur, S. 2001, "The epidemiology of overweight and obesity among Australian children and adolescents, 1995-1997", *Australian and New Zealand Journal of Public Health*, vol. 25, no. 2, pp. 162-169.

Cole, T. J. 1997, "3-in-1 weight-monitoring chart", *Lancet*, vol. 349,no. 9045, pp. 102-103.

Cole, T. J., Bellizzi, M. C., Flegal, K. M., & Dietz, W. H. 2000, "Establishing a standard definition for child overweight and obesity worldwide: international survey", *British Medical Journal*, vol.320, pp. 1240-1243.

Davies, P. S., Gregory, J., & White, A. 1995, "Physical activity and body fatness in pre-school children", *Int.J.Obes.Relat Metab Disord.*, vol. 19,no. 1, pp. 6-10.

Drewett, R. F., Corbett, S. S., & Wright, C. M. 1999, "Cognitive and educational attainments at school age of children who failed to thrive in infancy: a population-based study", *J Child Psychol.Psychiatry*, vol. 40,no. 4, pp. 551-561.

Eck, L. H., Hackett-Renner, C., & Klesges, L. M. 1992, "Impact of diabetic status, dietary intake, physical activity, and smoking status on body mass index in NHANES II", *Am.J Clin.Nutr.*, vol. 56,no. 2, pp. 329-333.

Edmunds, L. & Waters, E. 2000, "Childhood obesity," in *Evidence-based pediatrics*, V. Moyer, ed.

Glenny, A. M., O'Meara, S., Melville, A., Sheldon, T. A., & Wilson, C. 1997, "The treatment and prevention of obesity: a systematic review of the literature", *International Journal of Obesity*, vol. 21, pp. 715-737.

Hall, D.M.B. 1996, *Health for All Children*, 3rd edn, Oxford University Press, New York.

- Hamill, P. V., Drizd, T. A., Johnson, C. L., Reed, R. B., Roche, A. F., & Moore, W. M. 1979, "Physical growth: National Centre for Health Statistics percentiles.", *American Journal of Clinical Nutrition*, vol. 32, no. 3, pp. 607-629.
- Lazarus, R., Wake, M., Hesketh, K., & Waters, E. 2000, "Change in body mass index in Australian primary school children, 1985-1997", *International Journal of Obesity*, vol. 24, pp. 679-684.
- Lusky, A., Barell, V., Lubin, F., Kaplan, G., Layani, V., Shohat, Z., Lev, B., & Wiener, M. 1996, "Relationship between morbidity and extreme values of body mass index in adolescents", *Int. J Epidemiol.*, vol. 25, no. 4, pp. 829-834.
- Maes, H. H., Neale, M. C., & Eaves, L. J. 1997, "Genetic and environmental factors in relative body weight and human adiposity", *Behav. Genet.*, vol. 27, no. 4, pp. 325-351.
- Mulligan, J. & Voss, L. D. 1999, "Identifying very fat and very thin children: test of criterion standards for screening test", *BMJ*, vol. 319, no. 7217, pp. 1103-1104.
- Panpanich, R. & Garner, P. 1999, "Growth monitoring in children", *The Cochrane Library (Oxford)*, vol. Issue 3 Oxford: Update Software.
- Power, C., Lake, J. K., & Cole, T. J. 1997, "Measurement and long-term health risks of child and adolescent fatness", *Int. J. Obes. Relat Metab Disord.*, vol. 21, no. 7, pp. 507-526.
- Raynor, P. & Rudolf, M. C. 2000, "Anthropometric indices of failure to thrive", *Arch Dis Child*, vol. 82, no. 5, pp. 364-365.
- Raynor, P., Rudolf, M. C., Cooper, K., Marchant, P., & Cottrell, D. 1999, "A randomised controlled trial of specialist health visitor intervention for failure to thrive", *Arch Dis Child*, vol. 80, no. 6, pp. 500-506.
- Robinson, T. N. 1999, "Reducing children's television viewing to prevent obesity: a randomized controlled trial", *Journal of the American Medical Association*, vol. 282, no. 16, pp. 1561-1567.
- Troiano, R. P. & Flegal, K. M. 1998, "Overweight child and adolescents: Description, Epidemiology, and Demographics", *Pediatrics*, vol. 101, pp. 497-504.
- Wake, M., Salmon, L., & Waters, E. 2000, "Health status of overweight/obese and underweight children: a population based study", Abstracts for the Royal Australasian College of Physician Annual Scientific Meeting A51, 33.
- Whitaker, R. C., Wright, J. A., Pepe, M. S., Seidel, K. D., & Dietz, W. H. 1997, "Predicting obesity in young adulthood from childhood and parental obesity", *New England Journal of Medicine*, vol. 337, no. 13, pp. 869-873.
- Willett, W. C., Dietz, W. H., & Colditz, G. A. 1999, "Guidelines for healthy weight", *N. Engl J Med*, vol. 341, no. 6, pp. 427-434.
- World Health Organization Division of Noncommunicable Diseases. 1998, *Obesity - preventing and managing the global epidemic. Report of a WHO Consultation on Obesity, Geneva June 1997*. WHO.
- Wright, C., Avery, A., Epstein, M., Birks, E., & Croft, D. 1998b, "New chart to evaluate weight faltering", *Arch Dis Child*, vol. 78, no. 1, pp. 40-43.
- Wright, C. M. 2000, "Identification and management of failure to thrive: a community perspective", *Arch Dis Child*, vol. 82, no. 1, pp. 5-9.
- Wright, C. M., Callum, J., Birks, E., & Jarvis, S. 1998a, "Effect of community based management in failure to thrive: randomised controlled trial", *British Medical Journal*, vol. 317, no. 7158, pp. 571-574.
- Wright, C. M., Matthews, J. N., Waterston, A., & Aynsley-Green, A. 1994, "What is a normal rate of weight gain in infancy?", *Acta Paediatr*, vol. 83, no. 4, pp. 351-356.
- Wright, C. M. & Talbot, E. 1996, "Screening for failure to thrive: What are we looking for?", *Child: Care, Health & Development*, vol. 22, no. 4, pp. 223-234.
- Wright, J. A., Ashenburg, C. A., & Whitaker, R. C. 1994, "Comparison of methods to categorize undernutrition in children", *J Pediatr*, vol. 124, no. 6, pp. 944-946.

