An international review of the Cost-Effectiveness of expanded Neonatal Bloodspot Screening and its implications for the Netherlands

(heel)Pricking Babies around the Globe

Master Thesis Health Sciences/Health Care Management

L.J.B. Leewis

An international review of the Cost-Effectiveness of expanded Neonatal Bloodspot Screening and its implications for the Netherlands

(heel)Pricking Babies around the Globe

Author:
L.J.B. Leewis
s0074659
Master thesis Health Sciences
Faculty of Management and Governance
Twente University
l.j.b.leewis@student.utwente.nl / loes.leewis@rivm.nl

Supervision:

Prof. dr. M.J. IJzerman	(University of Twente)
Dr. M.M. Boere-Boonekamp	(University of Twente)
Dr. M. van den Berg	(RIVM-VTV)

Source coverpicture: *http://www.metaboleziekten.nl/ziekten/pku/images/hielprik.jpg*

Abstract

Introduction

In 2008 the Center for Population Screening (CvB) made an inventory of Dutch cost-effectiveness evaluations for all its screening programs. This inventory showed that so far, no evaluation had been conducted to determine the cost-effectiveness of the heel prick program. In 2007 the Dutch neonatal heel prick screening program was expanded from 3 to 17 conditions and therefore received priority for determining its cost-effectiveness. This report is meant to show what is known about the cost-effectiveness of the neonatal heel prick program through cost-effectiveness evaluations done abroad.

Method

To gain information on international cost-effectiveness evaluations a systematic literature review was conducted. To ensure that no evaluations would be missed (e.g. non published) national screening agencies were contacted with an information request. The resulting evaluations were subjected to transferability criteria and quality guidelines to determine whether their information could be relevant for the Netherlands.

Results

The literature review showed that 20 cost-effectiveness evaluations had been done on neonatal heel prick screening for at least one of the conditions screened for in the Netherlands. Out of these evaluations, 7 concerned tandem-mass-spectrometry (MS/MS) screening for an expanded panel of conditions, with the remaining 11 evaluations concerning one or two conditions. These 11 evaluations are therefore not directly relevant for providing information on the cost-effectiveness of the *expanded* Dutch program. Out of the 7 evaluations on expanded programs, only 4 evaluations met the basic transferability criteria and (most) quality guidelines.

Conclusion

Almost all evaluations, no matter the number or type of conditions screened for, yielded incremental cost-effectiveness ratios (ICER) of less than €20.000 per (quality adjusted) life year gained with the exception of two evaluations. As all evaluations differed, yet still yielded similar results, there is no reason to assume that the Dutch program would not be cost-effective as well. This suggests that there is no urgency to conducting a new cost-effectiveness analysis for the Dutch neonatal heel prick program.

Preface

A year ago my interest in finding a master thesis topic concerning prevention led me to mr. Dirk Stemerding. He introduced me to Pepita Groeneveld (Ministry VWS) who introduced me to Barbara Hoebee (RIVM-CvB). Together we came up with a construction that allowed me to work on my thesis while getting to know both the RIVM and the Ministry of VWS. My research and my time at VWS brought me into contact with many interesting people, and many thanks goes out to my colleagues at both the RIVM and VWS for their help, information and (lacking a good English word) "*gezelligheid*".

The report you are now reading is the conclusion of my research, and also the conclusion of my master studies in Health Sciences at the University of Twente. When I started the research for this report I was somewhat of a stranger to cost-effectiveness evaluations. I knew the basics of how it was done and how to read the results, but the actual content was unfamiliar territory. This gave me a challenge, for I first had to understand the evaluations before I was be able to analyse them.
Fortunately I did not have to meet this challenge alone. I owe a great deal of thanks to my supervisor at the RIVM, Dr. Matthijs van den Berg, for his patience (especially in reading my many versions of this report), explanations, criticism and good discussions. Without his guidance this report would have turned out quite different!

There are several others to thank here. Prof. dr. Maarten IJzerman, for taking over as supervisor from the University when Dirk Stemerding left and Dr. Magda Boere-Boonekamp for agreeing to be second supervisor partway through my research. The two supervisors complemented each other quite well, as the feedback I got from them always concerned mostly different parts of my report.
I also want to thank drs. Eugenie Dekkers, program coordinator for the Heel Prick, dr. Bert Elvers and dr. Gerard Loeber for providing all the information I needed on the Dutch program and helping me with additional contact information for (inter)national screening agencies. Also a big thank you to Paul van Gils for speeding up the making of new conversion tables to convert all my cost-effectiveness results to 2008 euros. From the national screening agencies my thanks goes to Bridget Wilcken, Rodney Pollitt, Elise Houssin, Jean Louis Dhondt, Ulrika von Döblen, Masaru Fukushi, Brad Therrell, Emma Scott and Louise Galloway for providing information on their national screening programs.

Last, but definitely not least, I want to thank my family, friends and colleagues (at all workplaces) for their support, interest, comments and distractions when I needed them.

I wish you happy reading!

Loes Leewis

Bilthoven, June 2009

Contents

Samenvatting

In 2008 heeft het Centrum voor Bevolkingsonderzoek (CvB) een inventarisatie gedaan om in beeld te brengen welke Nederlandse kosteneffectiviteit studies er beschikbaar waren voor haar bevolkingsonderzoeken. Uit deze evaluatie bleek dat er tot nu toe nog geen studie was gedaan om de kosteneffectiviteit van het hielprik programma in beeld te krijgen. Vanwege de recente uitbreiding (2007) van screenen op 3 aandoeningen naar screenen op 17 aandoeningen was er prioriteit om de kosteneffectiviteit van juist dit programma in beeld te brengen. Voor dit rapport is onderzoek gedaan om inzicht te krijgen in bestaande kosteneffectiviteit studies die in het buitenland gedaan zijn, en vervolgens is er gekeken hoe deze resultaten in Nederland gebruikt kunnen worden.

Voor informatie over buitenlandse studies is er een systematisch literatuuronderzoek gedaan. Om zeker te zijn dat alle studies (ook niet gepubliceerde studies) meegenomen zouden worden is er ook contact geweest met nationale screening organisaties met een verzoek om informatie. De gevonden studies zijn aan de hand van vertaalbaarheids criteria en kwaliteitsrichtlijnen beoordeeld om hun bruikbaarheid voor Nederland in kaart te brengen.

Uit het literatuuronderzoek kwamen 20 kosteneffectiviteitstudies naar voren die betrekking hadden op neonatale hielprik screening voor tenminste één van de aandoeningen waar in Nederland op gescreend wordt. Van deze 20 studies gingen er maar 7 over uitgebreide hielprik programma's (met tandem-massa-spectrometrie(MS/MS). De overige 11 studies hadden betrekking op één of twee aandoeningen, terwijl informatie over hele programma's gewenst was. Van de 7 studies over uitgebreide programma's waren er maar 4 die voldeden aan alle vertaalbaarheids criteria en (de meeste) kwaliteitsrichtlijnen.

Bijna alle studies, ongeacht het aantal aandoeningen dat meegenomen werd, gaven een incrementele kosteneffectiviteit (ICER) van minder dan €20.000 per (voor kwaliteit gecorrigeerd) gewonnen levensjaar. Er waren maar twee studies die boven deze grens vielen, één studie van een programma dat onbehandelbare aandoeningen meenam, en één studie over maar één aandoening. Hoewel alle studies op meerdere punten van elkaar afweken kwamen ze allemaal tot ongeveer hetzelfde resultaat. Er is daarom geen reden om aan te nemen dat het Nederlandse programma niet kosteneffectief zou zijn. Dit geeft aan dat er geen urgentie is om op korte termijn een nieuwe kosteneffectiviteit studie in Nederland uit te schrijven. Als meer specifieke informatie over de kosteneffectiviteit van het Nederlandse programma gewenst is, kan het de moeite waard zijn om de resultaten van de 4 overgebleven studies daadwerkelijk te vertalen naar de Nederlandse situatie. Hierbij is het goed mogelijk dat niet alle informatie over kosten en effecten in Nederland bekend is. Het is aan te bevelen, gezien de jaarlijkse rapportage over monitoring en evaluatie van het programma, om informatie over kosten en effecten mee te nemen in die rapportage.

Vanuit de studies kwam een belangrijk discussiepunt naar voren betreffende het aantal aandoeningen waarvoor gescreend wordt. Een aantal studies stelt dat het toevoegen van aandoeningen aan een hielprik programma niet altijd gunstig is voor de kosteneffectiviteit. Het screenen op aandoeningen waar geen geschikte test of geschikte behandeling voor is, en aandoeningen met erg lage incidentie kunnen er zelfs voor zorgen dat een programma minder, of zelfs helemaal niet kosteneffectief meer is.
Dus, hoewel er geen urgentie is om een nieuwe kosteneffectiviteit studie uit te voeren is het wel belangrijk dat vóór het toevoegen van nieuwe aandoeningen altijd de impact op kosteneffectiviteit wordt onderzocht.

Summary

In 2008 the Center for Population Screening (CvB) made an inventory of Dutch cost-effectiveness evaluations for all its screening programs. This inventory showed that so far, no evaluation had been conducted to determine the cost-effectiveness of the heel prick program. In 2007 the Dutch neonatal heel prick screening program was expanded from 3 to 17 conditions and therefore received priority for determining its cost-effectiveness. This report is meant to show what is known about the cost-effectiveness of the neonatal heel prick program through cost-effectiveness evaluations done abroad.

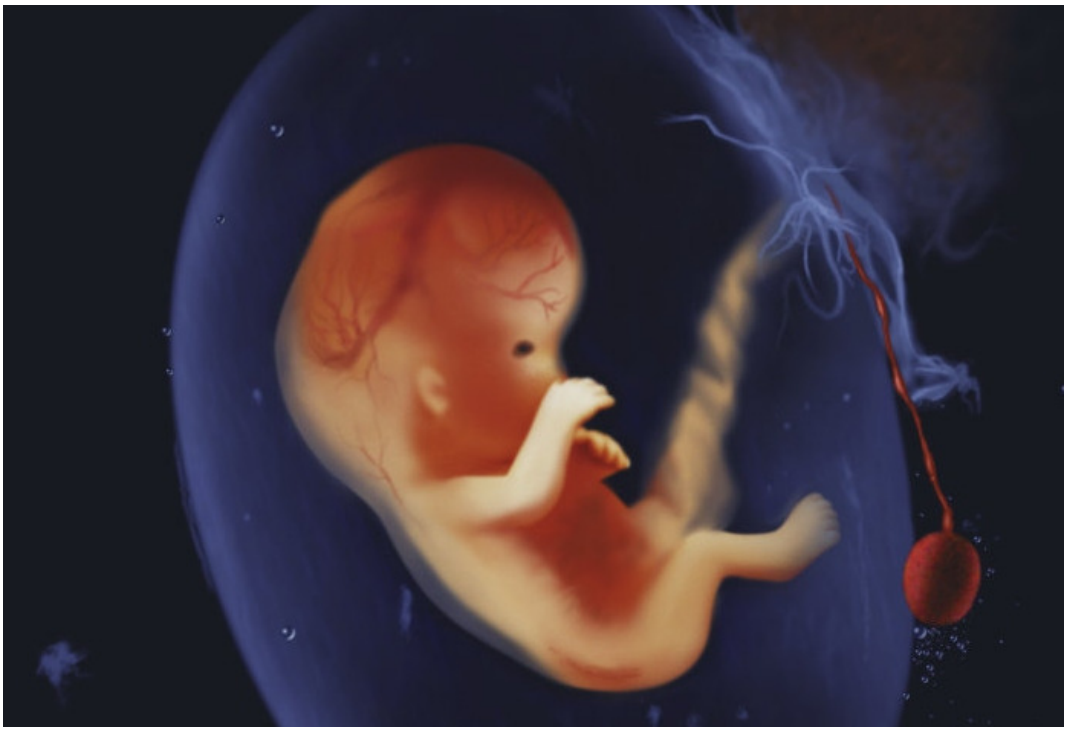
To gain information on international cost-effectiveness evaluations for this report, a systematic literature review was conducted. To ensure that no evaluations would be missed (e.g. non published) national screening agencies were contacted with an information request. The resulting evaluations were subjected to transferability criteria and quality guidelines to determine whether their information could be relevant for the Netherlands. The information from national agencies was also used for a background analyses to show differences in (inter) national programs.

The literature review showed that 20 cost-effectiveness evaluations had been done on neonatal heel prick screening for at least one of the conditions screened for in the Netherlands. Out of these evaluations, 7 concerned tandem-mass-spectrometry (MS/MS) screening for an expanded panel of conditions, with the remaining 11 evaluations concerning one or two conditions. These 11 evaluations are therefore not directly relevant for providing information on the cost-effectiveness of the *expanded* Dutch program. Out of the 7 evaluations on expanded programs, only 4 evaluations met the basic transferability criteria and (most) quality guidelines.

Almost all evaluations, no matter the number or type of conditions screened for, yielded incremental cost-effectiveness ratios (ICER) of less than €20.000 per (quality adjusted) life year gained. Only two evaluations yielded a higher ICER, one that included untreatable conditions and the other on only one condition. As all evaluations differed, yet still yielded similar results, there is no reason to assume that the Dutch program would not be cost-effective as well. This suggests that there is no urgency to conducting a new cost-effectiveness analysis for the Dutch neonatal heel prick program. However, if more exact data is wanted on the cost-effectiveness of the Dutch program, it can prove worthwhile to transfer the results from the 4 final evaluations to fit the Dutch situation. There is a possibility that not all information on costs and effects is known in the Netherlands. It is recommendable, as the Dutch program is evaluated and monitored yearly, to include information on costs and effects in the yearly evaluations.

A point of discussion came forth from the evaluations, saying that increasing the number of conditions screened for will not necessarily improve the cost-effectiveness. Screening for conditions that have no suitable test, no suitable treatment or extremely low incidence might actually cause a program to cease being cost-effective or at least reduce cost-effectiveness. Therefore, while there is no urgency to conducting a new cost-effectiveness evaluation, the impact on cost-effectiveness should be assessed before new conditions are added to the program.

1 Introduction



1.1 Introduction

The Center for Population Screening (CvB) of the National Institute for Public Health and the Environment (RIVM) is a relatively new addition to the RIVM, having been realized in 2006. The CvB organizes and executes the eight national disease prevention programs of the Netherlands as an agent for the Ministry of Health. All these national programs are aimed at diseases or conditions that, if detected early, can be treated to prevent symptoms or lessen those symptoms *(except for the screening for Down syndrome)*. All programs are designed to cater to certain target groups (e.g. Breast cancer screening for women between 50-75 years old) and are completely voluntary. In order to ensure that these programs are all working as they are intended, they are regularly evaluated. One of the factors considered in these evaluations is cost-effectiveness[1].

In 2008 the CvB conducted an evaluation to determine if any appropriate cost-effectiveness evaluations had been done in the Netherlands for the national population screening programs[2], and whether this information was still up to date (*Rapportage Meerjarenplanning Kosten-Effectiviteits studies*, 2008). The report concluded that at present there are three screening programs that required additional review as they had not been included in any recent national economic evaluations. These three are the expanded *neonatal heel prick program*, *the neonatal hearing screening* and the *prenatal screening for infectious diseases*. This report focuses on the neonatal heel prick program[3] because of the recent expansion.

[1] http://www.rivm.nl/en/aboutrivm/organization/public_health/cvb/index.jsp
[2] These are the cervical cancer screening, the breast cancer screening, the national influenza prevention program, the hereditary hypercholesterolemia program, prenatal screenings (infectious diseases, erythrocyte immunization/Down syndrome) and neonatal screenings (hearing and heel prick)
[3] The heel prick screening uses a blood sample collected from the heel of a newborn to test for rare conditions that if treated early can prevent irreparable damage to the health of the child.

