Newborn screening for inborn errors of metabolism: principles, policies and weighing the evidence

J V Leonard

C Dezateux

Abstract

Newborn screening for metabolic disorders has become a contentious issue. The aim of screening is to identify individuals at risk and start treatment before they become ill. To this end newborn screening programmes are well established in many countries and recent technological developments have lead to an expansion of these programmes. These require careful evaluation, both of the process and the outcome. The original Wilson and Jungner criteria for evaluation are still valid but, in this review, three main points are particularly considered. The burden and the natural history of the disease need to be defined. The test should predict accurately those who would develop clinical disease but current screening programmes detect many with 'mild' disease, the importance of which is often unclear. This is particularly relevant when assessing any improvement in outcome which should be seen in terms of the advantages and problems for both the individual and the family.

Keywords genetic testing; healthcare evaluation mechanisms; health policy; high-throughput screening assays; infant, newborn; mass screening; metabolism, inborn errors; Neonatal screening; public health practice; tandem mass spectrometry

Introduction

Newborn screening for inborn errors of metabolism is now well established in developed countries worldwide. As noted by Wilson and Jungner in their 1968 WHO report on the Principles and Practice of Screening for Disease (Box 1), screening seems both intuitive and attractive: it aims to detect and manage serious diseases in order to secure an outcome better than that which might be achieved following clinical presentation or diagnosis.

The recognition that presymptomatic diagnosis and treatment could profoundly alter the outcome for phenylketonuria (PKU) drove Guthrie to develop a screening test based on dried blood spots that was cheap and feasible for mass screening of newborn infants. Screening for PKU was first introduced in the 1960s and was followed, in the 1970s, by screening for congenital hypothyroidism. Newborn screening programmes for these two disorders are now almost universal in high- and middle-income countries throughout the world. While regarded as examples of effective preventive medicine, almost half a century later experience with newborn screening for these conditions exemplifies some of the difficulties in establishing that all those identified and treated as a consequence of screening do in fact need treatment, and the nature of the benefit conferred.

> "The central idea of early disease detection and treatment is essentially simple. However the path to its successful achievement (on the one hand, bringing to treatment those with previously undetected disease and, on the other, avoiding harm to those persons not in need of treatment) is far from simple though sometimes it may appear deceptively easy."
>
> *Wilson and Jungner, 1968, WHO*

Box 1

More recently there has been a marked expansion of newborn screening programmes, driven by the development of technologies adaptable for high through-put analyses of biomarkers in newborn dried blood spots, notably tandem mass spectrometry (MS-MS). It is likely that future expansion will be driven by the development of new treatments for rare disorders or by new approaches to identify risk for or susceptibility to more complex or chronic diseases, as much as by new technologies. Currently parents of newborns in many countries are now offered testing for more than 30 disorders, many of them very rare. However, not all countries have implemented 'expanded' newborn screening on this scale, reflecting different screening policies and approaches to their evaluation.

In this contribution, we review the criteria by which proposed screening programmes are assessed and discuss aspects that are specific to the assessment of newborn screening for rare conditions such as inborn errors of metabolism. We highlight some of the challenges in obtaining and evaluating the evidence needed to inform screening policies for these conditions. Detailed information about screening for specific disorders is covered by other contributors to this mini-symposium.

Definition of screening

Wald defined screening as the '*systematic application of a test or enquiry to identify individuals at sufficient risk of a specific disorder to warrant further investigation or direct preventive action, amongst persons who have not sought medical attention on account of symptoms of that disorder.*' [Wald N. Guidance on terminology. *J Med Screening* 1994; **1**(1): 76].

While the rationale for screening is driven primarily by concern to improve outcome for affected individuals, in this definition, Wald reminds us that all those offered screening do not usually have any concerns or symptoms related to that condition that has so far prompted them to seek medical care. In doing so, he highlights an implicit and ethical imperative to do no harm to those screened. The implication is that the potential benefits of screening should be positively balanced in relation to potential harms. This requires a judgement based on a range of complex and often imperfect information.

J V Leonard *PhD FRCP FRCPCH is Professor Emeritus at UCL Institute of Child Health, 30 Guilford Street, London WC1 1EH, UK.*

C Dezateux *MD FMedSci is Professor of Paediatric Epidemiology and Director of MRC Centre of Epidemiology for Child Health, London, UK.*

Approaches to evaluating screening programmes

In their original WHO publication, Wilson and Jungner distinguished the evaluation of screening procedures from the evaluation of effects of screening (namely reduced morbidity and mortality). Newborn screening as a process, not just a test: the various stages in this process require careful assessment to ensure that the perceived advantages are genuine and outweigh any potential harm. While most would agree that there need to be demonstrable benefits to screening, views vary regarding the types of benefits to be considered, the weighting given to those benefits, and the evidence of benefit which is needed before screening policy can be made and programmes implemented. For example, some argue that early diagnosis *per se* is a legitimate goal of screening, irrespective of evidence of improved health outcomes.