7. LIMITATIONS OF THIS REVIEW

A formal systematic review of one single topic typically takes at least six months to carry out, and it is not possible to carry out such reviews of all the topics covered. What we aimed to do was report on all existing systematic reviews and then document the evidence as fully as possible for other topics, with the depth of each review determined by priority and accessibility of available evidence. Decisions limiting the depth of any review were clearly documented.

The topics covered in this document are those recommended by the project steering committee; there may be other topics or conditions that were beyond the scope of this review.

As stated in the Terms of Reference, development of Clinical Guidelines and a full consultative process was beyond the brief of this project.

8. SUMMARY OF FURTHER RESEARCH RECOMMENDATIONS

1. ATLANTO AXIAL INSTABILITY

- Natural history of atlanto-axial instability in Down Syndrome
 - Reduction in burden of disease through preventive management
 - Accuracy of radiograph interpretation
 - Harms (particularly quality of life) related to banning from particular activities
-

2. CONGENITAL ADRENAL HYPERPLASIA

- Optimal screening regime
 - Interpretation and cut off points for abnormal results
 - Overall population benefits of screening could be assessed by multi-centre randomised controlled trial, following the Wisconsin cystic fibrosis model.
-

3. CARDIAC DISEASE

- Evaluation and continuing quality improvement of programs
 - Methods to improve the sensitivity and specificity of the clinical examination in detecting important cardiac disease
-

4. CONGENITAL HYPOTHYROIDISM

- Optimum screening cutpoints to maximise sensitivity and specificity of blood spots taken in the first 48 hours of life
 - Randomised controlled trials using a broad range of outcome measures of :
 - medium vs high initial doses of thyroxine
 - use of T3 as a partial substitute for thyroxine in children
 - Quality of life of children and adults with congenital hypothyroidism
-

5. CYSTIC FIBROSIS

A recent Cochrane review states that long term follow up and analysis based on 'intention to treat' is required and note that this is under way in RCTs in the US and the UK. HTA reviewers call for re-analysis of existing research information.

Possible research areas:

- Evaluation of the consequences of detecting carriers as a result of neonatal screening
 - Utilisation of early detection in screened individuals to evaluate specific early treatments with RCTs.
 - Development and evaluation of audit procedure to ensure parents give informed consent for neonatal screening
 - Continuation of research into gene replacement therapy
 - Ongoing study of parent-reported outcomes, including impact of screening on reproductive decisions and quality of life
-

6. DEVELOPMENTAL DYSPLASIA OF THE HIP

- Evaluation and continuing quality improvement of programs
- monitoring of the incidence of surgery and late presentations of CDH through population based data collection. If there is no reduction in these outcomes following improvements to the screening program, the role of clinical screening should be further evaluated
- Further examination of the role of ultrasound screening in infants *at risk* of developmental dysplasia of the hip. Results from the UK RCT, when available, should inform this process
- Teanby et al recommend a prospective multicentre trial of untreated dysplasia diagnosed by ultrasound. This would allow treatment to be advocated for certain types of dysplasia if the trial confirms persistence or progression of dysplasia (Teanby & Paton 1997)

7a. CONDUCTIVE HEARING LOSS

- Long term effects of persistent conductive hearing loss associated with moderate or greater hearing loss
- Long term effects of persistent conductive hearing loss on other aspects of language such as auditory processing

7b. PERMANENT CHILDHOOD HEARING IMPAIRMENT

Many questions about infant hearing screening remain unanswered or partly answered. Possible research questions for Australia include:

- What are the effectiveness and efficiency of different models of early hearing screening at population level, weighed against short and long term harms?
- What proportion of disability can be prevented by early detection and management, and for whom?
- What aspects of early intervention and management are most crucial, and for whom?
- How do we most effectively deliver appropriate and ethical counselling and education to affected families?
- What is the role of genetic analyses in screening and management programs?
- Does early detection and management of mild and/or unilateral sensorineural hearing loss improve outcomes (ie in infancy/preschool years)?
- What is the prevalence of mild, unilateral and high-frequency hearing impairment in Australian school aged children, and what are the educational impacts?
- Are there opportunities in childhood and adolescence to prevent progressive hearing loss?

8. HYPERTENSION

- What is the association between childhood hypertension and adult morbidity?
- Can active management of childhood hypertension reduce adult morbidity and mortality?
- If screening for hypertension in children is found to reduce the burden of disease, can the reliability of the screening test be improved and what are the appropriate cutpoints?

9. IRON DEFICIENCY

- We recommend population-based surveys of Australian children approximately every five years. Surveys should use appropriate sampling strategies to access both low and high risk children, and should be designed to inform public health nutritional policy for young children.
- Public health studies into effective strategies to promote optimal dietary intake of iron among high risk and lower risk children. Outcomes should include (1) iron status and (2) developmental and behavioural outcomes.
- Research into simple parent-report diet questionnaires that might be used to reliably identify children with poor dietary patterns, who are likely to be at greater risk for iron deficiency.

10. LEAD TOXICITY

None recommended.

11. PHENYLKETONURIA

As per recommendations:

Continuation of current universal newborn screening programs for phenylketonuria is recommended.

Screening programs should incorporate population based monitoring and management programs for women of childbearing age, aiming to systematically prevent avoidable intellectual disability in their infants. Such monitoring should meet defined, stringent quality and reporting standards, and include reporting of long term outcomes for these children.