1.2 Social and Scientific Relevance

In 2007 the Dutch neonatal heel prick program was expanded from testing for only three rare conditions to testing for seventeen. The only cost-effectiveness analyses (CEA) that have been done for the Dutch neonatal heel prick program concerned the addition of one condition to the existing program. The previous program of screening for three conditions was never evaluated for incremental cost-effectiveness compared to symptomatic diagnosis (no screening). No CEA have been done to evaluate the expansion of the Dutch heel prick screening program, as the program has only been operational for two years. However, the Netherlands is not the first country to implement an expanded screening program. In other countries economic evaluations have been done to determine the cost-effectiveness of expanded neonatal screening.

Conducting a new CEA would be expensive and time-consuming, and is therefore not undertaken lightly. As the heel prick program was recently expanded to testing for almost six times the previous number of conditions, it is important to know whether the program is still cost-effective. It is important to find the best available evidence, and to determine if this evidence can help in deciding whether or not further cost-effectiveness information needs to be gathered for the heel prick program in the Netherlands.

1.3 Research Questions

This report aims to answer the following research question:

What is known about the cost-effectiveness of the neonatal heel prick worldwide, and (how) can this information be used for decision making about the cost-effectiveness of the expanded neonatal heel prick program in the Netherlands?

This research question can be divided into two sections:

1. What is known about the cost-effectiveness of the neonatal heel prick worldwide
    a. What economic evaluations have been done worldwide?
    b. Is there evidence of the cost-effectiveness of expanded heel prick programs?

2. (how) Can this information be used for decision making about the cost-effectiveness of the expanded neonatal heel prick program in the Netherlands?
    a. Are these economic evaluations transferable to the Netherlands, and do they meet Dutch quality standards?

Course materials, information from the CvB, literature and (informal) interviews were used to gather background information and theoretical data on heel prick programs. To answer the questions in section 1 a systematic literature review was done. To gather the data for the review a search was done using the Pubmed database and national agencies worldwide were contacted to collect the required data.

For section 2 of the research question existing theories and techniques for determining quality and transferability of economic evaluations were used, and the information gathered was assessed by converting the results to a common currency.

1.4 Report structure

Chapter 2: The neonatal heel prick
Chapter two will elaborate on the neonatal heel prick. This chapter contains information on the current Heel Prick program in the Netherlands, including the 2007 expansion of the program. It also gives information on neonatal screening programs conducted in other countries to show some of the differences in screening programs worldwide.

Chapter 3: Economic evaluations
The third chapter of this report will focus on economic evaluations in health care. In this chapter different theories on generalizability and transferability will be listed, as well as the Dutch quality criteria for (conducting) economic evaluations.

Chapter 4: Systematic review
Chapter 4 presents the results of the systematic review of economic evaluations done on (expanded) neonatal heel prick screening programs. The data from chapter 3 on criteria for economic evaluations was used to determine to what extent the evaluations found in the systematic review match the criteria on quality and transferability for economic evaluations in the Netherlands.

Chapter 5 and 6: Conclusion and Discussion
The previous chapters are followed by a conclusion and discussion on whether a statement about the cost-effectiveness of the Dutch neonatal heel prick program can be made based on the international cost-effectiveness results.

2 The Neonatal Heel Prick program

2.1 Introduction

A *heel prick program* focuses on rare conditions[4] that, if detected in the first week of a newborn's life, can be treated to stop or lessen the consequences of the conditions. The first disease to be detected and diagnosed with the heel prick was Phenylketonuria (PKU), a disease that, if not detected early, causes (severe) brain damage and prohibits the child from leading a "normal" life. While a treatment for PKU was discovered in 1934, there was no means for early detection until 1963. At that time, microbiologist Robert Guthrie designed dried blood spot testing using a heel prick card to collect blood samples. It is a relatively simple procedure that is done by examining the blood taken from a heel of a new born child (between 4 and 7/8 days after birth) using a "heel prick kit". After the introduction of the testing method, dried blood spot testing was rapidly implemented in many countries, and is now used to screen for more and more metabolic diseases and enzyme deficiencies around the globe. Though the heel prick is the most common technique used for neonatal screening, other sample collection techniques are sometimes used. Examples of these are blood taken from a vein on the hand and naval cord blood *(limited number of conditions)*.

[4] e.g. fatty acid oxidation disorders, organic acidemias, urea cycle disorders and amino acidemias

2.2 Neonatal heel prick screening in the Netherlands

In the Netherlands neonatal screening using the heel prick started with PKU in 1974. The screening was expanded with Congenital Hypothyroidism (CH) in 1981, Congenital Adrenal Hyperplasia (CAH/AGS) in 2000 and fourteen more conditions in 2007 *(§2.3)*. A listing of all the conditions that are screened for in the Netherlands can be found in *Appendix B.*

The heel prick is carried out by an obstetric nurse or a trained screener in the home of the newborn. If the birth takes place in the hospital and the newborn is there during the first week, the heel prick is carried out in the hospital. In the Netherlands it is possible to indicate that you do not want to receive information about being a carrier of Sickle Cell Disease (SCD) (one of the conditions screened for). It is also possible to opt-out of giving permission for using the blood samples in scientific research after screening. The blood samples are sent to one of the five certified laboratories where testing for conditions takes place. When the test results come back, the newborn's parents will only receive results if an abnormality was found. In some cases the sample can be unclear, or warrant further testing. In that case a second heel prick is carried out within two weeks of the first heel prick, and, in contrary to the first heel prick, parents always receive the results of this second heel prick. If an abnormality is found, the newborn is referred to a general practitioner, and then to a hospital for further testing to ensure that the results were true-positive for the condition. As the results can not always be a hundred percent accurate, sometimes false-positives or false-negatives occur. In the first case, a child showed abnormal test results but does not have the condition. In the second case a child has the condition, but it is not detected in the screening. To ensure that this does not happen often, research is always done to determine the best "cut-off point" that will lead to the least amount of false results. However, for most conditions only limited information on (e.g.) specific test characteristics is available, and this can limit the evidence base for adding a new condition.

2.3 Why are conditions included in a neonatal heel prick program?

Before new conditions are added to the existing program, research is done to determine whether a condition meets the requirements of population screening. The criteria for screening that are used most often are those set up by Wilson and Jungner in 1968. These criteria state that any disease or condition has to meet ten criteria before it can be included in a screening program. In a recent article in the bulletin of the World Health Organization (WHO) these criteria were summarized *(Box1)*[5].

Box 1. Wilson and Jungner classic screening criteria (1968)

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 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 continuing process and not a "once and for all" project.

[5] http://www.who.int/bulletin/volumes/86/4/07-050112/en/index.html, Andermann et al, 2008

The WHO article focuses on the changing environment of screening and the emergence of new criteria that are deemed important to consider when implementing a screening program *(Box 2)*.[6]

Box 2. Synthesis of emerging screening criteria proposed over the past 40 years

1) The screening program should respond to a recognized need.
2) The objectives of screening should be defined at the outset.
3) There should be a defined target population.
4) There should be scientific evidence of screening program effectiveness.
5) The program should integrate education, testing, clinical services and program management.
6) There should be quality assurance, with mechanisms to minimize potential risks of screening.
7) The program should ensure informed choice, confidentiality and respect for autonomy.
8) The program should promote equity and access to screening for the entire target population.
9) Program evaluation should be planned from the outset.
10) The overall benefits of screening should outweigh the harm.

Box 2 shows that there is currently a shift in criteria that is more concerned with quality assurance and informing patients. While the original Wilson and Jungner criteria are still relevant, the changing health care environment has placed new demands on screening programs. Patients have more means and interests in finding information through new media, such as internet "self-help" websites. This autonomous patient feels in charge of his or her own health (care) and wants to be well informed about that own health (or that of their child). This means that the reasons for including conditions in a screening program are changing too. Where the focus was always on the test and treatment, more attention now goes to the need to know. In the Netherlands, however, the focus is still primarily on the benefit to the newborn. Other studies (e.g. van der Wilk et al, 2007) have stated that some countries include conditions that don't necessarily meet the Wilson and Jungner criteria, but are included in a program because parents wish to know (e.g. for family planning).

2.4 Expansion of the Dutch neonatal screening program

In 2003 the Ministry of Health sent out a request to the Health Council of the Netherlands (GR) to evaluate what conditions would meet all the criteria for newborn screening. A committee was assembled, and in 2005 the committee published her findings (Neonatale Screening, 2005).
The most important criteria they used were:

- the disorder should be clearly described
- there should be a suitable detection method
- the treatments should be available and accessible
- there has to be direct benefit to newborns
- participation should be voluntary (after receiving all information)
- informed consent is asked of parents (acting on behalf of newborn)

Based on these criteria the neonatal screening committee listed three categories of conditions:

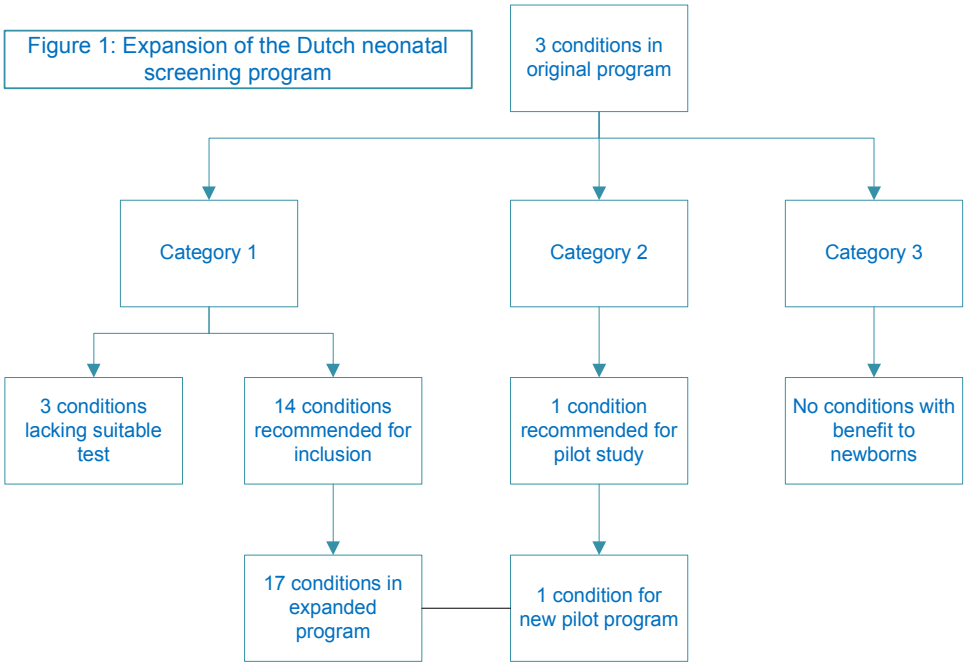
Category 1: Disorders for which considerable, irreparable damage can be prevented
Category 2: Disorders with a lesser degree of preventable damage or where evidence on benefits of screening is inconclusive
Category 3: Disorders where neonatal screening does not prevent damage to health

[6] http://www.who.int/bulletin/volumes/86/4/07-050112/en/index.html, Andermann et al, 2008

In **Category 1** seventeen conditions were identified as being detectable through newborn screening and having a well established treatment program. Of these, fourteen were recommended for inclusion in the neonatal screening program. The other three (cystinosis, carnitine palmitoyl transferase deficiency type 1 and carnitine transporter deficiency) lacked a suitable test for conclusive detection of the conditions through newborn screening.
In **Category 2** only one condition was considered for inclusion, Cystic Fibrosis (CF). The recommendation included that before the condition was added to the existing program more research needed to be done into a testing method with high enough specificity. This high specificity is needed to ensure the least amount of false-positives and chance findings (e.g. carriers of the condition).
As screening for conditions in **Category 3** would provide no benefit for newborns none of these conditions were included in the recommendations.

Based on the GR report the Ministry of Health accepted the recommendations *(figure 1)* to include the Category 1 conditions in the program and started a pilot program for CF to determine what the best testing method would be[7]. The assumption was made that the addition of these conditions to the existing program would result in more than doubling the number of newborns found annually.

Figure 1: Expansion of the Dutch neonatal screening program



The expanded screening program[8] was implemented in January 2007 and recently evaluated by TNO, an organization for applied scientific research in the Netherlands. It showed that in 2007 approximately 180.000 babies were born in the Netherlands, of which 99.75% underwent neonatal screening. In 2007 eleven newborns were diagnosed with CAH, fifty-seven with CH, ten with PKU, forty-one with SCD and sixty with other metabolic disorders (TNO, 2008). This gives a total of 179 newborns found, of which 101 newborns with an “expanded” condition, proving that the assumption of more than doubling the number of newborns found with a condition annually was correct.

[7] http://www.minvws.nl/images/pg-2635399_tcm19-102060.pdf
[8] screening for CAH, CH, PKU, BIOT, GAL, GA-I, HMG, HCS, HCY, IVA, LCHADD, MSUD, MCADD, 3-MCC, SCD, TYR-I and VLCHADD

2.5 Actors in the Netherlands

There are several actors involved with the organization and execution of the neonatal heel prick program in the Netherlands. These actors supervise, evaluate, advise, coordinate, conduct, represent and support the program *(figure 2).* (Draaiboek neonatale hielprik screening 2007,www.rivm.nl/hielprik).

Ministry of Health, Welfare and Sports

The Ministry of Health (MoH) has several roles when it comes to the national screening programs. The most important role is governance. This means that the MoH decides about the content of the programs, financing, political aspects, legislation and authorization. In its roles the ministry is advised by the *Centre for Population Screening (CvB),* the *Health Council of the Netherlands (GR)* and the *Council for Public Health and Health Care (RVZ).* The MoH also finances the coordination of the screening programs at the RIVM-CvB.

The neonatal heel prick is one of the programs governed by the MoH. The Minister of Health decides if and when new conditions are added to the program, and when conditions are taken out. Taken out a condition can happen when, for example, a certain method of testing is not accurate enough.

Coordinating Actors

The MoH governs the coordinating actors and takes their advice on matters concerning the execution of the program. The Coordinating actors for the heel prick are the RIVM-CvB, the RIVM Laboratory for Infectious diseases and Screening (LIS) and the RIVM regional offices.

RIVM (CvB)

The CvB takes care of the national coordination of the neonatal screening program, and acts as a focal point between policy and practice. They are also in charge of the yearly reports and analyses of data.

RIVM (LIS)

The LIS handles the national coordination and quality monitoring of the screening laboratories. Aside from the national role, the LIS is also a screening laboratory and functions as a reference laboratory. The LIS also stores and safeguards the used heel prick cards.

The RIVM regional offices (RCP)

The regional offices (formally known as ent-administrations) take care of the regional coordination of the national screening program and are in charge of:

- Distribution of information materials, regional education and the heel prick sets.
- Giving the orders to screeners to perform the heel prick when a newborn is registered.
- Referring children, through the GP, to specialized pediatricians.
- Supporting the maintenance of the quality of care.
- Being the rally point for all questions of professionals concerning the heel prick.

Advisory actors

The advisory actors for the CvB are the four Advisory committees on Neonatal Screening that give advice through the Program committee Neonatal Heel prick Screening (PNHS). The advisory actors are made up of experts from relevant areas in the field, and organizations with authority in their area of expertise or their network. In its role, the CvB also gives advice to the MoH based on information from (a branch of) the PNHS. The GR and RVZ advise the government and the parliament on matters concerning public health. Advice is usually given as a response to a request from a Minister or Secretary of State, but the actors can also give unsolicited advice if they sense a danger to the public health of people in the Netherlands.