Wilson and Jungner were the first to outline a broad and systematic approach to evaluation and in their original report identified 10 major criteria which needed to be addressed (see Box 2). In the United Kingdom these criteria have been extended into a framework comprising 22 criteria which are used for the evaluation of all screening programmes by a National Screening Committee. In certain countries, specific policies have been published about newborn blood spot screening, for example, by the Human Genetics Society of Australasia or the American College of Medical Genetics. All these frameworks have common elements and there is broad consensus that decisions should be informed by the evaluation of scientific evidence. In this article we have identified three main areas for evaluation, summarized as (1) the burden of the disease for which screening is being offered; (2) the clinical validity of the screening test and (3) the clinical utility of the screening programme. Below we discuss these in the context of newborn screening for inborn errors of metabolism and highlight some specific challenges in programmes for rare diseases. We conclude by considering some additional issues related to rare diseases, drawing on the UK National Institute for Health and Clinical Excellence (NICE) guidance on the principles to be used when applying social value judgements to policy evaluations and guidance.

The Wilson and Jungner criteria for evaluating screening programmes

1 The condition sought should be an important health problem.
2 There should be an accepted treatment for patients with recognized disease.
3 Facilities for diagnosis and treatment should be available.
4 There should be a recognizable latent or early symptomatic stage.
5 There should be a suitable test or examination.
6 The test should be acceptable to the population.
7 The natural history of the condition, including development from latent to declared disease, should be adequately understood.
8 There should be an agreed policy on whom to treat as patients.
9 The total 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.
10 Case finding should be a continuous process, not a "once and for all" project.

Wilson and Jungner, 1968, WHO

Box 2

Epidemiological considerations: the burden of disease

The importance of a condition for which screening is offered relates not just to its frequency but also to its consequences. Wilson and Jungner noted that 'phenylketonuria is extremely uncommon but warrants screening on account of the very serious consequences if not discovered and treated very early in life.'

A rare disease has been defined as a condition which affects less than five people in 10,000: by this definition almost all inborn errors of metabolism are rare. Many are, in fact, very rare which pose problems in acquiring reliable and unbiased information about their frequency, natural history, clinical outcome and the effects of treatment. In the UK all paediatricians contribute to the surveillance of rare diseases through a monthly active reporting scheme run by the British Paediatric Surveillance Unit. This has proved a valuable mechanism for studying the epidemiology and early clinical course of a wide range of candidate conditions for newborn screening, including galactosaemia, medium chain acyl CoA dehydrogenase deficiency (MCADD), glutaric aciduria type 1 and congenital adrenal hyperplasia. This scheme can also be used to obtain useful information about the burden of clinically presenting and diagnosed disease and importantly helps to identify whether there is a window of opportunity for screening to make a difference, referred to by Wilson and Jungner as a latent or early symptomatic phase which allows time for diagnosis and initiation of definitive treatment and management.

While the natural history of those with severe disease may be well documented, the course of those with 'mild' disease may not: such individuals may present infrequently to a clinician or may not be ascertained at all clinically but are nevertheless identified by screening. The problems associated with the wide range of the clinical phenotype are discussed later.

Surveillance studies of rare diseases may also provide information about their geographical variation, but this is not usually helpful in determining whether screening should be offered to geographically defined populations. It can in practice be difficult to determine whether an apparent geographical cluster of a rare disease ascertained by active surveillance relates to a true difference in its frequency, to differential ascertainment or reporting, or to availability of specialist services and referral patterns. While targeting higher risk populations that may be geographically isolated or separated by custom or religion is an attractive proposition, this presupposes a robust strategy for selecting those at higher risk. For example, although tyrosinaemia type 1 was recognized to be more prevalent in one area of Quebec — Saguenay-Lac St Jean — in practice screening is offered throughout the province. Similarly where the risk of a rare disease is higher in certain ethnic groups it may appear attractive to consider using ethnicity as a basis for offering screening. However the difficulties of ascertaining ethnic origin in contemporary populations with high rates of migration and inter-ethnic union make such selection unreliable, as has been demonstrated by the progressive abandonment of 'selective' newborn screening strategies for sickle cell disorders in the United States.