12. SCOLIOSIS

- Quantification of the proportion of children with scoliosis who will have significant progression of their spinal curves, and in which children significant progression will occur.
 - Randomised controlled trials to quantify the effectiveness of current therapies on an intention-to-treat basis.
-

13. UNDESCENDED TESTES

- Accuracy of the clinical examination as a screening tool
 - Assessment of effectiveness of treatment and the optimum age for treatment
 - The natural history of retractile and ascending testes
 - Evaluation and continuing quality improvement of programs
 - Effectiveness of neonatal screening programs
 - Role of parental participation in surveillance for undescended testes
-

14. URINALYSIS

None recommended.

15. VISION

- Assessment of the extent of disability and burden of disease attributed to amblyopia, and the benefits and harms of amblyopia treatment.
 - A clearer understanding of the relative risks associated with different levels of refractive abnormalities (and the potential use of different cutoff levels in screening) such as hypermetropia, astigmatism and myopia.
 - The potential role of family history and parental observation questionnaire instruments in targeted screening in the Australian context.
 - Understanding about the current level of 'usual care' in order to be able to evaluate the added effects of vision screening programs.
 - High quality evaluations of selected screening measures (such as the HOTV chart with confusion bars), and of the most appropriate age(s) for screening.
 - If such evidence suggested that vision screening were warranted, and that there were vision screening tests with adequate characteristics, then further controlled research (preferably by RCT) in community samples would be required.
-

16. DENTAL HEALTH

- Outcomes for treated vs untreated dental caries in the deciduous teeth
 - The effectiveness of preventive strategies, and how to maximise uptake of those that are effective
-

17. DEVELOPMENT

- Research to demonstrate the link between participation in developmental screening, with subsequent intervention where indicated, or not and subsequent development, educational and social outcomes of children.
 - Further research into better developmental screening tests is encouraged. However, this should take second priority to research at program level. Given the nature of child development it would appear unlikely that there will be significant improvements in test characteristics in the near future from those already achieved.
 - Research into screening programs in high risk or targeted groups.
 - Consideration of new models of early identification of developmental delay and disability, combined with intervention that is responsive to the child's needs. This might involve aspects of what is in some places called surveillance or monitoring, as well as the use of screening tests. There could be less reliance on screening tests; they would be used more as adjuncts when a course of action is not already clear.
-

18. LANGUAGE

- High quality cohort studies to better delineate the natural history of early language delay and its outcomes (language, social, academic) for at least the first seven years of life in Australian children. A major aim would be to study tools to reliably detect children with language delay that does not resolve and/or is associated with adverse outcomes. Until such tools exist, screening programs cannot be recommended.
 - Community based RCTs of early language intervention via parents for very young children (8-24 months) at risk of language delay.
-

19. HEIGHT

- Retrospective study in areas that have discontinued school entry height screening of secular trends in age of diagnosis of growth hormone deficiency and Turner syndrome, with and without screening
 - Properties of height measurement as a screening tool at various centile cutpoints for various ages (sensitivity, specificity, PPV, NPV, NNT)
 - RCT to determine whether school entry height screening is effective in:
 - a) lowering of age of diagnosis of growth hormone deficiency and Turner syndrome in school age children
 - b) benefits and harms, uptake of referrals, costs
 - Regular, systematic population-based surveys of height to monitor secular trends in height for the whole population and for subgroups at particular risk.
-

20. WEIGHT

- *Failure to thrive*: further research into how to accurately define and how to manage FTT should be a prerequisite to any screening activities.
- *Obesity*: regular, systematic population-based surveys of height, weight and BMI are recommended to monitor secular trends in overweight and obesity for the whole population and for subgroups at particular risk.

Prevention appears to be the most promising approach to childhood obesity. For established childhood obesity, further research into better definitions and effective management strategies should be a prerequisite to any screening activities. Research is urgently needed to define:

- a) which children are most at risk of obesity persisting into adulthood
 - b) which children are most at risk of adverse outcomes of obesity in adulthood
 - c) effective management strategies suitable for large scale use in the primary care setting
 - d) the acceptability, benefits, harms and uptake of such activities
-

9. THE LANGUAGE OF EARLY DETECTION: TIME FOR REVIEW?

It has long been accepted that efforts at prevention and early detection form a core component of health services. For some conditions it is possible to prevent any manifestations of what might otherwise be a serious or fatal condition. For others, early detection and early treatment can alter the natural course of the condition to prolong life expectancy and/or improve the quality of life. Early detection has also been considered to be desirable for conditions related to development and behaviour. There is a growing body of evidence documenting the effectiveness of early intervention for children at biological and/or environmental risk of poor or sub optimal developmental outcomes (Shonkoff & Hauser-Cram 1987).

In recent years, three loosely-related terms have commonly been used to describe these activities: screening, surveillance, and health promotion. In this chapter, we review the current status of these terms, and then make some suggestions to better focus these terms.

9.1. CURRENT TERMINOLOGY

9.1.1. SCREENING

Screening tests have been used in clinical practice and in population programs for many years, and the number of conditions screened for gradually expanded. A screening test is not intended to be diagnostic, but sorts out those who are likely to have the condition from those who likely do not.

As can be seen in the body of this report, there is clear evidence for the effectiveness of screening for a number of conditions, many of which are readily detectable in the neonatal period. There are well-established criteria against which to benchmark screening tests, suitability of conditions for screening, and overall effectiveness of screening programs – these are outlined in Chapter 4. Implicit in these criteria is that one either “has” or “does not have” a disease or condition, and that if one has, it can and will be treated effectively. This assumption underlies several of the most fundamental properties of screening programs, such as sensitivity, specificity, positive and negative predictive values, and number needed to treat. Screening programs that work well target conditions in which this distinction can be readily made for the great majority of children (eg PKU, congenital hypothyroidism, congenital adrenal hyperplasia). But what of conditions for which early detection and intervention may seem a useful goal, but for which no such distinction can readily be made - ie where there is considerable difficulty into categorising into 'has' or 'does not have' the condition?