Supporting actors

The supporting actors are the committees and working groups that support the Program committee Neonatal Heel prick Screening (PNHS). These committees consist of experts from relevant areas in the field and organizations with authority in their network or area of expertise. The supporting actors give advice to the PNHS and the CvB about different aspects of the program, and signal when (parts of) the program might be in need of adjustment.

Executing actors

Through the regional coordination (RCP) notice is sent to the executing actors when a baby is born, that a heel prick needs to be conducted. The executing actors are the heel prick conductors (*obstetric nurse, screener*), the screening laboratories and the general practitioners. They are in charge of conducting the heel-prick and ensuring that proper follow-up action is taken if needed. While the heel prick is usually conducted in the home of the newborn, it is possible that the heel prick takes place in a hospital. In that case the heel prick sample will be taken by hospital staff.

The Obstetric Nurse

The nurse informs parents during their first consultation and the delivery consult. An obstetric nurse can also be a screener if he or she conducts the heel prick.

Screener

The screener has to ensure that the parents have previously received information about the heel prick and, if necessary, to provide additional information. The screener has to fill in the information on the heel prick card and then conduct the heel prick. The final role of the screener is to send the heel prick card to the screening laboratory.

Screening laboratory

The laboratory conducts the examination of the heel prick blood sample, and guards for ambiguous or abnormal results. When a test shows ambiguous results, a second heel prick sample will be taken from the newborn and sent to a reference laboratory.

General Practitioner (GP)

The GP is in charge of referring patients to specialists if the heel prick shows abnormal results. The GP is also in charge of supporting parents in case of abnormal results. The GP refers patients to a pediatric specialist. The pediatrician sees the patient after a referral from the GP and does additional testing to confirm if a newborn is really afflicted with the condition that showed abnormal results. If this is indeed the case, the specialist will begin follow-up treatment immediately and, if needed, give referrals for other specialists (such as dieticians).

Representative actors/ interest groups

The representative actors and the interest groups represent different groups of patients and professionals. Examples are the National Association for Pediatricians (NVK), the National General Practitioners Association (NHG) and the associations representing patients with a condition that is screened for in the heel prick program. These actors are usually represented in the advisory committees mentioned before.

Evaluating actors

TNO – QoL (Quality of Life)

At present, TNO handles the yearly monitoring of the heel prick program. They collect the data needed from regional offices and pediatricians and use this data to conduct an epidemiological evaluation of the program. Important factors that are monitored are the number of false-positives/false-negatives and the number of children detected through the heel prick for each of the conditions. With time, this role could be fulfilled by a national database, in which all pediatricians and GPs can contribute data.

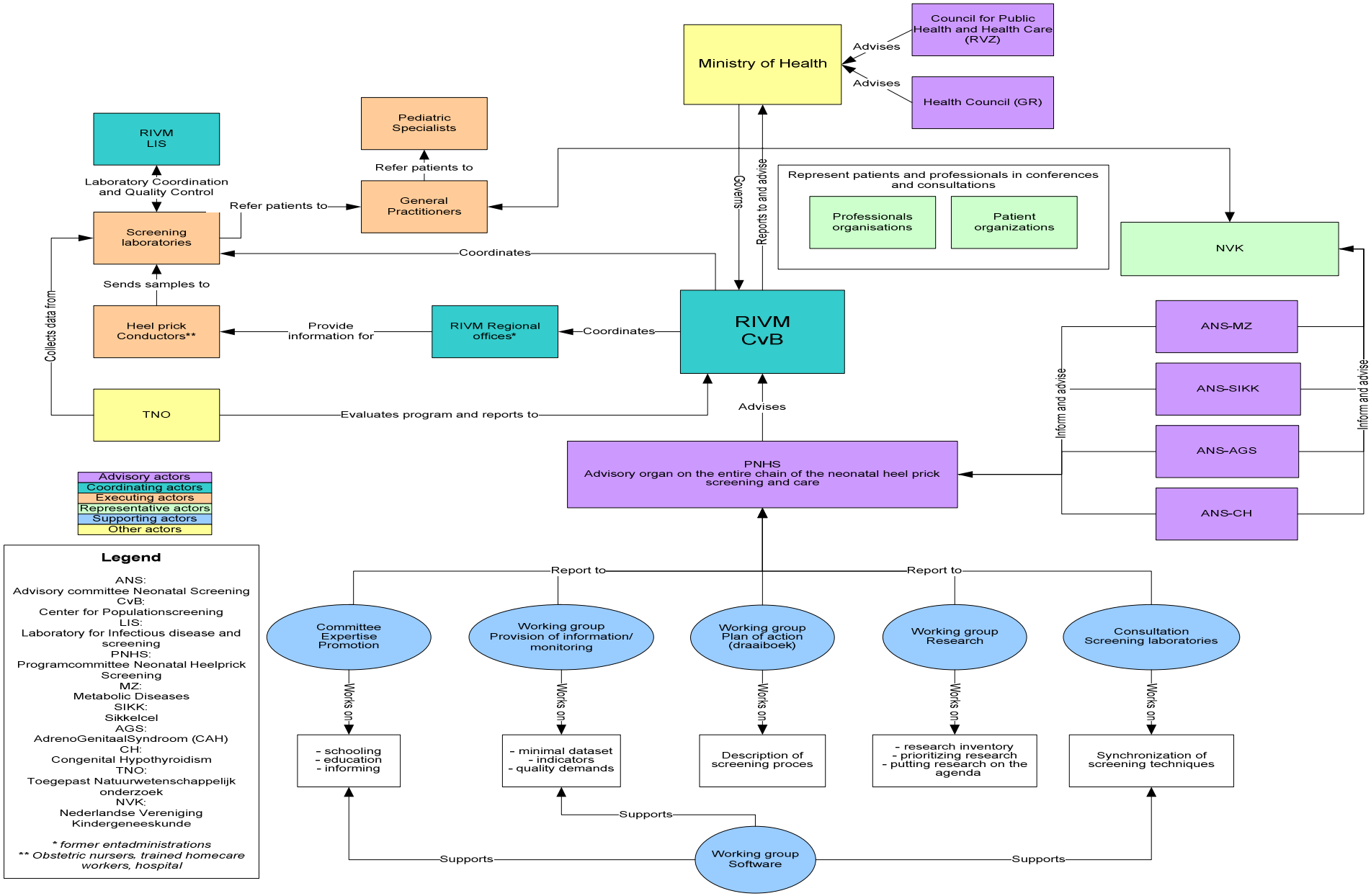


Figure 2: Organization of the Heel Prick Program in the Netherlands

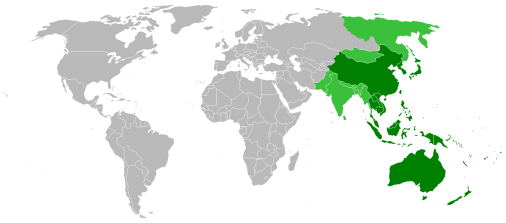
2.6 International Programs

Most countries around the globe have at least some form of a newborn screening program. However, these programs vary greatly, with some programs that have full national coverage, and some programs having such low coverage that they can not even be called programs. If participation or coverage is low a (newborn) screening program will have less impact. The chances of not finding an affected child will increase with a low participation or low coverage. The differences in coverage show that in the developed countries there is almost always a national program, and that some of the developing countries have a regional or pilot program but there are also countries with no structured program at all. In developing countries participation and coverage of newborn screening is extremely low, often due to lack of facilities such as laboratories, equipment and skilled staff (Padilla&Therrell, 2007) and other health problems that require more immediate attention (e.g. AIDS and healthy drinking water). As these developing countries differ greatly with the Netherlands they are not mentioned in detail in this report. However, to give an indication that even in developed countries there are differences between programs, a number of screening programs will be described. A selection was made to include each continent except for Africa, as the prevalence of conditions and availability of health care differs too much with other continents.

Latin America (Borrajo, 2007)

In Latin America most countries still do not have a structured national newborn bloodspot screening program (Borrajo et al. 2007). Cuba was the first to implement a national screening program in 1986, where other countries only started implementation on a request basis. Brazil's neonatal screening program started in 1976 on a request only basis, with national screening implemented in 2001. This program is mandatory and run by the government. Mandatory screening includes screening for CH (1:2453), PKU (1:25294), CF and Hemoglobinopathies. While Ms/Ms screening is available it is currently on a request basis only. Even though the program is mandatory, it has a participation rate of only 80.2 percent. Borrajo et al (2007) mention that since the program was started in 2001 coverage has been expanding rapidly, but has not yet reached full national coverage which would explain that participation is not yet 100 percent.

Asia-Pacific



About half of the world's newborns are born in the Asia-Pacific region *(approximately Mongolia to New Zealand)*. This is mostly due to India and China falling in this region. This region has the greatest differences between countries when it comes to newborn screening as it includes both developed and developing countries. Out of the developed countries in this region Australia and Japan are discussed in more detail, as one represents an Asian country with a structured program and one a Pacific country.

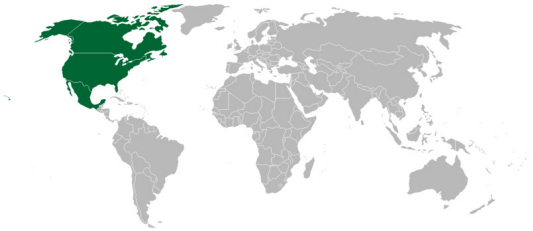
In *Australia*[9] approximately 260.000 babies are born annually, of which at least 98 percent participate in neonatal bloodspot screening. Screening has been organized and funded by the government since 1976 and currently includes voluntary screening for PKU (1:14000), CF (1:2900), MCADD (1:19750), GA-I (1:60000) and a range of other conditions tested with MS/MS technology. Overall around 1 in 3684

[9] Data provided by Professor Bridget Wilcken, NSW Biochemical Genetics and Newborn Screening Service

newborns is born with a condition that is tested for using newborn screening. The heel prick blood sample is taken by nurses in hospitals, and in case of a home birth by a midwife. The testing takes place in five screening laboratories and is centralized for each state. A national committee (Human Genetics Society of Australasia) advises the states. Quality Assurance is based on national criteria, though the five heads of the laboratories are responsible for their own quality control.

National Newborn Screening (NBS) in *Japan*[10] started in 1977 and currently covers over 99 percent of the population. This means that approximately 1.100.000 newborns are screened each year for PKU (11-16)[11], HCY(1-2), GAL(1-2), MSUD(1-2), CH (500) and CAH(50) of which an average of 1:2000 newborns have one or more conditions. Screening for the first four conditions was started in 1977 with CH added in 1979 and CAH in 1988. The program is fully funded by the government. A nurse or doctor at a birthing institute conducts the heel prick and sends the sample to one of the forty-eight decentralized screening laboratories. The individual laboratory directors are responsible for quality assurance. Laboratories are evaluated monthly through the centralized national *External Quality Control Program*.

North America



In North America lies the origin of newborn blood spot screening, where the first programs were developed by Robert Guthrie around 1962. Laboratories and research in the USA and Canada also developed most of the equipment that is currently used to test heel prick samples. These first programs were not, and are still not, organized on a national level with a universal standard (Therrell et al, 2006; Bradford, 2007). Each state or province has its own program, varying from testing for three conditions in one state to more than fifty in others *(National Newborn Screening Status Report, 18/05/2009)*[12]. In 2004 a consensus was reached in the USA, and it became mandatory (by law) for each state to screen for a core panel of twenty-nine conditions (Watson et al, 2006). Currently twenty-six states meet this criterion[13]. The original implementation of neonatal bloodspot screening spread gradually from state to state, with the first program started in 1962, and the last state program started in 1985. In 1978 a national Quality Assurance (QA) program was initiated through a national laboratory and the NNSGRC.[14] When DNA screening and MS/MS testing became possible in the late 1980s, the screening programs in the USA were the first to be expanded.

Approximately four million babies are born annually in the USA with an average of 1:750 having one of the conditions tested for. Screening currently includes PKU, CH, GAL and SCD in all state programs, CAH and BIOT in most programs, and a growing number of states use MS/MS testing. The heel prick sample is taken at the hospital or by a midwife if the child is not in the hospital. About 25 percent (certain states) of all newborns receive double testing (with different technologies), with the second sample taken at the office of a physician. In total there are thirty-five screening laboratories nationwide, meaning that while every state has its own program, not every state has its own testing facility. The Centers for Disease Control and prevention (CDC) provides external proficiency testing for all laboratories. As state programs vary, so does the financing of the programs. Only four states have

[10] Data provided by Mr. Masaru FUKUSHI, Sapporo City institute of Public Health
[11] Cases found annually
[12] http://genes-r-us.uthscsa.edu/nbsdisorders.pdf
[13] http://www.marchofdimes.com/peristats/tlanding.aspx?reg=99&top=12&lev=0&slev=1
[14] National Newborn Screening and Genetics Resource Center, http://genes-r-us.uthscsa.edu/

complete public financing, the other states combine public funding with a fee. In twenty states this is a fee of <$50 and two states have a fee >$100 with the highest charged fee being $139.33. In some cases this fee is covered by health insurance, and if patients are uninsured hospitals pay the required fee, though this differs per state.[15] Screening takes place on a voluntary basis, though a written statement is required if parents choose not to participate in the screening of their child.