Typical of this are the diverse problems that may occur in child development (including language) and behaviour. Screening for these problems is discussed in detail in Chapter 6.3.18. The terms “development” and “behaviour” artificially collapse constructs which are multidimensional along numerous axes. The concept of “having” or “not having” a problem in any one of these domains is mostly artificial and therefore problematic. Nonetheless, often driven by requirements of service systems to determine eligibility, generations of professionals have attempted to make such distinctions, and in turn to develop screening tests that are “valid” in terms of sensitivity, specificity, and positive predictive value. The extent to which this has failed can be appreciated in the commonly-held view that, for developmental screening tests, sensitivity of 70% and specificity of 70–80% is “acceptable” - values which would not be admissible when considering introduction of a new biochemical screen for a metabolic condition. Yet most developmental screening tests fail to reach even this low level. We are even less well off when it comes to distinguishing diagnosable behavioural problems from the normal and expected developmental problem behaviours. We should not blame the screening tests for this. While one can tinker with sensitivity at the expense of specificity, and vice versa, this begs the question: perhaps child development and behaviour are areas that are fundamentally unsuitable for screening, and an alternative approach is needed.

9.1.2. SURVEILLANCE

Given the issues and problems described above, it has been suggested that for child development and behaviour the concept of screening be replaced by surveillance (Hall 1996; Dworkin 1989).

There is no consensus on how surveillance is defined. For many it has a narrow definition, focusing on early detection of problems. For others it has been expanded to embrace many other activities. Hall noted that existing conceptualisations of surveillance could include “the oversight of the physical, social, and emotional health of children; measurement and recording of physical growth; monitoring of developmental progress; offering and arranging intervention when necessary; prevention of disease by immunisation and other means; and health education” (Hall 1996). However, he noted this concept to be problematic in its breadth.

Dworkin defined surveillance as “a flexible, continuous process that is broader in scope than screening, whereby knowledgeable professionals perform skilled observations on children throughout all encounters

during child health care (Dworkin 1989)". While screening consists of administering a test to a child, often as an activity divorced from routine care and excluding parental input, surveillance in this context actively involves parents by eliciting and responding to their concerns, putting these concerns into a historical and family context, and providing advice and information. Standard screening tests could be used as a component of surveillance, but were not the central focus (Glascoe & Dworkin 1993). Surveillance could be employed with individual children, but "also in the process of monitoring the health of a whole community of children" (Hall 1996).

By acknowledging and potentially addressing the biological and environmental factors that can influence child health and development, surveillance may seem to overcome many of the limitations of screening. However, its very breadth poses the danger of making both the term and the process vague and meaningless. Further, through multiple encounters rather than just one, its fundamental goal may still be similar to that of screening - to identify children who do or do not have developmental delay or a behaviour problem. What activities are included and what excluded? Do we include any or all activities that might conceivably improve children's health? Should we try to develop a more precise scientific definition with specified, clearly definable activities? How well do efforts at early detection sit with prevention activities and health education?

The term surveillance may have another different connotation for consumers. There is anecdotal information, difficult to quantify, that some parents perceive that the term surveillance somehow involves professionals evaluating and judging their competence as parents, or checking up on them. Such perceptions, even if wrongly conceived, nevertheless are of concern; if widely held, they may undermine population programs of surveillance in the epidemiological sense.

9.1.3. ACTIVITIES DESIGNED TO PREVENT PROBLEMS AND PROMOTE HEALTH

There are a number of activities, often subsumed under the umbrella of health surveillance, that are designed to either to prevent problems from occurring or to promote or enhance health outcomes. These include making sure that children are fully immunised, ensuring a safe environment to prevent injuries, reading to young children to promote early literacy, encouraging breastfeeding and subsequent good nutrition, discouraging parental smoking, anticipatory guidance, and providing parents with accurate information on the basis that well informed parents will be more likely to make appropriate decisions in relation to their children.

Some of these activities have a strong evidence base (eg immunisation, breast feeding); for others there are some limited data suggesting they are likely to be effective in facilitating improved outcomes (eg early literacy, injury prevention programs); some appear to be intuitive yet at present there is no compelling evidence as to their effectiveness in improving outcomes (anticipatory guidance, the provision of information to parents).

As for child health surveillance, there appears to be little consensus as to which preventive and health promotion activities should be systematically incorporated into child health programs and how their outcomes should be judged. To date, with the notable exception of immunisation, such activities have tended to be ad hoc and time-limited, and to occur outside a quality framework.

9.2. MOVE AWAY FROM THE CONCEPT OF PASS/FAIL

For many conditions in childhood where early support is likely to be of benefit – development, language, behaviour, family psychosocial issues – their very nature is such that there will never be a suitable screening test. These are complex, multi-dimensional areas that are not appropriate to categorise into pass/fail on the basis of a test; such a mutually exclusive categorisation flies across the face of the nature of children's development and of family functioning.

Development, language, behaviour, attention, family stress, quality of parenting and other variables exist on a continuum. At one end of the spectrum are those children and families where deficit, delay or dysfunction clearly exists (eg Down Syndrome or other chromosomal disorders), and where it is likely that this will be recognised by parents or professionals in the course of routine contact with a high-quality health and education system. At the other end of the spectrum are well-functioning children in well-functioning families without biological or environmental risk factors, and where good outcomes are likely without any specific intervention. Neither of these groups are likely to need a specific screening or identification program, though the former will certainly need appropriate assessment and management systems and facilities.

Many children and families exist toward the lower end of this continuum. They are not in the clearly abnormal range for any condition, yet various poor outcomes are frequent. Children in these groups, variously described as "borderline", "low-functioning", "high risk" or in the "gray" or "intermediate" zone, may have very real additional needs but can be virtually invisible to screening tests/programs that by

definition recognise only normal or abnormal. Yet by virtue of their position on the normal curve they are usually more numerous than those who have a clearly-defined disorder. While risk and protective factors may be fluid, some will have a combination of biological and environmental risk which is relatively stable and potentially leads to poorer outcomes – the concept of “double jeopardy” (Escalona 1982; Parker, Greer & Zuckerman 1988).

Many of these children and families could benefit from some sort of support or intervention. For example, individual maturational differences are the most likely reason that a two year old might have limited language. If there is no intervention, two thirds of these children will have language which is considered “normal” or “within normal limits” by age four, although half of these will continue to carry subtle evidence of their earlier difficulties. Early language support might benefit the child who would otherwise go on to have true language impairment. It might also benefit the children who “recover”, not only in terms of ultimate language functioning but also in decreasing the chances of subsequent learning difficulties. Similar scenarios can be constructed for other domains such as behaviour, where a child may not fit into a diagnostic category but where intervention at a young age may be beneficial in terms of improved short and long-term outcomes. Additional support may also allow additional scrutiny of progress, and may facilitate earlier detection of and intervention for a “true” condition or disability.

9.3. CLARIFICATION OF TERMINOLOGY

We would suggest greater precision in the way these terms are used.

9.3.1. SCREENING

“Screening” should be confined to those conditions and tests that meet criteria for screening and for which the benefits of screening have been clearly demonstrated. It should be noted that most of these conditions and tests occur in the neonatal period, and are likely to be implemented exclusively by health professionals in health care settings. These are situations where pass/fail criteria are appropriate, where it is almost always clear what further action needs to be taken on the basis of a failed screening test, and where management instituted as a result of screening is expected to substantially alter outcomes for that child.

Even though there is strong evidence for the effectiveness of screening with these conditions, it is still important to establish quality processes for the tests themselves, their interpretation, the expertise of the testers, and the program as a whole including systems and standards for referral, tracking, management, support and reporting. Many clinical conditions currently screened for do not approach ideal system standards (eg congenital dislocation of the hip, undescended testes, ophthalmic screening).

9.3.2. SURVEILLANCE

We would argue for a more restrictive definition of surveillance, linked to its original intention of early detection of specific problems.

Surveillance should refer to a set of activities designed for the early detection of clearly-defined and specific problems which, unlike those appropriate for one-off screening, would not be expected to be reliably detected at a single point because they may develop or fluctuate over time. Essentially, surveillance would equate to periodic screening over time; “failing” the screen one or more times would indicate a high likelihood of having a disorder. Unlike adult periodic screening programs, such as for breast and cervical cancer, the instruments used might vary at different ages to be appropriate to the child’s age and/or developmental stage.

Such conditions might include short stature, overweight/obesity, hypertension, and vision and hearing problems in the school years, if the evidence were there to support these activities. We would also argue that, like screening, surveillance activities should adhere to the evidence-based principle that each such activity should lead to more benefit than harm. Generally research in this area of child health is almost non-existent. Thus, few if any of these activities can currently be recommended on the existing evidence; this does not mean that detection of these problems is not important.

9.3.3. ACTIVITIES DESIGNED TO PREVENT PROBLEMS AND IMPROVE OR PROMOTE HEALTH

These activities should include a broad range of issues, for which different approaches may be needed. However, these activities are likely to be implemented in a range of settings by many of the professionals who come into contact with children and their families, not just in the health care system. They may be divided into universal and targeted activities.

Universal activities, directed at all children, may:

- promote healthy lifestyles, for example food choices, physical activity, safety in the home, and immunisation.

- directly enhance health & well-being, for example promotion of breastfeeding, early language/literacy, anticipatory guidance for behaviour and sleep management, and immunisation.

In contrast, targeted activities are directed at children “at risk” of specific adverse outcomes. Risk may be determined by activities such as monitoring BMI, eliciting parent concerns, and determining psychosocial or sociodemographic risk factors. While the process of determining risk may for some issues resemble screening, it differs in that it aims to be inclusive rather than exclusive, the response may be flexible or graded, and that services may be offered on the basis of risk alone – a diagnosable disorder does not need to be identified (although ideally the “at risk” group would contain all or almost all of those with true disorder). Importantly, the “borderline” or “grey” areas become issues worthy of consideration in their own right. Typically, far more children fall into these borderline zones than have a diagnosable disorder; thus the potential for health benefit at a population level from secondary prevention activities delivered in the primary care or community setting is large.

Issues for which promising outcomes have already been demonstrated in the community include language enhancement programs for those with very early communication delay, high-quality day care for children at psychosocial and cognitive risk, and simple infant sleep management programs. Others urgently require research, for example secondary prevention of overweight.

This flexible, longitudinal process would seem particularly suited to areas such as child development or to psychosocial and other risk factors that exist on a continuum and which are neither easily nor desirably categorised into pass/fail. It could consist of any activities that lead to identification of risk – eliciting parent concerns, physical examination, informal observations, obtaining information from other sources, measurement of growth, administration of tests and procedures, referral for further assessment. These activities may be initiated by professionals but involve a partnership with parents. While they are a core component of primary health care, they may also take place in non-health care settings – child care centres and family day care schemes, preschools and kindergartens, schools – by all professionals who come into contact with young children and their families.

These activities should be continued on an individual and population basis, and further research encouraged to demonstrate efficacy and effectiveness of various strategies. We would argue that prevention and health promotion should be regarded as an integral component of a high quality primary care health system, and not separate from service delivery. Like screening, they should be conducted within an evidence based framework; where evidence of overall benefit does not exist, new programs should not be implemented outside the research context.

9.4. PRINCIPLES AND PROCESSES FOR SERVICE DELIVERY

Prevention, early detection and health promotion activities are inexorably linked with the delivery of services to children and their families. The three-tier system suggested – screening, surveillance, and prevention/promotion – should be considered a core component of service delivery. All children and all families would benefit from access to a universal platform of primary services – health, childcare, and education - where these activities are adopted. We suggest that consideration be given to the following principles and procedures to provide an optimal context for the detection of problems at an early stage, as well as providing interventions and support to children and families who are likely to benefit from it.