	PKU[1]	CH	CAH	CF	SCD	GAL	BD	MCAD/ ms/ms	G6PD	DMD
Albania	--	--	--	--	--	--	--	--	--	--
Armenia	--	2005	--	--	--	--	--	--	--	--
Austria	1967	1978	2001	1997	--	1967	1986	2002	--	--
Azerbaijan	--	--	--	--	--	--	--	--	--	--
Belarus	1978	1991	--	--	--	--	--	--	--	--
Belgium	1964	1974	1984	1981	2003	1968	1987	1998	1979	--
Bosnia-Herzegovina	2000	2000	--	--	--	--	--	--	--	--
Bulgaria	1978	1993	--	--	--	--	--	--	--	--
Croatia	1978	1985	--	--	--	--	--	--	--	--
Cyprus	1989	1989	--	--	--	--	--	--	--	--
Czech Republic	1975	1985	2006	2005	--	--	--	--	--	--
Denmark	1975	1978	--	--	--	--	--	2002	--	--
Estonia	1993	1996	--	--	--	--	--	--	--	--
Finland	--	x	--	--	--	--	--	--	--	--
France	1967	1978	1981	2002	1996	--	--	--	--	--
Georgia	--	2007	--	--	--	--	--	--	--	--
Germany	1969	1981	1996	1990	--	1969	1996	2000	1996	--
Greece	x	x	--	--	--	--	--	--	x	--
Hungary	x	x	--	--	--	x	x	--	--	--
Iceland	1972	1979	--	--	--	--	--	2007	--	--
Ireland	1966	1979	--	--	--	1972	--	--	--	--
Italy	1973	1979	1991	1981	--	1977	1989	2003	1982	--
Latvia	1985	1995	--	--	--	--	--	--	--	--
Liechtenstein	1965	1978	1992	--	--	2003	2003	2004	--	--
Lithuania	1975	1993	--	--	--	--	--	--	--	--
Luxembourg	1968	1978	2001	--	--	--	--	--	--	--
Macedonia	x	x	--	--	--	--	--	--	--	--
Malta	--	x	x	--	--	--	--	--	--	--
Moldova	1989	1989	--	--	--	--	--	--	--	--
Montenegro	--	2007	--	--	--	--	--	--	--	--
Netherlands	1974	1981	2000	p	2007	2007	2007	2007	--	--
Norway	1965	1979	--	--	--	--	--	--	--	--
Poland	1964	1983	--	--	--	--	--	--	--	--
Portugal	1979	1981	--	--	--	--	--	2005	--	--
Romania	x	x	--	--	--	--	--	--	--	--
Russia	1984	1989	2002	2002	--	2006	--	--	--	--
Scotland	1965	1979	--	2003	--	--	--	--	--	--
Serbia	1982	1983	--	--	--	--	--	--	--	--
Slovakia	1972	1985	2005	2007	--	--	--	--	--	--
Slovenia	1979	1979	--	--	--	--	--	--	--	--
Spain	1968	1979	1990	1999	2003	p	p	2002	--	--
Sweden	1965	1980	1986	--	--	1967	2002	--	--	--
Switzerland	1965	1978	1992	--	--	2003	2003	2004	--	--
Turkey	2002	2004	2004	2004	2005	2002	2002	2005	2004	--
Ukraine	x	p	--	--	--	--	--	--	--	--
Un.Kingd.(excl. Scotl./Wales)	1969	1981	--	1979	1978	--	--	--	--	--
Wales	1970	1981	--	1996	--	--	--	--	--	1990
Total (excl. pilots)	40	43	16	13	6	12	10	12	5	1

[1] PKU= phenylketonuria, CH = congenital hypothyroidism, CAH = congenital adrenal hyperplasia, CF = cystic fibrosis, SCD = sickle cell disease, GAL= galactosemia, BD = biotinidase deficiency, MCAD = medium chain acyl-CoA dehydrogenase deficiency, G6PD = glucose-6-phosphate dehydrogenase deficiency, DMD= Duchenne Muscular Dystrophy
[2] n.d. = no data available
[3] x = nation wide program, starting year unknown
[4] p = pilot program or in part of the laboratories

Table 1: Conditions screened for in Europe (Loeber, 2009)

[15] Data from http://www2.uthscsa.edu/nnsis/ NNSIS USA screening information system

Europe

In Europe almost all countries screen for at least three conditions, PKU, CH and CAH. Almost all European screening programs cover more than 80 percent of newborns, though not all country data is available. More and more countries are adopting MS/MS screening and expanding the range of conditions tested for. Within the continent, the prevalence of conditions varies greatly. On prevalence Loeber (2007) mentions that PKU varies from 1:3000 to 1:30000 and CH from 1:1300 to 1:13000. *Table 1 (above)* shows the conditions screened for and when screening for them was started. Screening differs greatly, in part due to the differences in prevalence, in part because of the organization of national programs and partly because of the choices countries make about their national programs. As Europe hosts a wide variety of ethnic groups, it is no surprise that these programs differ greatly, for countries would choose to start their screening programs by including conditions with the highest prevalence. To give an example of different programs in Europe, the French and the Swedish program will be discussed here.

In *France*[16] 840.000 babies are born annually, of which approximately 100 percent receive neonatal screening. Babies are screened for PKU (1:16500), CH (1:3500), CAH (1:18500), (targeted screening) SCD (1:790/1:450 (colonies)) and CF (1:4400). The PKU program was started in 1972, with CF added last in 2002. Since the start of the program, 31.600 newborns were diagnosed through newborn screening. The heel prick is conducted by nurses in a hospital, and results are sent to one of twenty-three laboratories nationwide. Each of these laboratories is a regional association, and Quality Control is done through a technical committee of AFSSAPS[17]. Besides the regional coordination, there is a central organization that governs the regional delegations. The heel prick is funded by the National Health Insurance without an additional private charge.

While most countries in Europe screen with the heel prick technique, some countries use a different technique. In *Sweden*[18] for example dried blood spot cards are filled with blood taken from a vein on the back of the hand. Sweden screens approximately 100.000 newborns every year for PKU (1:17000), GAL (1:12500), CH, CAH (1:12500) and BIOT (1:58000). These samples are taken mostly at hospitals by midwives. All samples are tested at a centralized, accredited laboratory, and the program is coordinated by the head of that laboratory. More than 99,5% of newborns take part in the national screening program, and heel prick screening is fully funded by the government.

2.7 Conclusion

This chapter has shown that there are many differences in the conditions that are screened for around the world. While some conditions like PKU and CH are screened for in most developed countries, there are great differences in expanded (MS/MS) newborn screening. The conditions screened for depend on the incidence of those conditions and the available means and materials for screening. For a country to adopt a new condition, it has to be absolutely sure that all testing materials are accurate and that enough follow-up professionals are available to provide a good quality of care.

[16] Data provided by Elise Houssin and Professor Jean-Louis Dhondt, French association for the screening and Prevention of Infant Handicaps (AFDPHE)
[17] http://www.afssaps.fr/ / www.afdphe.asso.fr
[18] Data provided by Ulrika von Döblen, head of PKU laboratory, Centre for inherited metabolic diseases

Especially countries that are not fully economically developed can have difficulties in finding the means and organizational facilities necessary to create national newborn screening programs. Taking into account all countries discussed in this chapter, the Netherlands is one of the front runners of expanded newborn screening in Europe.

While all countries state that they maintain the Wilson and Jungner criteria for screening, some have included conditions for which there is currently no well established treatment program. For example, some states in North America screen for more than 50 conditions, including conditions that, at present, have no treatment possibilities. In the Netherlands the choice was made that, while it is possible to reveal that a child *has* an untreatable condition, specific detection of these conditions through screening would not be reported. As this knowledge would not help the child in question, screening for untreatable conditions would be ineffective. These findings concur with the RIVM report *Leren van de Buren* (van der Wilk et al, 2007).

In the next chapter the methods of economic evaluation and the criteria for economic evaluations in the Netherlands will be discussed.

3 Economic Evaluation in Healthcare

3.1 Introduction

As resources are usually scarce, economic evaluation can be important in deciding whether a new technology should be implemented or not. An economic evaluation compares costs and effects of multiple interventions *(usually an old and a new or future intervention)*, resulting in an incremental cost-effectiveness ratio (ICER) for the new intervention. The ICER is calculated through division of the difference in costs by the difference in effects.

There are three main approaches for conducting economic evaluation, differing in the way the effects are measured. These are cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility analysis (CUA). *CBA* measures all benefits of an intervention in monetary terms. While this gives a broad view about the intervention, it is often controversial to express health benefits solely in monetary units. *CEA* measures benefits in natural effects such as life years gained (LYG) or cases detected (CD). CUA also measures benefits in natural effects, but takes into account the utility value of the effects. For example, where CEA looks purely at the number of life years gained, CUA looks further and includes the quality of life in those gained life years (e.g. if the affected person is able to function on his/her own or is able to work). The results of a CUA are measured in Quality Adjusted Life Years (QALY).

This means that for a CBA the ICER will be a cost to benefit ratio, for CEA it will be (e.g.) costs per LYG and for CUA costs per QALY. In the Netherlands an intervention is considered to be cost-effective if the ICER is no higher than €20.000 per QALY, though there is some discussion about this threshold being too low (Pomp et al., 2007).

3.2 Transferability and Generalization of economic evaluations

Results from economic evaluations done in one country will not necessarily be the same, or equal to results from that same evaluation if done in another country. To determine whether results from evaluations done in different countries are usable, there are many criteria for determining the transferability and generalizability of economic results. In this paragraph a number of theories on possible criteria will be listed.

Transferability
Transferability pertains to the setting in which an evaluation takes place. Results can only be transferred from one setting to another if the circumstances are similar. For example, results from a developed country can not easily be transferred to a developing country because of the differences in means and availability of health care. Transferring means applying the results to fit your own setting, but if the original setting differs in too many ways, transferring results might not be possible.

Welte et al (2004) presented a decision chart for assessing transferability of data from economic evaluations between countries. In this chart they take into account several factors on which evaluations can differ *(Methodological characteristics, Healthcare system characteristics and population characteristics)*. With regards to these factors, Welte names *knock-out criteria.* He divides these criteria into general and specific *knock-out criteria*. Knock-out criteria consist of factors that if (not) present make the transfer of evaluations impossible. There are three general knock-out criteria, which, when one of them is present makes transferability near to impossible. These criteria are:

a) When the evaluated intervention is not comparable to the one that will be used in the country to which the results must be transferred.
b) When the comparator (the (old) intervention that is already in use) is not available in the country to which the results must be transferred.
c) When the study found does not meet the quality standards of the country to which the results must be transferred.

Besides these three *general* knock-out criteria, any of the transferability factors from their decision chart can become knock-out criteria if the factor is present but cannot be assessed due to insufficient data.

Boulenger et al (2005) build on the works of Welte et al. to create a sub-checklist with the most important factors that have to be considered when attempting to transfer economic evaluation results.
While these overlap with Welte et al., they also introduce a new factor, namely the amount and detail of information that is provided in the journal article. When there is little to no detailed information on costs and effects that led to the final results, transferability will be impaired.

Generalizability
If a study is generalizable, the results can be extended to reflect more than just the study population. For example, a national survey does not require all citizens to participate, but aims to include a number of people that are ‘representative’ of the national population make-up. The results from the survey can be generalized to state that they reflect the general opinion of the entire nation, not just the survey group.

Oostenbrink et al. (2004, p.118) state that it is not possible to generalize results of economic evaluations without at least a minimal amount of additional research in the country to which the results are to be transferred. The reason for this is that countries differ to a degree on several factors:

- demographic and epidemiological factors
- cultural factors
- availability of health care and differences in medical practice
- financing for health care and financial stimulants for healthcare suppliers and patients
- absolute and relative differences in prices

Urdahl et al (2006) list a number of factors that can be of aid in determining whether a study is generalizable or not. The generalizability can be evaluated by determining whether certain factors are also relevant for the decision maker (the country to which data is to be transferred). These factors are:

- Data sources
- Resource data
- Unit costs
- Preferences and utilities
- Robustness (e.g. sensitivity analysis)
- Differences in compliance rates (e.g. participation)
- Were results compared to other relevant evaluations?

Urdahl et al focus on the *context* of the evaluations. Most of these factors are also included in the CvZ criteria mentioned in 3.3.

3.3 Quality guidelines for economic evaluation in the Netherlands

There are several criteria for evaluating the quality of economic evaluations. These criteria pertain to the way an economic evaluation was conducted and the information that is provided in the reports of economic evaluations. For the Netherlands criteria for conducting economic evaluations were set up by the CvZ (2006) in the guidelines for Pharmaco-economic research. While these guidelines pertain to economic evaluations of pharmaceuticals, they can also be interpreted to evaluate other health interventions. The guidelines they have set up are for:

- the chosen perspective of the evaluation
- the choice which alternative interventions to compare
- the technique used for the analysis
- the time horizon for the analysis
- the cost identification, measurement and valuation
- the valuation of quality of life and QALYs
- the type of model used to present the results
- Incremental analysis
- the used discount rate for future costs and effects
- the measurement of uncertainties and risks
- the use of experts

The chosen perspective: According to the Dutch guidelines, an economic evaluation should always be conducted using a societal perspective and should always include all costs and effects, no matter who the benefits or costs are for. A societal perspective includes not just the direct and indirect health care costs, but also takes into account other costs and benefits such as education and loss (or gain) of productivity.

The choice of alternatives for comparison: The new intervention should be compared with the standard treatment, or if there is no treatment yet available, compared to no intervention at all.

The analysis technique: This guideline states that a CUA is most preferred. If this is not possible a CEA should be undertaken. If a new intervention is being analyzed the possible alternatives and therapeutic value need to be assessed before any other analysis can take place.

The time horizon for the analysis: The time horizon used for an economic evaluation must allow for valid and reliable conclusions about the costs and effects of interventions. The costs and effects of both the old and new interventions have to be measured across the same time horizon.

The cost identification, measurement and valuation: This guideline refers to the *"handleiding voor kostenonderzoek"* (Oostenbrink et al. 2004) for criteria that an economic evaluation should meet concerning four types of costs: Direct costs within the health care sector (medications/treatments, etc.), Direct costs outside the health care sector (travel expenses, waiting times, etc.), Indirect costs within the health care sector (medical costs for life years gained that are unrelated to the treated condition) and Indirect costs outside the health care sector (productivity loss, education, etc.).

The valuation of quality of life and QALY: If quality of life is an important factor in the analysis, evaluation of the *state* of health has to take place to determine the QALY value of an intervention. These valuations can be done in two different ways: through patient surveys, assessing their own quality of life, or by using models. The criteria states that while QALY are preferred, results can also be measured in Life Years Gained (LYG) without the adjustment for the state of health in those life years.

The type of model used to present the results: To be able to support policymaking, a model has to be transparent and preferably based on easily accessible, peer reviewed publications. The model should be as simple as possible, while still including all the information necessary for a complete and thorough economic evaluation. Creating or using a model is almost unavoidable in economic evaluations. A model is needed to translate results to costs and effects, to evaluate the costs and effects beyond the scope of clinical research, and to compare effectiveness and costs between means that have not been compared directly in previous research.

Incremental analysis: This guideline states that the incremental differences in costs and effects of the compared interventions (ICER) must be reported in detail as this shows the cost-effectiveness.

The used discount factor for future effects and costs: When an analysis spans above the width of one year the effects and costs generated after the first year must be discounted. In the primary analysis the costs should be discounted with a constant discount rate of 4 percent. Future effects should be discounted with a constant discount rate of 1,5 percent.

The measurement of uncertainties and risks: Sensitivity analyses should be used to determine to what extent the results of an evaluation depend on assumptions that were made for evaluating the uncertainties and risks of an intervention.

The use of an expert panel: If there is certain data missing from an analysis that is needed to complete the used model a panel of experts can be consulted. The choices made about including members in a panel have to be described in the economic evaluation report and have to be based on methods accepted for scientific research.

What also has to be considered is what the future effects will be. With neonatal screening, most of the beneficial effects take place in the future. Costs also take place in the future, because, as children with a certain condition become older they will also be susceptible to illnesses that they would otherwise not have become old enough to contract.

Aside from these factors there are some challenges named by the *World Health Organization* (WHO). The WHO (2003, p.4) names four challenges of generalization that have risen now that CEA is used more often to assess health interventions. The most important of these is that *"the non-existence of international guidelines for doing CEA has made it difficult to generalize results of economic evaluations, as they were conducted using different sets of guidelines".* Due to the lack of international guidelines for conducting economic evaluations, it is often difficult to generalize and transfer results from one country to another. As long as there is no international standard that provides a minimum of criteria that all economic evaluations have to meet, transferability will always be a problem. This can, and has, lead to a lot of overlap in evaluations even though countries may have similar programs. These challenges were kept in mind while studying the economic evaluations from the systematic review and were used for discussion purposes.

3.4 Criteria and guidelines used in this report

The previous paragraphs show that while there are many different sources of information regarding generalization and transferability, they all show a certain overlap. A combination of the sources was used for this report. When determining the criteria to use, the main concerns were that the theories were regarding (or applicable to) the Netherlands.

With regards to transferability the knock-out criteria from Welte et al. (2004) were chosen to determine if economic evaluations met the basic criteria for transferability. The third Welte criterion pertains to the quality of the evaluations, for which the Dutch CvZ (2006) criteria were used.

For generalizability the factors from the Oostenbrink et al. (2004), Urdahl et al. (2006) and the extra factor of Boulenger et al., level of information detail in study articles were chosen. These sources combined give a good insight into areas that might be problematic for generalization. The factors were grouped into three categories:

- Incidence/Prevalence
- Practice variation (e.g. coverage, participation, test method)
- Cost (screening and treatment)

4 Results

4.1 Introduction: Methodology

Search strategy and inclusion criteria.
A Pubmed/Medline search was conducted starting with the key words cost-effectiveness and newborn screening, then expanded and varied depending on whether the added term generated more results. The search details and final search term are listed in *Appendix A*.