Early detection of problems, early intervention and health promotion activities should not be confined to the health sector. One of the advantages of an accessible and universal service system is that it provides numerous opportunities over a period of time to detect emerging or established problems and to implement activities designed to improve outcomes.

All professionals working with children and parents should engage the parents actively in discussions regarding their children. Formal instruments such as the Parents’ Evaluation of Developmental Status (Glascoe 1997) can be used for this purpose.

All parents are likely to benefit from being provided with credible, age-appropriate and culturally relevant information about their child’s health, development and behaviour. Where parents have voiced concerns, these are addressed in a timely and appropriate manner taking into account the context. Context factors include the child (medical and developmental history, biological risk, age), family and environment (environmental risk), the setting (childcare, community nursing, general practitioner, preschool, school) and the provider (level of training expertise).

Most children and families will benefit from some level of support – this will range from the provision of advice to parents through to more intensive intervention targeted to children with moderate to severe problems. The level and type of support/enrichment/ intervention again will be dependent on the child, the parents (level of knowledge, confidence, support available, level of concern), and available services.

Risk and protective factors are dynamic and their balance will likely change over time. There is variability with respect to these factors for different children, families, and in different age periods. The specific balance of risk and protective factors is complex and rarely fixed. The higher the number of risk factors, the more likely there is to be a poorer outcome if there is no additional support or intervention. A combination of biological and environmental risk factors is especially problematic. (Escalona 1982; Parker, Greer & Zuckerman 1988).

For development and behaviour, we recommend a conceptual move away from tests and from categorising children into pass/fail groups, towards the concept that most children and families would benefit from ongoing contact with a universal system that is responsive to their needs. Ideally parent concerns and risk factors would be systematically elicited and addressed; a range of graded interventions offered in context; longitudinal follow-up would occur to take into account the changing nature of development and risk and protective factors; seamless referral and follow-up systems would be put into place in community networks; and the whole system would be underpinned by a system of quality assurance to ensure that structures and processes are consistent with contemporary knowledge.

REFERENCES

- Dworkin, P. H. 1989, "British and American recommendations for developmental monitoring: The role of surveillance", *Pediatrics*, vol. 84, no. 6, pp. 1000-1010.
- Escalona, S. K. 1982, "Babies at double hazard: Early development of infants at biologic and social risk", *Pediatrics*, vol. 70, pp. 670-676.
- Glascoc, F. P. 1991, "Can clinical judgement detect children with speech-language problems? ", *Pediatrics*, vol. 87, pp. 317-322.
- Glascoc, F. P. 1996, "The importance of discussing parents' concerns about development", *Ambulatory Child Health*, vol. 2, pp. 349-356.
- Glascoc, F. P. 1997, "Parents' concerns about children's development: Prescreening technique or screening test?", *Pediatrics*, vol. 99, pp. 522-528.
- Glascoc, F. P., Altemeier, W. K., MacLean, W. E. 1989, "The importance of parents' concerns about their child's development", *Am J Dis Child*, vol. 143, pp. 855-858.
- Glascoc, F. P., Dworkin, P. H. 1993, "Obstacles to effective developmental surveillance: Errors in clinical reasoning", *J Dev Behav Pediatr*, vol. 14, no. 5, pp. 344-349.
- Glascoc, F. P., Dworkin, P. H. 1995, "The role of parents in the detection of developmental and behavioural problems", *Pediatrics*, vol. 95, pp. 829-836.
- Glascoc, F. P., MacLean, W. E., Stone, W. L. 1991, "The importance of parents' concerns about their child's behaviour", *Clin Pediatr*, vol. 30, pp. 8-11.
- Hall, D. M. B. 1996, *Health For All Children (3rd Edition)*, Oxford University Press, Oxford.
- Parker, S, Greer, S, Zuckerman, B. 1988, "Double Jeopardy: The impact of poverty on child development", *Pediatr Clin North Am*, vol. 35, pp. 1227-1240.
- Shonkoff, J. P, Hauser-Cram, P. 1987, "Early intervention for disabled infants and their families: A quantitative analysis", *Pediatrics*, vol. 80, pp. 650-658.
- Starfield, B, Borkworf, S. 1969, "Physicians' recognition of complaints made by parents about their children's health", *Pediatrics*, vol. 43, pp. 166-172.

10. GLOSSARY

Effectiveness

Effectiveness is the outcome of a specific intervention under normal, everyday conditions. Muir Gray (1997) puts this in the context of a screening program:

‘The effectiveness of any screening program is determined by:

- The sensitivity of the series of tests applied to the population
- The effectiveness of the therapy offered to those individuals discovered to have the condition

Thus screening effectiveness = test accuracy + therapeutic effectiveness’

Efficacy

Efficacy is the outcome of a specific intervention under ideal conditions.

Efficiency

Efficiency is a measure of the economy with which an intervention of known efficacy and effectiveness is carried out. This is frequently used synonymously with cost-effectiveness, which in most cases is appropriate. As well as the actual costs of a screening program, a measure of efficiency or cost-effectiveness must consider the costs of any potential harms of the intervention versus the benefits, and the opportunity costs of other interventions that are foregone in favour of the program in question.

False negative

An individual incorrectly identified by a negative test result who does actually have the condition.

False positive

An individual incorrectly identified by a positive test result who does not actually have the condition.

Negative Predictive Value (NPV)

The negative predictive value of a test is the proportion of people with a negative test result who are actually free of the disease. It is the probability that a person with a negative test is a true negative.

Number needed to test

The number of individuals in the target population that must be tested in order to find one case of the condition being screened for.

Positive Predictive Value (PPV)

The positive predictive value of a test is the proportion of people with a positive test result who actually have the disease. It is the probability that a person with a positive test is a true positive.

The positive predictive value is related to the prevalence of the condition being screened for. If the condition has a high prevalence in the population, a test with good sensitivity and specificity will have a high PPV. If the condition has a low prevalence, a test will have a low PPV.