The initial search presented 1079 articles, and after adding the limits of evaluations published after 1995 and written in the English or Dutch language 712 evaluations remained. These 712 titles were downloaded into a database, where the first selection was made. If the titles suggested that the study was about newborn bloodspot screening and its evaluation they were included. In this first selection evaluations that appeared to be on only costs or only effects were included to ensure that no evaluations were left out.

All evaluations related to newborn screening not done by heel prick and evaluations that did not include an evaluation of a program were excluded. This gave a result of 138 articles. Of these articles the second selection was made by reading all abstracts of the first selection. Only evaluations that were actually cost-effectiveness/cost-benefit or cost-utility analyses were included. Evaluations pertaining to carrier screening and neonatal screening other than heel prick (e.g. hearing) were excluded. This resulted in 41 articles. The 41 remaining articles were downloaded or, where not available online, requested and then read in full to determine their relevance for this research. Articles that were purely reviews of CEA and articles limited to a condition that is not screened for in the Netherlands were excluded. This resulted in the inclusion of 20 economic evaluations. To ensure no articles were missed, references of the twenty evaluations were searched and the automatic NCBI search updater was used regularly. The search as described above was also conducted on other search engines. Both additional searches (references and other search engines) did not generate new articles to include in this report. Whenever it was uncertain if a study should be included, supervisors were consulted to ensure no evaluations were wrongfully in- or excluded.

Appraisal of Quality and Transferability

The results were evaluated on two factors, quality and transferability. The appraisal was based on Welte's Knock-out criteria for transferability and the CvZ quality guidelines for economic evaluations as described in chapter 3 and summarized in figure 3.

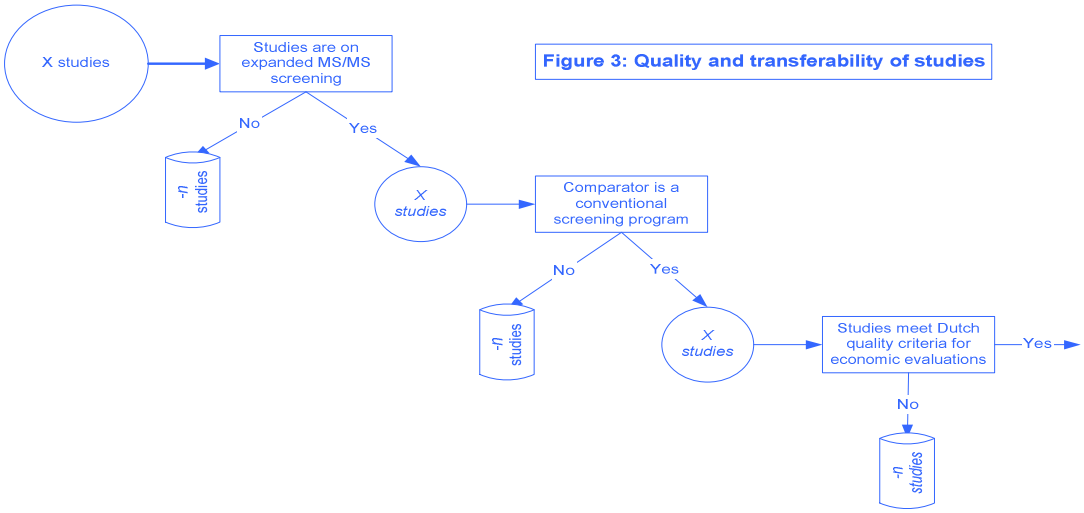


Figure 3: Quality and transferability of studies

The first Welte criterion states that the intervention evaluated has to be comparable to the intervention in the country to which results are to be transferred. The second criterion states that the comparator has to be the existing situation within the country for which the results are to be transferred. The third Welte-criterion pertains to the quality of the evaluations reviewed. In this report the quality criteria used are the Dutch CvZ guidelines for economic evaluations. The results will be described in 4.3.

4.2 Results of the literature search

As mentioned above, twenty evaluations were identified. These evaluations were done on varying conditions, from evaluations of only one condition to evaluations of up to fifty conditions. All the conditions except for CF and SCD are screened for using MS/MS technology. In the following paragraphs the evaluations have been divided into three groups; evaluations on MS/MS screening for one or two conditions, evaluations on CF or SCD screening *(non-MS/MS but included in Dutch program)* and evaluations on expanded MS/MS screening. Table 2 gives a summary of these results.

Evaluations on MS/MS neonatal screening for 1 or 2 conditions

Insinga et al. (2002) evaluated the cost-utility of screening for MCADD through MS/MS compared to no screening. This was done under the assumption that, if screening for MCADD would be cost effective, screening for additional conditions would be as well. A societal perspective was used, though no indirect costs were included in the analysis. Decision tree modeling was done and univariate and threshold sensitivity analyses were conducted. This led to a result of $41.862 per QALY (best case $6008 per QALY). The study was conducted under the assumption that if MCADD screening was cost-effective, screening for thirteen other fatty acid oxidation and organic acidemia disorders in the screening program would also be cost-effective as the MCADD screening would cover the initial costs of MS/MS testing. It is mentioned that the detection of the thirteen other conditions would bring the cost-effectiveness to $15252/QALY though not how this ratio was calculated. ***Venditti et al. (2002)*** used CEA and CUA to evaluate adding MCADD to an existing MS/MS screening program using a societal perspective. The results were measured for a lifetime horizon of twenty (20) and seventy (70) years to reflect the lowest and highest life expectancies. They used a Markov model to process their data and univariate, multivariate, and threshold sensitivity analysis were used to test the strength of the results. For a 20 year horizon the costs were $11.000 per Life Year Gained (LYG) and $5.600/QALY. For a 70 year horizon the costs were significantly lower at $300/LYG and $100/QALY. Sensitivity analysis showed that the upper limits were $17.100/QALY ($33 800/LYG) for the 20-year horizon and $6900/QALY ($13 000/LYG) for the 70-year horizon. ***Tran et al. (2007)*** used CUA to determine whether MCADD screening with MS/MS was more cost-effective than detection through clinical, symptomatic diagnosis (no screening). A public sector perspective was adopted, with a lifetime horizon set at seventy-seven years. The analysis used a decision tree and besides univariate and multivariate sensitivity analysis, robustness of results was tested using best case/worst case scenarios. This gave a result of CA$ 2.676/QALY with all sensitivity variations falling below a threshold of CA$20,000/QALY. The sensitivity analysis showed that the ICER varied from $1029/QALY (best case) to $11463/QALY (worst case) and was most sensitive to changes in screening cost and specificity. The fourth study done on only MCADD was a study in the Netherlands by ***Van der Hilst et al. (2007)***. This was a CEA done using a Markov model and a societal perspective. Screening for MCADD was compared to screening for the current conditions at the time (PKU, CH, and CAH). To measure risks and uncertainties a Monte Carlo simulation was done, along with univariate sensitivity analysis. This gave a result of $1.653/LYG. The sensitivity analysis yielded an ICER of $14,839-$4345/LYG.

There was one study done on screening for only CAH. ***Yoo et al. (2008)*** evaluated the cost-effectiveness of screening for CAH compared to no screening. The assumption was made that each avoided infant death would save thirty-two life years. They used a decision tree and the perspective of the health care sector. The robustness of their results was tested using probabilistic sensitivity analysis, resulting in a cost per Life Year Saved (LYS) of $30.900 (best case), $2.866.200 (worst case) and a base-case ICER of $292.000/LYS. Mentioned in the study is that if more (indirect) benefits are taken into the equation the cost-effectiveness would improve greatly.

One study was conducted to examine the cost-benefit of screening for galactosemia (GAL) compared to no screening, the current practice at the time. This CBA was conducted in the Philippines by ***Padilla et al. (2003).*** It is unclear what method of testing was used as it is not mentioned explicitly. Therefore we can not be sure that MS/MS testing was used. A societal perspective was used, and a sensitivity analysis was done on incidence and discount rate. The result was a C-B ratio of 5.14:1, indicating that screening was not cost-beneficial in the Philippines at the current known incidence rates.

Pandor et al. (2004) conducted a CEA using data from a systematic review of all available literature. They compared the cost-effectiveness of screening for PKU with and without additional conditions using MS/MS testing to screening with conventional technologies. A health care provider perspective was used, and a Markov model was developed. Almost all data was obtained from existing evaluations. While implementing MS/MS testing for only PKU was cost-effective, adding MCADD made it cost-saving. Adding more conditions to the PKU-MCADD combination was still cost-effective but would cease to be cost-saving. The results were a mean of 59 LYG and £23.312 saved when combining PKU and MCADD screening with MS/MS. Several Monte Carlo simulations were done, and the results mentioned above are the mean of the data produced through those simulations. ***Pandor et al. (2006)*** generated similar results to the evaluation mentioned above with a similar CEA. The results were an incremental cost of –£17,298 (–£129,174, £66,434) for 57.3 (28.0, 91.4) LYG on a cohort of 100.000 newborns when screening for PKU and MCADD with MS/MS.
Geelhoed et al. (2005) compared screening for PKU and CH to no screening. Using a public sector perspective a CEA was conducted using a mathematical model. A univariate sensitivity analysis was performed. The results yielded an annual savings of AU$2.879.776.

Evaluations on neonatal screening for CF and SCD

As stated before, CF and SCD can not be screened for using MS/MS technology. As SCD is in the expanded Dutch program, and there is currently a pilot in the Netherlands to determine whether CF should be added to the expanded screening program, the cost-effectiveness of CF and SCD screening is still relevant for this report.
There are different ways to screen for CF, as the two evaluations below will show. ***Simpson et al. (2005)*** evaluated the cost-effectiveness and cost-utility of screening for CF against the current screening of only PKU and CH. The screening method they evaluated was IRT (*ImmunoReactiveTrypsin*) + DNA testing. They used a public sector perspective and a Markov model for calculations. Univariate and multivariate sensitivity analysis were done to test the results; a cost-effectiveness of: £6864 per QALY (£7.474-£6.532/QALY) and £5387 per case diagnosed. There was also a Dutch cost-effectiveness study done by ***van den Akker et al. (2006)*** on CF screening, and what would be the best strategy for testing. The study compared no screening to IRT+IRT, IRT+DNA, IRT+DNA+IRT and IRT+DNA+DGGE testing. A societal perspective was used, though it was unclear what indirect costs were included, as no details for these costs were mentioned in the article. A decision model was used, with univariate and probabilistic sensitivity analysis to review the robustness of the results. For the four strategies the results were (compared to no screening): €24800/LYG, €38300/LYG, €39800/LYG and €330000/LYG, making the IRT+IRT the most preferred and cost-effective testing strategy. The difference in results can be explained partially by the confirmatory sweat test and partially by the perspective. Van den Akker et al. maintained a sweat test cost of almost double what Simpson et al. used, and while van den Akker used a societal perspective including costs such as genetic counseling, Simpson et al. excluded societal costs.

Two of the evaluations found through the review were solely on the cost-effectiveness of screening for SCD, focused on a difference between universal and targeted screening. ***Gessner et al. (1996)*** compared no screening to four different screening strategies[19]; targeted screening with complete follow-up (TSCF), universal screening with complete follow-up (USCF), targeted screening with selective follow-up (TSSF), and universal screening with selective follow-up (USSF). The perspective of the public sector was adopted, and modeling was done using a decision tree. The used time horizon was

[19] Universal screening means testing all newborns while targeted screening would involve testing only newborns racially identified as black, with selective follow-up meaning follow-up only for those newborns that have clinically significant traits.

45 years for costs (and savings), and 1.75 years for outcomes under the assumption that 1.75 is the age at which the condition would have been diagnosed clinically if no screening had taken place. For that reason, the outcomes were not discounted. Results were subjected to extensive univariate and multivariate sensitivity analyses. The incremental costs per death averted ranged from $206.000(TSSF) to $1.780.000(USCF). Out of the four strategies TSSF was the most cost-effective. It was mentioned that if productivity costs would be included the overall cost-effectiveness would improve. The life expectancy used was forty-five years. ***Panepinto et al. (2000)*** also evaluated screening for SCD, and compared no screening to targeted or universal screening. A public sector perspective was adopted and modeling took place using a Markov model. A univariate sensitivity analysis was performed. The results were measured in LYS, resulting in $6709/LYS when comparing targeted screening to no screening and $30.760/LYS when comparing universal to targeted screening. Taking all variations in factors into account the sensitivity analysis showed a maximum of $205.000/LYS and a minimum of $8.000/LYS. A mean life expectancy of forty-five years was assumed. The differences in results in the two evaluations can be explained using the sensitivity analyses. These stated, especially in *Panepinto et al.* that the results were highly sensitive to certain environmental factors. Though both evaluations were done in the USA, one was based on data from Alaska, the other on data from Denver.

Evaluations on expanded MS/MS neonatal screening

Pollit et al. (1997) evaluated the current screening practice against screening with MS/MS using a cost-utility analysis. The study does not give an overall cost-utility-ratio but states based on the economic evaluation results that MS/MS testing would be cost-effective, possibly cost-saving and recommends pilot evaluations.

Schoen et al. (2002) evaluated screening for MSUD, MCADD, other fatty acid disorders, GA-I, MMA, urea cycle disorders and HCY compared to no screening. No modeling was done, the perspective was that of the health care sector and univariate sensitivity analysis was used to evaluate the results. These results were $5.827 ($11.419-$736) per QALY.

Autti-Rämö et al. (2005) examined screening for CAH, MCADD, LCHADD, PKU and GA-I with MS/MS versus the current screening of only CH with cord blood. No specific model was used and a public sector perspective was adopted for the cost-utility analysis. A best and worst practice scenario was calculated, resulting in a cost-effectiveness of €5520-€25560/QALY.

Carroll et al. (2005) looked at screening for PKU, CAH, CH, BIOT, MSUD, GAL, HCY and MCADD and evaluated the cost-utility of screening separately for each condition and combining screening for all conditions using MS/MS. Both scenarios were compared to no screening at all. A societal perspective was used, though indirect costs were not included. A decision tree was used to model the data and the results were tested for uncertainties using univariate, multivariate and threshold sensitivity analysis. Screening for all conditions except for GAL and CAH proved to be cost saving when screened with MS/MS or separately. With a threshold of $20.000/QALY gained only testing for GAL and CAH was, besides not being cost-saving, also not cost-effective. At the base-case scenario testing with MS/MS was the most cost-effective alternative, being cost-saving, and included all aforementioned conditions. The sensitivity analysis yielded results of $310.56/QALY for conventional screening and $4838.71/QALY for screening with MS/MS as the worst scenario.

Feuchtbaum et al. (2006) conducted CEA, CUA and CBA to determine whether screening for all MS/MS detectable disorders was cost-effective compared to screening for only PKU, the current practice. No explicit model was used and neither costs nor benefits were discounted. A third party payer perspective was used and univariate sensitivity analysis was performed on the results. The results were a CBR 1:9.32 ($4.34-$11.67), $132.000/case detected and $708.000/LYS. In the base-case analysis

949 QALY were gained at a savings of $1628/QALY with the MS/MS program. Under the worst case scenario MS/MS was not cost saving but cost $14.922/QALY.

Cipriano et al. (2007) looked at including up to twenty-one conditions in the national screening program. They first gathered data on the cost-effectiveness of screening for each of the conditions separately and compared that to screening using MS/MS. Following these results a scenario was calculated in which PKU would carry all the costs of setup and equipment. This was followed by adding more conditions up to a point where adding the 15th condition would have a marginal cost-effectiveness of more than $300.000/LYG. The screening would cease to be cost-effective mostly because the added condition has a highly expensive treatment program with a low amount of LYG. Cipriano et al. used a third party payer perspective, a decision tree model and univariate sensitivity analysis. The results showed that screening would be effective for PKU plus fourteen conditions. According to Cipriano et al. screening would be cost-effective for PKU, MMA, MSUD, HMG, PPA, VLCADD, GA-I, IVA, MCADD, 3-MCC, LCHADD and four different types of carnitine disorders at a cost per LYG of CA$68.346. The sensitivity analysis on all factors showed an ICER of $41.300 (best case) to $2.042.500 (worst case).

The most recent study, conducted by **Norman et al (2009)** evaluates the cost-effectiveness of MS/MS screening. They used a health sector perspective and two control groups to compare results. A wide range of MS/MS detectable conditions was included. The outcomes were measured on a time horizon of four years due to lack of scientific data on the outcomes after that age. For all patients that were still alive after four years a life expectancy of 66.2 years was used as the horizon. The results were a cost-effectiveness of AU$10.779/LYG making MS/MS screening cost-effective. Sensitivity analysis showed that, when looking at all factors, cost-effectiveness varied between AU$7.969/LYG and A$58.036/LYG.

4.3 Quality and Transferability

In the previous paragraph all twenty evaluations from the literature search were listed. While these evaluations are all about neonatal bloodspot screening, not all are suitable for transferability and not all meet the Dutch quality standards. Figure 2 (4.1.2) showed the criteria against which the measure evaluations would be measured. This figure will now be used to evaluate the evaluations found in the literature search.

Transferability knock-out criteria

To determine whether the evaluations were transferable they were subjected to the three knock out criteria of Welte et al (2004) to determine which evaluations are definitively unsuited for transferability.

The three criteria to meet for this report were:

a. That the evaluated intervention was an expanded program using MS/MS testing
b. That the comparator used was a conventional screening program as this was the situation in the Netherlands
c. That the evaluations met the criteria of the CvZ guidelines for economic evaluations in the Netherlands.

Criteria *a* states that the evaluation has to be on expanded neonatal screening *programs*. The results show that seven evaluations match the first criterion. So, according to criterion *a*, thirteen evaluations were knocked-out. Criterion *b* states that the comparator has to be a conventional (existing) screening program. Three evaluations did not meet this criterion, having used no screening program as the comparator. For the four evaluations that met the first two criteria, the third criterion will be described in 4.3.2.

Author/ Criteria	*Publication Year*	*Societal perspective*	*Intervention*	*Comparator*	*Method*	*Time horizon*	*Measure*	*Use of Model*	*Discount factor*	*Result*
Gessner et al	1996	no	SCD	no screening & 4 diff. strategies	CEA	Lifetime on costs only	DA/MRA	Yes	none (b), 5% (c.)	$206000-1780000/DA
Pollitt et al	1997	no	Wide range MS/MS	Current program	CUA	Lifetime	LYS	Yes	6	£31-8339/LYS
Panepinto et al	2000	no	SCD	2 diff. strategies	CEA	Lifetime	LYS	Yes	3	$6709-30760/LYS
Insinga et al	2002	yes	MCADD	no screening	CUA	Lifetime	QALY	Yes	3	$41862/QALY
Schoen et al	2002	no	Wide range MS/MS	no screening	CUA	Lifetime	QALY	No	3	$5827/QALY
Venditti et al	2003	yes	MCADD	Current program	CUA	Lifetime	LYG/QALY	Yes	3	$300-11000/LYG, $100-5600/QALY
Padilla et al	2003	yes	GALT	no screening	CBA	Lifetime	Cost avoided	No	7	5,14:1
Pandor et al	2004	yes	PKU& MCADD	Current program	CEA	Lifetime	LYG	Yes	6	-£395/LYG
Carroll et al	2005	Yes	Wide range MS/MS	no screening	CUA	Lifetime	QALY	Yes	3	MS/MS cost-saving compared to no and conventional testing
Simpson et al	2005	no	CF	Current program	CUA	Lifetime	QALY	Yes	2 (b), 6(c.)	£6864/QALY
Autti-Rämö et al	2005	yes	Wide range MS/S	Current program	CUA	Lifetime	QALY	Yes	5	€5520-25560/QALY
Geelhoed et al	2005	no	PKU & CH	no screening	CEA	Lifetime	DA	Yes	5	AU$2,9 million saved
Feuchtbaum et al	2006	no	Wide range MS/MS	no screening	CUA	Lifetime	QALY/ LYS	No	Unknown	-$1628/QALY
Van den Akker et al	2006	yes	CF	no screening	CEA	Lifetime	LYG	Yes	3	€24800-39800/LYG
Pandor et al	2006	no	PKU& MCADD	Current program	CEA	Lifetime	LYG	Yes	1,5 (b), 6(c.)	-£302/LYG
Tran et al	2007	no	MCADD	no screening	CUA	Lifetime	QALY	Yes	3	CA$2676/QALY
Van der Hilst et al	2007	yes	MCADD	Current program	CEA	Lifetime	LYG	Yes	4	$1653/LYG
Cipriano et al	2007	yes	Wide range MS/MS	Current program	CEA	Lifetime	LYG	Yes	3	CA$68346/LYG
Yoo et al	2008	no	CAH	no screening	CEA	Lifetime	LYS	Yes	3	$255700-292000/LYS
Norman et al	2009	no	Wide range MS/MS	Current program	CEA	Lifetime	DA/LYG	Yes	1,5 (b), 6(c.)	AU$10779/LYG

Table 2: Evaluations on expanded neonatal bloodspot screening (using MS/MS testing)

Quality Guidelines

The quality guidelines the evaluations had to meet are the CvZ *Quality guidelines for economic evaluation in the* Netherlands. The study results will be discussed per guideline and are summarized in table 2 *(above)*.

- *the chosen perspective of the evaluation*

Out of the four evaluations that passed through the first two knock-out criteria, two used a societal perspective, though the level of cost data included varied greatly. The other evaluations used a public health or third party payer perspective, excluding societal costs such as time lost, productivity lost and education costs. These costs have to be included in order to have a societal perspective as required for cost-effectiveness evaluations according to the national criterion.

- *the choice which alternative interventions to compare*

Regarding the comparator, the criteria state that this should be the current standard treatment in the country to which the evaluation pertains. All four evaluations met this criterion.

- *the technique used for the analysis*

The type of analysis used varied between evaluations. One of the evaluations conducted a CUA (using QALY), the others conducted a CEA (using LYG). While both are valued analyses, the fact that results were measured in different units will make pooling those results more difficult. However, both methods are acceptable for this criterion.

- *the time horizon for the analysis*

With regards to time horizon, the criterion states that, besides being relevant for the intervention, the time horizon should be the same for measured costs and measured benefits. As neonatal screening can prevent death the preferred time horizon would be lifetime. All evaluations matched this criterion.

- *the cost identification, measurement and valuation*

The details of costs measured varied greatly between evaluations. While some showed how all the costs and benefits were calculated, others gave only a summary, limiting possibilities for transferability.

- *the valuation of quality of life and QALYs*

The results were given in costs per: Quality Adjusted Life Years (QALY), Life Years Gained (LYG), Life Years Saved (LYS) and Deaths Averted (DA). Three evaluations measured costs in LYG, one study measured costs in QALY.

- *the type of model used to present the results*

The next criterion pertains to the use of models. The criterion says that the use of a model is unavoidable to translate the results into costs and benefits, and to compare these results. All remaining evaluations used a model, though Autti-Rämö et al. did not specify details of that model. The most frequently used models were decision trees and Markov models.

- *Incremental analysis*

All four evaluations gave an incremental cost-effectiveness ratio (ICER)

- *the used discount rate for future costs and effects*

Regarding the use of a discount factor the remaining evaluations all used discounting. The criterion states that future costs should be discounted at 4% and future benefits at 1,5%. One study used 3%, one study used 6%, one discounted costs and benefits at 5% and the fourth study discounted costs at 6% and benefits at 1,5%. The problem with the use of the same discounting for costs and benefits is that, regarding the criterion, the benefits are discounted at a rate that is too high, making it difficult to generalize results when the undiscounted costs and benefits are not reported. However, this can be positive, because it would mean that the actual cost-effectiveness ratio is more favorable than the study results suggest.

- *the measurement of uncertainties and risks*

To measure uncertainties and risks all evaluations performed sensitivity analysis on some or all of the factors. Therefore all evaluations (though to a different extent) match this criterion.

- *the use of experts*

The last criterion refers to the use of experts if data is uncertain or incomplete. All four evaluations used expert opinion to verify data that was unclear.

To summarize, out of the four evaluations that passed the first two Welte knock-out criteria, only two meet all Dutch quality guidelines. The other two *(Pollitt et al. and Norman et al.)* only failed to meet one of the criteria, as they did not use a societal perspective. While there is general consensus about the merit of a societal perspective, most economic evaluations still do not include societal costs and benefits. The above mentioned factors are meant to be guidelines, not explicit criteria. As such, meeting all other guidelines and criteria, all four evaluations will be used in the following analysis.

4.4 Incremental cost effectiveness results converted to 2008 euros[20]

From the twenty evaluations originally included, only four met (almost) all quality guidelines and transferability criteria set for this report. For these evaluations the results were calculated into 2008 euros to better be able to compare them[21] *(table 4)*. The evaluations with an asterix (*) did not use a societal perspective. When societal costs and benefits are excluded, the ICER tends to be less favorable. This will be kept in mind during the analysis.

Author	Year	Country	2008 Euros	Unit	included disorders
Autti-Ramo et al	2005	Finland	5800	QALY	CAH, MCADD, LCHADD, PKU, GA-I
Norman et al*	2009	Australia	6700	LYS	aminoacidurias, urea cycle disorders, organic acidurias, MCADD, other fatty acid oxidation defects
Pollitt et al*	1997	UK	7600	LYS	PKU, CH, TYR-I, HCY, MSUD, Urea cycle disorders, GAL, CF, CAH, Duchenne, BIOT, MPA, PPA, IVA, branched chain acyl-CoA metabolism defects, MCADD, LCHADD, GA-I
Cipriano et al	2007	Canada	51100	LYS	PKU, MMA, MSUD,HMG,PPA,VLCADD,GA-I, IVA,MCADD,3-MCC,LCHADD,4x Carnitine

Table 4: Converted study results from evaluations. **did not use societal perspective*

[20] All cost conversions by Paul van Gils (RIVM-VTV)
[21] All results, unless reported otherwise, reflect the ICER with regards to a pre-existing, non-MS/MS screening program.

Table 4 shows that three of the remaining evaluations show favorable results towards the cost-effectiveness of their evaluated programs[22]. However, these results were measured in different units (LYG/QALY) and vary from €5800/QALY to €51100/LYS for expanded MS/MS testing programs with a median of €7150/LYS. The median of all (17) evaluations that used QALY or LYS/LYG was €6900/LYS (-€610/LYG to €230100/LYS[23]). The differences in results can be caused by several factors; the number and type of conditions screened for, the incidence of the conditions, the number of tests done annually per MS/MS machine, the number of laboratories that need an MS/MS machine, and so on. It is difficult to say what factors are most responsible for the differences, as not all evaluations included the same costs and benefits, and the level of detail also varied between evaluations. While almost all evaluations stated that screening was very likely to be cost-effective, the evidence for these claims differed greatly. Some evaluations gave detailed descriptions of included costs and benefits, and other evaluations gave only grouped costs (e.g. screening or organization). Also, not all evaluations mention the exact conditions screened for, giving instead the group of conditions such as urea cycle disorders or fatty acid disorders. This suggests that even when these evaluations match the basic transferability criteria on all other aspects, it will be difficult to generalize (or transfer) the results. If detailed unit costs are given there are more possibilities to "substitute" data from other countries and convert results.

The outlier, Cipriano et al. can be explained by a number of factors. In this evaluation, costs included a second MS/MS analysis for all positive screens *before* contacting the patient, confirming the diagnosis with other technologies before making a final diagnosis and giving all patients three months of treatment and follow-up while awaiting confirmation of the results. While the other evaluations also include the last aspect (to a lesser degree), the extra costs of confirmation tests with other technologies would have a tremendous impact on cost-effectiveness as the other evaluations did not include this. Also, Cipriano et al. assumed a higher percentage of false-positives *(at CA$300 per false-positive diagnosis)* than the other evaluations. While the follow-up costs for false-positives are equal to those made for true-positives, they lack the balance of lifetime benefits. So, if more false-positives were assumed the higher cost of 51100/LYS can also be attributed to that. This higher number of assumed false-positives might also be explained through the included conditions. Some of the conditions Cipriano et al. included in their calculations were not included in the Dutch program because the GR stated that they lacked a suitable test (*Neonatale Screening, 2005*). The less specific a test is, the higher the chance for false-positives, which leads to a less favorable cost-effectiveness.

4.5 Additional factors that influence transferability

Besides the quality and transferability criteria used in the previous chapter, other factors play an important role when considering the transfer of cost-effectiveness results. These include differences in incidence of conditions, variations in costs, and practice variation such as participation and test method (Oostenbrink et al.; Urdahl et al.; Boulenger et al.).

Incidence is a very important factor for transferability. If the incidence of a condition is high, more cases will be detected. If more cases are detected (e.g. 1:1.000) the benefits are higher than if few cases (e.g. 1:100.000) are detected. Incidence levels vary between diseases, but also per country as a whole. If more conditions are screened for in a country, the overall incidence of newborns that have one of those conditions increases too, therefore (in theory) also increasing the cost-effectiveness of screening.

[22] *Cipriano et al is the only study showing an ICER that is higher than the generally used threshold of >€20.000/LYG or QALY.*
[23] This was the result of Yoo et al. The closest result was €51100/LYS (Cipriano et al.)

Regarding practice variation, when **participation** is lower, the probability of detecting a child with a condition decreases. The costs of screening will be lower because fewer tests are done, but the benefits will also be less because the chance will increase of children with conditions not being detected through screening.
Another aspect of practice variation is the **variation of conditions** included in the screening program. Different conditions will yield different cost-effectiveness results. For example, if a condition is included for which there is no suitable test, the cost-effectiveness will be a lot less favorable. A last factor to consider regards the **costs of screening and treatment.** In a recent study by Norman et al (2008) a number of the evaluations included in this review were set out and compared regarding costs of screening and treatment as used in the original evaluations. It stated that the costs differ according to location (country), cost components and also the transparency of the calculations. It showed that the costs of the screening test varied between $10-20 per screen with a few at <$10 per screen. This difference may be explained through the included costs, as some evaluations included all costs (organization, set-up, labor) and others only included the costs of expanding a pre-existing MS/MS program. While the evaluations included in Norman et al. (2008) were not the same as those included here, the overall trend of costs is similar. Regarding treatment costs there are more variations. This can be explained with the fact that the evaluations did not include the same conditions, and some treatments would be more expensive. Also, the evaluations using a societal perspective will have higher costs (e.g. education) but those should weigh out against higher benefits (e.g. gained productivity).