Sensitivity

The sensitivity of a test is the proportion of people with the condition who are correctly identified by a positive test result. It is a measure of the true positive rate.

Specificity

The specificity of a test is the proportion of people who do not have the condition who are correctly identified by a negative test result. It is a measure of the true negative rate.

Yield

The yield of a screening program is the number of new cases detected for every 100 cases screened.

The Incremental yield is the number of new cases that would not have been detected without the screening program.

11. APPENDICES

A. REFERENCE GROUP

Professor David Hall, Professor of Paediatrics, University of Sheffield
Sheffield Children's Hospital, Sheffield S10 2TH, UK
Ph: 0114 271 7344; Fax: 0114 275 5364

Professor Paul Dworkin, Professor of Paediatrics, St Francis Medical Centre
115 Woodland Street, Hartford CT 06095 USA
Ph: 203 714 5020; Fax: 203 714 8054

Professor Frances Page Glascoe, Child Development Centre, Vanderbilt University
2100 Pierce Avenue, Room 426MCS, Nashville Tennessee 37232-3572 USA
Ph: 615 936 0249; Fax: 615 269 6967

Dr Garth Alperstein, Community Paediatrician, Central Sydney Community Health Services
Queen Mary Building – Level 6, Camperdown NSW 2050
Ph: 02 9515 3270; Fax: 02 9515 3282

Dr Katrina Williams, Staff Specialist, Clinical Epidemiology Unit
The Children's Hospital at Westmead, Locked Bag 4001, Westmead NSW 2145
Ph: 02 9845 2006; Fax: 02 9845 3389

Professor Colin Morley, Professor/Director of Neonatal Medicine
Neonatal Services Division, Royal Women's Hospital, 132 Grattan Street, Carlton VIC 3053
Ph: 03 9344 2524; Fax: 03 9347 2731

A/Professor Gay Edgecombe, Chair of Community Health Nursing
RMIT, GPO Box 2476V, Melbourne VIC 3000
Ph: 03 9660 3825; Fax: 03 9660 3066

Dr Neil Wigg, Director, Community Child Health Services;
Chair of the Chapter of Community Child Health, Division of Paediatrics
PO Box 1507, Fortitude Valley QLD 4006
Ph: 07 3250 8555; Fax: 07 3252 7895

Professor Chris del Mar, Professor of General Practice; Director of the Centre for General Practice
University of Queensland, Herston QLD 4006
Ph: 07 3365 5379; Fax: 07 3365 5130

Ms Carolyn Briggs, Senior Lecturer, Faculty of Nursing, Midwifery & Health; Uni of Technology, Sydney
President, Australian Association of Maternal, Child and Family Health Nurses (Inc)
Ph: 02 9514 5136

Professor George Patton, Director, Centre of Adolescent Health, University of Melbourne
Royal Children's Hospital, Flemington Road, Parkville VIC 3052

Dr Lyndall Whitecross (replacing Dr Gerard Connors), Chair of the Preventive Health Subcommittee,
RACGP
Bluff Road Medical Centre, Sandringham VIC 3191
Ph: 03 9589 6244

Dr John Aloizos, Chair, Australian Divisions of General Practice
83 Davrood Street, Robertson QLD 4109
Ph: 07 3274 1886; Fax: 07 3274 1930

Professor Bob Williamson, Director, Murdoch Children's Research Institute
Royal Children's Hospital, Flemington Road, Parkville VIC 3052

Ms Meredith Carter, Consumer Representative, Health Issues Centre
Level 11, 300 Flinders Street, Melbourne VIC 3000
Ph: 03 6914 0500; Fax: 03 6914 0511

B. SEARCH STRATEGIES

Search for evidence moving down through the levels of evidence

Finding existing systematic reviews

1. Search Cochrane database of systematic reviews
2. Search “Best Evidence”
3. Search “DARE”
4. Search Medline, HealthStar, CINAHL and ClinPSYC using “Systematic Review” filter

If systematic review found: search for literature published after the review

1. Search Cochrane database of RCTs
2. Search “Best Evidence”
3. Search “DARE”
4. Search Medline, HealthStar, CINAHL & ClinPSYC

If no systematic review found: search for randomised controlled trials

5. Search Cochrane database of RCTs
6. Search “Best Evidence”
7. Search “DARE”
8. Search Medline, HealthStar, CINAHL & ClinPSYC using “RCT” limit option

If no randomised controlled trials found: search for literature reviews, well designed studies (eg cohort studies) and articles reporting the accuracy of screening tests

1. Search Medline
2. Search HealthStar
3. Search CINAHL
4. Search ClinPSYC

C. WEBSITES

Key websites

Organisation	Address
ACP Journal Club	http://www.acponline.org/journals/acpjic/pastiss.htm
Adolescent Health Online (AMA)	http://www.ama-assn.org/adolhth/gapslink/gapslink.htm
Agency for Health Care Policy and Research (AHCPR)	http://www.ahcpr.gov/
American Academy of Child & Adolescent Psychiatry	http://www.aacap.org/web/aacap/
American Academy of Pediatrics	http://www.aap.org/
Bandolier EBM Annuals	http://www.jr2.ox.ac.uk/Bandolier/
BioMedNet search	http://www.biomednet.com/
Cambridge Public Health	http://fester.his.path.cam.ac.uk/phealth/phweb.html
Canadian Institute of Child Health	http://www.cich.ca/html/menu.htm
Canadian Medical Association	http://www.cma.ca/
Canadian Medical Association Journal	http://www.cma.ca/cmaj/index.htm
Canadian Public Health Association	http://www.cpha.ca/
Canadian Task Force on Preventative Health Care	http://www.ctfphc.org/index.htm
CDC Wonder	http://wonder.cdc.gov/
Centre for Evidence-based Child Health	http://www.ich.bpmf.ac.uk/ebm/ebm.htm
Centre for International Child Health	http://cich.ich.ucl.ac.uk/
Children's House Health Floor (WHO)	http://childhouse.uio.no/health/index.html
CliniWeb	http://www.ohsu.edu/clinweb/search.html
Cochrane Collaboration	http://www.cochrane.org/ http://hiru.mcmaster.ca/cochrane/cochrane/cc-broch.htm
Cochrane Library	http://www.hcn.net.au
Abstracts of Cochrane Reviews	http://som.flinders.edu.au/fusa/cochrane/revabstr/mainindex.htm
Core library for evidence based practice	http://www.shf.ac.uk/uni/academic/R-Z/scharr/tr/core.html
Current controlled trials	http://www.controlled-trials.com
Effective Healthcare Bulletins	http://www.york.ac.uk/inst/crd/ehcb.htm
Effectiveness Matters	http://www.york.ac.uk/inst/crd/em/html
Evidence Based Medicine	http://www.bmjpg.com/data/ebm.htm
Evidence Based Mental Health	http://cebmh.warne.ox.ac.uk/cebmh/journal/
Evidence Based Nursing	http://www.bmjpg.com/data/ebn.htm
Global ChildNet	http://ediecpst.sfu.ca/qcnet/index.html
Guide to Clinical Preventive Services (Report of US Preventive Services Task Force)	http://text.nlm.nih.gov/frs/pick?collect=cps&dbName=0&cc=1&t=930199919
Health Canada	http://www.hc-sc.gc.ca/english/
Health Evidence Bulletins – Wales	http://www.uwcm.ac.uk/uwcm/b/pep/
Health Services Research Unit, Oxford University	http://hsru.dphpc.ox.ac.uk/
Institute for Child Health Policy	http://www.ichp.ufl.edu/
International Child Health: a digest of current info.	http://www.urmc.rochester.edu/IPA/pubs/inch.htm
Internet Database of Evidence-based Articles (IDEA) topic list	http://www.ohsu.edu/bicc-informatics/ebm/ebm-topics.htm
MedFinder Smart Medical Web Search	http://www.netmedicine.com/medfinder.htm
Medline	http://www.medline.cos.com
National Library of Medicine (multiple database search) (HealthStar)	http://www.nlm.nih.gov http://igm-01.nlm.nih.gov/cgi-bin/doler?account=++&password=++&datafile=healthstar
NHS Centre for Evidence-Based Medicine	http://cebmr2.ox.ac.uk
CatBank	http://163.1.212.5/cats/allcats.html
NHS Centre for Reviews and Dissemination	http://www.york.ac.uk/inst/crd/
NHS CRD incl: DARE (Database of Abstracts & Reviews of Effectiveness)	http://nhscrd.york.ac.uk/
NEED (NHS Economic Evaluation database)	
HTA (Health Technology Assessment database)	
NHS Health Technology Assessment Programme	http://www.hta.nhsweb.nhs.uk/ or http://www.soton.ac.uk/~hta/
North Thames Research Appraisal Group	http://cebmr2ox.ac.uk/ntrag/ntrag.html
OMNI database	http://omni.ac.uk
Paediatric Points of Interest	http://www.med.jhu.edu/peds/neonatology/poi.html
PEDBASE: Pediatric Database	http://www.icondata.com/health/pedbase/index.htm
Pediatric Internet Directory (wwwSLACK)	http://www.slackinc.com/child/pednet-x.htm

Pediatrics (Electronic Pages)	http://intl.pediatrics.org/
PedInfo	http://www.uab.edu/pedinfo/
SchARR-Lock's Guide to the Evidence	http://www.shef.ac.uk/uni/academic/R-Z/scharr/r/scebm.html
Systematic Reviews Training Unit	http://www.ich.ucl.ac.uk/srtu/frames.htm
The best of the pediatric internet (AAP)	http://www.aap.org/bpi/default.htm
The <i>Informed</i> Newsletter	http://www.ices.on.ca/docs/informed.htm
TRIP database	http://www.qwent.nhs.gov.uk/trip/test-search.htm
ATTRACT	http://www.qwent.nhs.gov.uk/trip/attract.html
Unit for Evidence Based Practice and Policy	http://www.ucl.ac.uk/primcare-popsci/uebpp/uebpp.htm#How
Virtual Children's Hospital	http://vch.vh.org/
WHO International Child Health & Development	http://www.who.int/chd/
World Health Organisation	http://www.who.int/
WHO publications	http://www.who.int/dsa/cat97/ztrs.htm
Clinical Practice Guidelines	
Agency for Health Care Policy and Research (AHCPR) guidelines	http://www.ahcpr.gov/clinic/
Centre for Evidence-based Mental Health guidelines	http://cebmh.warne.ox.ac.uk/cebmh/guidelines/
Clinical Practice Guideline Catalogue (AMA)	http://www.amda.ab.ca/cpg/index.html
Clinical Practice Guidelines – Ottawa General Hospital	http://www.ogh.on.ca/library/cpg.htm
Guideline appraisal project (GAP)	http://www.inforward.uaberta.ca/cpg/
National Guidelines Clearing House	http://www.guideline.gov/index.asp
National Guidelines Clearing House II	http://www.ahcpr.gov/
National Library of Medicine – Health Services/Technology Assessment Text (HSTAT)	http://text.nlm.nih.gov/frs/gateway
Virtual Hospital Clinical Guidelines	http://www.vh.org/Providers/ClinGuide/CGType.html

Other useful websites

Organisation	Address
Aggressive Research Intelligence Facility (ARIF)	http://www.hsrc.org.uk/links/arif/
Centre for Clinical Effectiveness	http://www.med.monash.edu.au/publichealth/cce/
Centres for Disease Control and Prevention	http://www.cdc.com
Cochrane Neonatal Collaborative Review Group	http://silk.nih.gov/SILK/COCHRANE/Cochrane.htm
Neonatology on the web	http://www.neonatology.org/
Neonatology Teaching Files and Guidelines	http://www.neonatology.org/syllabus/syllabus.html