Of the four evaluations included in table 4, one was conducted in the UK*, one in Canada, one in Finland and one in Australia*. All these countries have a screening participation rate of almost 100 percent and national coverage. With regards to incidence it is hard to compare the evaluations. In Finland for example, PKU incidence is so low that screening is not cost-effective, even though it is cost-effective in most other countries. In Canada the incidence differs per state, and for certain types of conditions only grouped information is available (e.g. hemoglobinopathies) without mention of the incidence for the separate conditions. For MS/MS testing the cost of screening (per sample) varied from €14.94 in Autti-Rämö et al. to €6.78 in Pollitt et al. to €2.82 in Cipriano et al. and €1.53 in Norman et al. In the original currencies, Autti-Rämö et al gave a cost of €14.23, the other three a cost of 3-4 of their own currency. The differences between the Finnish study and the others is easily explained, because Finland had to include set-up costs as their current screening is done using naval cord blood and not dried blood spots. Regarding treatment costs the differences in conditions included in each evaluation are too great to accurately compare the cost of treatment. Also, not all evaluations mention the treatment costs per condition. Aside from these factors, many of the lifetime treatment costs are based on assumptions, and therefore will be different for every evaluation.

4.6 Summary

In this chapter the results of the literature review were listed and the evaluations were submitted to criteria for quality and transferability. These results showed that while much information can be found, not all this information is relevant and the quality of results differs greatly. After subjecting the twenty economic evaluations found in the systematic review to Welte's three knock-out criteria, four evaluations meet the basic transferability criteria with only two evaluations meeting *all* quality guidelines. Figure 4 shows a summary of the number of evaluations that were knocked-out per criterion.

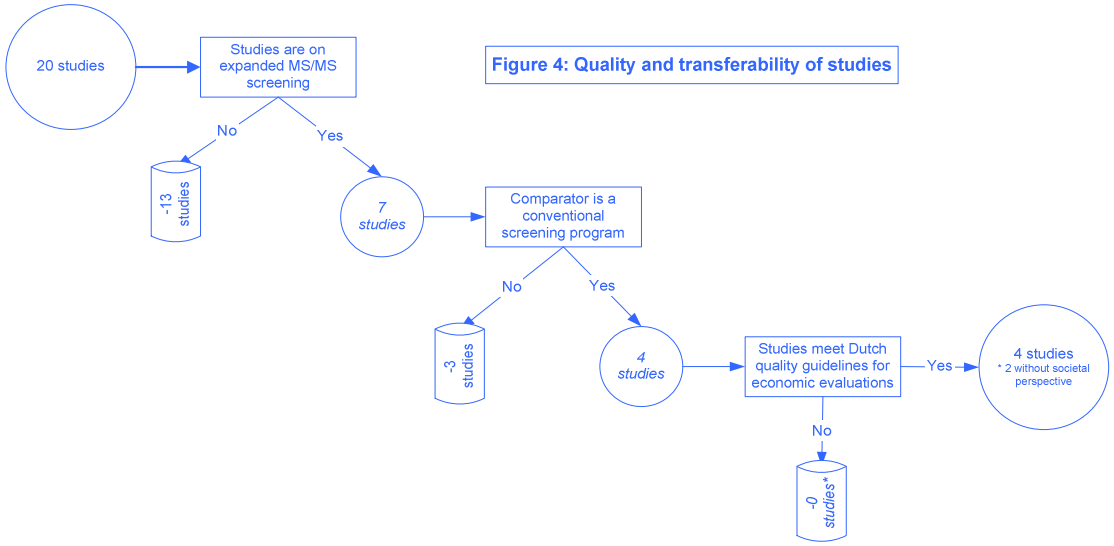


Figure 4: Quality and transferability of studies

Conclusion

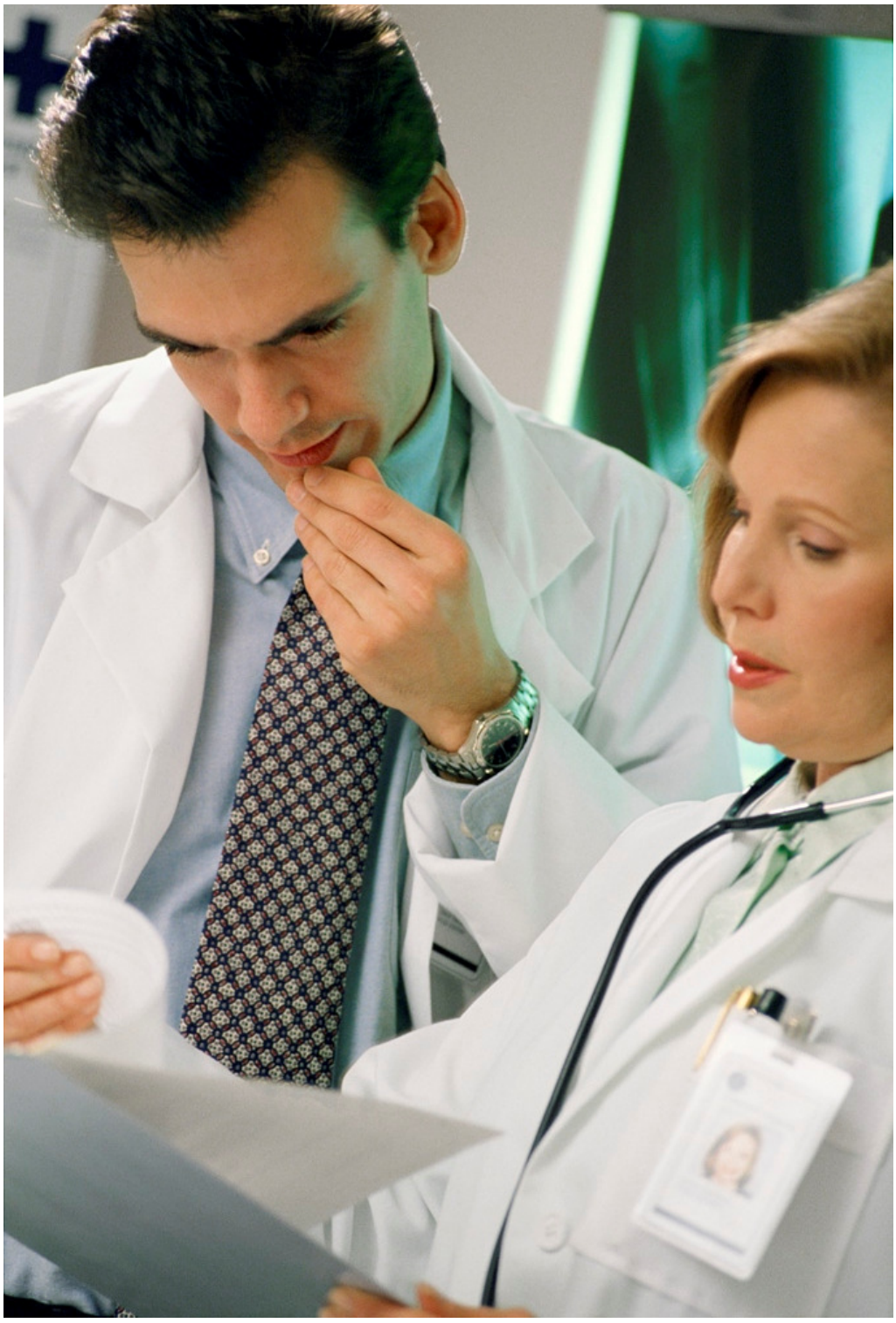
What is known about the cost-effectiveness of neonatal heel prick screening programs worldwide, and (how) can this information be used for decision making about the cost-effectiveness of the expanded neonatal heel prick program in the Netherlands?

To answer the first part of the research question a general background study on (inter)national heel prick programs was followed by a systematic literature review to collect all cost-effectiveness evaluations done on neonatal heel prick screening using MS/MS technology (or CF screening). This review yielded twenty evaluations that could aid in answering the question above. Out of these evaluations, four[24] survived the transferability knock-out criteria. These evaluations vary in results but generally agree that expanded neonatal bloodspot screening is cost-effective compared to conventional screening. The median of all evaluations using QALY or LYG/LYS was €6900/LYS, and for the four evaluations that passed the transferability criteria a median of €7150/LYS was found. Many of the variations in cost-effectiveness were due to differences in populations, conditions screened for and different incidences of conditions. In Finland for example, screening for only PKU was not cost-effective at all as the incidence was so low that screening might not even find one case each year.

For the second part of the research question we can state that while the evaluations differ in many ways, all (*including knocked-out evaluations)* arrive at approximately the same level of cost-effectiveness. If the threshold of €20.000/QALY (§3.1) is maintained, only one evaluation *(of an entire program)* had a result that was not cost-effective *(Cipriano et al. as discussed in §4.4)*. As the results show screening to be cost-effective in many varying circumstances, there is no reason to assume that the Dutch program would yield significantly different results. There is no reason, based on international economic evaluations, to believe that the Dutch program is not cost-effective. Therefore, there is no urgency to conducting a new cost-effectiveness analysis for the expanded neonatal heel prick program in the Netherlands.

[24] Two out of these four evaluations did not use a societal perspective

Discussion



The results present several points of discussion. First of all, this research was conducted to be an aid in decision-making on the necessity of conducting a new CEA for the expanded Dutch heel prick program. This means that by itself it can not give a conclusive answer as to whether or not that CEA should be conducted.

Objectivity and Conflicting interests

Whenever research is done, the possibility of biased results has to be kept as small as possible. An important factor is that the researchers are independent, with no financial relationships that might influence the end result. Table 5 shows the employment and funding information (where available) for all twenty evaluations from the literature review. Most evaluations were conducted primarily by universities, with the aid of government agencies, relevant hospital departments or both.

Author	Authors employed by	Funding	Objectivity
Carroll et al	University	Government funded research agency	Yes
Cipriano et al	University	Government funded research agency	Yes
Feuchtbaum et al	Government	Government funded research agency	Yes
Geelhoed et al	University	No specific project funding	Yes
Norman et al	University and Hospital	Government funded research agency	Yes
Pandor et al	University and Government	Government funded research agency	Yes
Pollitt et al	University and Hospital	Government funded research agency	Yes
Simpson et al	University and Government	Not stated	Yes
Tran et al	University and Government	Government funded research agency	Yes
Van den Akker et al	Research institute and Hospital	Foundation	Yes
Van der Hilst et al	University and Hospital	Government funded research agency	Yes
Autti-Rämö et al	University and Hospital	Government funded research agency	Unknown
Gessner et al	Government	Not stated	Unknown
Insinga et al	University and Government	Graduate traineeship	Unknown
Padilla et al	University, Gov. and Hospitals	Not stated	Unknown
Panepinto et al	University	Not stated	Unknown
Schoen et al	Medical center	Non governmental organization	Unknown
Venditti et al	University and Hospital	University	Unknown
Yoo et al	University and Government	Not stated	Unknown

Table 5: Researchers and funding

Eleven evaluations either stated that the views of the authors were not necessarily those of the funding agency, or more specific that the authors had no financial relationships relevant to the evaluation. The other nine did not mention objectivity in their evaluations.

For five evaluations no information on funding could be found, the other evaluations all received either continuous funding for research projects, or grants from government funded agencies. The funding agencies mentioned in the evaluations maintain strict policies on conflicting financial interests[25]. Those policies, combined with evaluations done mostly by universities, give the impression that the results of the evaluations included in this report are objective findings of the researchers. Also, in many cases the evaluations made conservative assumptions, suggesting that if there was a bias it will most likely show the results as less favorable than they really are.

Assumptions

A general limitation of cost-effectiveness evaluations with a lifetime horizon is the necessity of assumptions. Even though some information can be gathered on lifetime costs and benefits of the heel prick screening through treating pediatricians, specialists, homecare workers and insurance companies, this information is rarely in one place and even more rarely complete. The cause of this is a lack of long-term follow-up after screening (Norman et al, 2008). This makes it unavoidable for researchers to

[25] e.g. http://grants.nih.gov/grants/guide/notice-files/not95-179.html

make calculated assumptions as to the overall costs and *(especially treatment)* effects used in economic evaluations. These assumptions, no matter how scientifically sound, are still assumptions and as such introduce a bias into study results. This bias can be positive or negative, and can be leveled out somewhat through sensitivity analysis, but never completely eliminated. When the assumptions the evaluations are based on vary more, it also becomes more difficult to compare results. Also, as many evaluations use conservative assumptions, it is quite possible that the real cost-effectiveness is more favorable than that mentioned in the evaluation results.

Sensitivity of Results

Out of the four remaining evaluations the sensitivity analyses were studied to determine which factors were most sensitive to changes. The sensitivity of the results pertains to their robustness if factors of the equation change. A high sensitivity means that if there is even a small change in the value of a factor, it will affect the overall cost-effectiveness. The evaluations all showed that one of the most sensitive factors was the incidence of the conditions. Besides incidence, the cost and specificity of the MS/MS test were named as very sensitive. This means that if the equipment, testing or labor costs go up, the number of false-positives increases, or the incidence lowers, the cost-effectiveness of MS/MS screening will be less favorable.

Chance-findings

Another point of discussion is chance-findings. Chance-findings are conditions (or carriers of conditions) that are found through screening, though they are not specifically screened for.

Examples of this are carriers of SCD or CF, because the information about carriers is automatically generated when testing for the condition takes place. A carrier is a person who is not sick, but carries the gene for a certain condition. If a carrier wants to have a child with another carrier the chances of that child having the disease increase exponentially. Therefore, the knowledge of carrier status can possibly influence decisions about having *(additional)* children, though so far no evidence of this has been presented *(Workshop PNS conference April 3rd 2009 on SCD Carriers)*. However, it is possible that carrier knowledge could have a psychological impact and possibly diminish the quality of life of the carrier. That would suggest that the overall (indirect) costs of the screening could increase, while the effects remain the same, having a negative overall effect on cost-effectiveness.

Chance-findings can also have a positive effect, for example when conditions are found through screening that are not specifically screened for. The screening costs remain the same, the treatment costs increase, but over a lifetime the benefits of early detection will most likely outweigh those higher initial treatment costs.

Transferring results

Although the results of this report show that there is no urgency to conducting a (full) new economic evaluation, to gain more insight into the actual cost-effectiveness of the program it might be considered to conduct transfers of the results from the four remaining evaluations to fit the Dutch program. In that way, without conducting a full, new evaluation, it is possible to give a more accurate figure of the cost-effectiveness of the Dutch expanded program. Entering Dutch data into the existing study models of the evaluations that passed all *three* knock-out criteria can give a good indication to the actual cost-effectiveness of the Dutch program. By contacting the original authors it may be possible to introduce existing Dutch data to recalculate the models without having to create a new model. In doing this, two factors must be kept in mind. The first is that not all cost and effect information is readily available in the Netherlands, the second that there is a different modus operandi so used assumptions might have to be adjusted.

Monitoring and Evaluation

The Dutch neonatal heel prick program is monitored yearly by TNO. To gain more insight into the actual costs and benefits of the program it is advisable to include costs and benefits in the yearly monitoring. If all (health) care givers make note of the costs made for treatments and the health benefits detected in the children, it is possible to build a database with this information. Having the data stored in a central database can be of great aid if, in time, a new economic evaluation is conducted, as it would mean that the separate data would not have to be collected first. The data can also be of aid in the monitoring of the program because it would be easier to detect changes in costs and/or benefits. In order to effectively monitor costs and effects, a list would have to be made of all the costs and benefits that are desirable for inclusion in Dutch cost-effectiveness evaluations.

Number of conditions in screening package

All evaluations agree that screening for PKU only, with conventional testing methods, was cost-effective (with the exception of Finland (CH instead of PKU)). Evaluations that compared no PKU screening to conventional PKU screening *and* MS/MS PKU screening showed, that in most cases, MS/MS testing was more expensive, without many additional benefits. However, when adding just one more condition to the screening package, MS/MS testing became cost-effective compared to conventional screening. When looking at the evaluations that compared MS/MS testing to no screening at all, some evaluations even concluded that MS/MS screening was cost-saving! This does not mean that adding more conditions always means that the program becomes more cost-effective.

Regarding the cost-effectiveness of further expanding already expanded newborn blood spot screening programs, some evaluations mention that increasing the number of conditions is only cost-effective to a point. At this point adding conditions could increase the cost of treatment with no significant increase in health benefits. Screening for conditions that have no suitable test, no suitable treatment or extremely low incidence might actually cause a program to cease being cost-effective or at least reduce cost-effectiveness. If a condition is (partially) untreatable, many costs are made without the benefits that could justify those costs. Take for example Cipriano et al. who included (partially) untreatable conditions in the program selection. The higher ICER of that evaluation could very well be ascribed to the inclusion of untreatable conditions and conditions lacking a suitable test.

While screening for these rare or untreatable conditions would be cost-ineffective, there are also emotional benefits that should be considered. For parents whose child died shortly after birth, knowing sooner why their child passed away could provide emotional benefits that are significant even when the health benefits are not. Also, knowing sooner what is “wrong” with their child could relieve much worry for parents and have an impact on further family planning. The issue with this balance is that it is extremely difficult to decide how much emotional benefit is be needed to justify the costs made, and of course, how that benefit can be measured.

Therefore, while there is no urgency to conducting a new cost-effectiveness evaluation, the impact on cost-effectiveness should always be assessed before new conditions are added to the program.

List of abbreviations

3-MCC	3-Methylcrotonyl-CoA Carboxylase deficiency
AFDPHE	Association Française pour le Dépistage et la Prévention des Handicaps de l'Enfant
AWBZ	Algemene wet bijzondere ziektekosten (exceptional medical expenses act)
BIOT	Biotinidase deficiency
CAH (AGS)	Congenital adrenal hyperplasia
CBA	Cost-benefit Analysis
CD	Cases diagnosed/detected
CDC	Centre for Disease Control
CEA	Cost-effectiveness Analysis
CF	Cystic Fibrosis
CH	Congenital Hyperthyroidism
CvB	Center for Population Screening
CUA	Cost-utility Analysis
CvZ	College Voor Zorgverzekeringen
DA	Deaths averted
DNA	Deoxyribonucleic Acid
EE	Economic Evaluation
FAO	Food and Agriculture Organization
GA-I	Glutaric Acidemia type I
GALT	Galactosemia
GR	GezondheidsRaad (Health Council of the Netherlands)
HCS	Holocarboxylase synthetase
HCY	Homocystinuria
HMG	HMG-CoA-lyase Deficiency
ICER	Incremental Cost-effectiveness Ratio
ICUR	Incremental Cost-utility Ratio
IRT	ImmunoReactiveTrypsin
IVA	Isovaleric Acidemia
LCHADD	Long Chain Acyl-CoA Dehydrogense Deficiency
LIS	Laboratory Infectious disease and Screening
LNV	(Ministry of) Agriculture, Nature and Food quality
LYG	Life Years Gained
LYS	Life Years Saved
MCADD	MediumChain Acyl-CoA Dehydrogenase Deficiency
MMA	MethylMalonic Acidemia
MoH	Ministry of Health
MRA	Mental Retardation Avoided
MS/MS	Tandem mass spectrometry
MSUD	Maple Syrup Urine Disease
NBS	Newborn Bloodspot Screening
NHG	Nederlands Huisartsen Genootschap
NNSGRC	National Newborn Screening and Genetics Resource Center
NVK	Nederlandse Vereniging voor Kindergeneeskunde
PKU	Phenylketonuria
PNHS	Program committee Neonatal Heel prick Screening
PPA	ProPionic Acidemia
QA	Quality Assurance
QALY	Quality Adjusted Life Years
RIVM	Rijksinstituut Volksgezondheid en Milieu
RVZ	Raad Volksgezondheid Zorg
SCD	Sickle Cell Disease
TNO-QoL	Applied Scientific Research-Quality of Life
TSCF	Targeted Screening Complete Follow-up
TSUF	Targeted Screening Selective Follow-up
TYR-I	Tyrosinemias type I
UK	United Kingdom
UNEP	United Nations Environment Program
USA	United States of America
USCF	Universal Screening Complete Follow-up
USSF	Universal Screening Selective Follow-up
VLCHADD	Very Long Chain Acyl-CoA Dehydrogense Deficiency
VROM	(Ministry of) Social building, Regional Planning, and Environment administration
VWS	(Ministry of) Health, Welfare and Sports
WHO	World Health Organization

References

Akker van den-van Marle ME, Dankert HM, Verkerk PH, Dankert-Roelse JE. *Cost-effectiveness of 4 neonatal screening strategies for cystic fibrosis.* Pediatrics 118(3): 896-905. 2006.

Andermann A, Blancquaert I, Beauchamp S, Déry V. *Revisting Wilson and Junger in the genomic age: a review of screening criteria over the past 40 years.* WHO bulletin; volume 86, number 4:241-320. 2008

Antonanzas F., Rodriguez-Ibeas R, Juarez C, Hutter F, Lorente R, Pinillos M. *Transferability indices for health economic evaluations: methods and applications.* Eur. J. Health Econ. June 2008

Autti-Ramo I, Makela M, Sintonen H, Koskinen H, Laajalahti L, Halila R, Kaariainen H, Lapatto R, Nanto-Salonen K, Pulkki K, Renlund M, Salo M, Tyni T. *Expanding screening for rare metabolic disease in the newborn: an analysis of costs, effect and ethical consequences for decision-making in Finland.* Acta Paediatrica 94(8): 1126-1136. 2005.

Berg M van den, Wit GA de, Vijgen SMC, Busch MCM, Schuit AJ. *Kosteneffectiviteit van preventie: kansen voor het Nederlandse volksgezondheidsbeleid.* Ned.Tijdschr.Geneeskd.152:1329-1334. 2008.

Borrajo GJC. *Newborn Screening in Latin America at the Beginning of the 21st Century.* J. Inherit. Metab. Dis. 30:466-481. 2007.

Boulenger S, Nixon J, Drummond M, Ulmann P, Rice, S, Pouvourville G de. *Can economic evaluations be made more transferable?* Eur. J. of Health Econ. 6:334-346. 2005

Carroll AE, Downs SM. *Comprehensive cost-utility analysis of newborn screening strategies.* Pediatrics 117(5 Part 2): S287-S295. 2006.

Cipriano LE, Rupar CA, Zaric GS. *The cost-effectiveness of expanding newborn screening for up to 21 inherited metabolic disorders using tandem mass spectrometry: results from a decision-analytic model.* Value in Health 10(2): 83-97. 2007.

College voor Zorgverzekeringen. *Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie.* Versie april 2006. CVZ, Diemen. 2006.

CvB. *Draaiboek Neonatale Hielprik screening.* Versie 6.0. RIVM. 2007.

CvB. *Rapportage meerjarenplanning kosteneffectiviteitstudies.* RIVM, Bilthoven. 2008

Donaldson C, Currie G, Mitton G. *Cost-effectiveness analysis in health care: contraindications.* BMJ 325:861-894. 2002

Drummond MF, Jefferson TO. *Guidelines for authors and peer reviewers of economic submissions to the BMJ.* BMJ 313:275-283. 1996.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes. Third edition*: Oxford: Oxford University Press. 2005.

Feenstra TL, Baal PHM van, Wit GA de, Polder JJ, Hollander AEM de. *Kosteneffectiviteitsanalyses over de keten van preventie, cure en care.* RIVM report 270091003. Bilthoven. 2006.

Feuchtbaum L, Cunningham G. *Economic evaluation of tandem mass spectrometry screening in California*. Pediatrics 117(5 Part 2): S280- S286. 2006.

Geelhoed EA, Lewis B, Hounsome D, O'Leary P. *Economic evaluation of neonatal screening for phenylketonuria and congenital hypothyroidism.* Journal of Paediatrics and Child Health 41(11): 575-579. 2005

Gessner BD, Teutsch SM, Shaffer PA. *A cost-effectiveness evaluation of new-born hemoglobino-pathy screening from the perspective of state health care systems*. Early Human Development 45(3): 257-275. 1996

Health Council of the Netherlands. *Neonatal Screening.* The Hague: Health Council of the Netherlands, 2005; publication no. 2005/11.

Hilst CS van der, Derks TG, Reijngoud DJ, Smit GP, TenVergert EM. *Cost-effectiveness of neonatal screening for medium chain acyl-CoA dehydrogenase deficiency: the homogeneous population of the Netherlands*. Journal of Pediatrics 151: 115-120. 2007.

Holland WW, Stewart S, Masseria C. *Policy Brief; Screening in Europe.* World Health Organization on behalf of European Observatory on Health Systems and Policies. 2006.

Insinga RP, Laessig RH, Hoffman GL. *Newborn screening with tandem mass spectrometry: examining its cost-effectiveness in the Wisconsin newborn screening panel.* Journal of Pediatrics 141(4): 524-531. 2002.

Johns B, Baltussen R, Hutubessy R. *Programme costs in the economic evaluation of health interventions.* Biomedcentral: Cost-effectiveness and Resource Allocation 2003, 1:1.

Lapré RM, Rutten FFH, Schut FT. *Algemene economie van de gezondheidszorg.* 2e druk: Elsevier Gezondheidszorg, Maarssen. 2001.

Loeber JG. *Neonatal screening in Europe; the situation in 2004.* J. Inherit. Metab. Dis. 30:430-438. 2007.

Nixon J, Rice S, Drummond M, Boulenger S, Ulmann P, Pouvourville G de. *Guidelines for completing the EURONHEED transferability information checklists*. Eur. J. Health Econ. 2008 1618-7598

Norman R, et al. *International perspectives on the cost-effectiveness of tandem mass spectrometry for rare metabolic conditions*. Health Policy (2008), doi:10.1016/j.healthpol.2008.08.003

Norman R, Haas M, Chaplin M, Joy P, Wilcken B. *Economic Evaluation of Tandem Mass Spectrometry Newborn Screening in Australia.* Pediatrics 123:451-457. 2009.

Oostenbrink JB, Bouwmans CAM, Koopmanschap MA, Rutten FFH. *Handleiding voor kosten onderzoek, methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg.* College voor zorgverzekeringen, Diemen. 2004

Padilla CD, Dans LF, Estrada SC, Tamondong MR, Laceste JJ, Bernal R M. *Cost-benefit analysis of newborn screening for galactosemia in the Philippines.* Southeast Asian Journal of Tropical Medicine and Public Health 34(Supplement 3): 215-220. 2003.

Padilla CD, Therrell BL. *Newborn Screening in the Asia Pacific region.* J. Inherit. Metab. Dis. 30:490-506. 2007.

Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S. *Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review.* Health Technology Assessment 2004: 134

Pandor A, Eastham J, Chilcott J, Paisley S, Beverley C. *Economics of tandem mass spectrometry screening of neonatal inherited disorders.* International Journal of Technology Assessment in Health Care 22(3): 321-326. 2006.

Panepinto JA, Magid D, Rewers MJ, Lane PA. *Universal versus targeted screening of infants for sickle cell disease: a cost-effectiveness analysis.* Journal of Pediatrics 136(2): 201-208. 2000.

Pollitt RJ, Green A, McCabe CJ, *et al.* Neonatal screening for inborn errors of metabolism: cost, yield and outcome. *Health Technol Assessment* 1997; **1**(7).

Pomp M, Brouwer W, Rutten F. *QALY-tijd; Nieuwe medische technologie, kosteneffectiviteit en richtlijnen.* no 152. Centraal Planbureau. 2007.

Rijks Instituut voor Volksgezondheid en Milieu *(RIVM)*
http://www.rivm.nl/en/aboutrivm/organization/public_health/cvb/index.jsp
http://www.rivm.nl/en/aboutrivm/Missie_RIVM.jsp

Schoen EJ, Baker JC, Colby CJ, To TT. *Cost-benefit analysis of universal tandem mass spectrometry for newborn screening.* Pediatrics 110(4): 781-786. 2002.

Simpson N, Anderson R, Sassi F, Pitman A, Lewis P, Tu K, Lannin H. *The cost-effectiveness of neonatal screening for cystic fibrosis: an analysis of alternative scenarios using a decision model.* Cost-effectiveness and Resource Allocation 3:8. 2005

Therrell BL, Johnson A, Williams D. *Status of Newborn Screening Programs in the United States.* Pediatrics 117: 212-252. 2006

Therrell BL, Adams J, *Newborn screening in North America.* J. Inherit. Metab. Dis. 30:447-465. 2007.

Tran K, Banerjee S, Li H, Noorani HZ, Mensinkai S, Dooley K. *Clinical efficacy and cost-effectiveness of newborn screening for medium chain acyl-CoA dehydrogenase deficiency using tandem mass spectrometry.* Clinical Biochemistry 40(3-4): 235-241. 2007.

Urdahl H, Manca M, Sculpher MJ. *Assessing generalizability in Model-Based Economic Evaluation Studies.* Pharmacoeconomics 24(12):1181-1197. 2006

Venditti LN, Venditti CP, Berry GT, Kaplan PB, Kaye EM, Glick H, Stanley CA. *Newborn screening by tandem mass spectrometry for medium-chain acyl-CoA dehydrogenase deficiency: a cost-effectiveness analysis.* Pediatrics 112(5): 1005-1015. 2003.

Watson MS, Lloyd-Puryear M, Mann MY, Rinaldo P, Howell RR. *Newborn Screening: Toward a Uniform Screening Panel and System.* ACMG: Federal register volume 70:44. 2005.

Welte R, Feenstra T, Jager H, Leidl R. *A decision Chart for Assessing and Improving the Transferability of Economic Evaluation Results Between Countries.* Pharmacoeconomics 22(13):857-876. 2004.

Wit GA de, Verweij A, Baal PHM van, Vijgen SMC, Berg, M van den, Busch MCM, Barnhoorn MJM, Schuit AJ. *Economic evaluation of prevention: further evidence.* RIVM report 270091004, Bilthoven. 2007

Yoo BK, Grosse SD. *The Cost-effectiveness of Screening Newborns for Congenital Adrenal Hyperplasia.* Public Health Genomics 12:67-72. 2009.

Appendix A: Systematic PubMed search

First search criteria	Second search criteria	Number of results
Cost-effectiveness	Newborn screening	581
	~~OR heel prick~~	0 new→removed
	OR blood spot	591
OR cost-benefit		638
OR economic evaluation		762
OR cost-utility		772
	OR tandem mass spectrometry	794
	OR Ms/Ms	797
	OR PKU	805
	OR congenital Hypothyroidism	807
	OR CAH	818
	OR Cystic Fibrosis	1005
	~~OR MCAD(D)~~	0 new→removed
	OR thalassemia	1079
	~~Or galactosemia~~	0 new→removed
TOTAL without limits		**1079**
Limits		
	Publishing date after 1995	777
	Language English(703), Dutch(9)	712
TOTAL with limits		**712**
Further selection		
	Based on title	138
	Based on abstract	41
	Based on content	20
Final Selection		**20**

Final search term including limitations:

(cost-effectiveness OR cost-benefit OR economic evaluation OR cost-utility) AND (newborn screening OR blood spot OR tandem mass spectrometry OR Ms/Ms OR PKU OR congenital hypothyroidism OR CAH OR cystic fibrosis OR thalassemia) AND (("1995"[PDat] : "3000"[PDat]) AND (English[lang] OR Dutch[lang]))

Last search date: April 1st 2009.

Appendix B: Conditions screened for in the Netherlands

Information from the Dutch "hielprik spiekboekje".

CAH

CAH stands for Congenital Adrenal Hyperplasia. It is a rare heredic disorder that affects the hormone production in the adrenal glands. The adrenal glands produce three types of hormones; Cortisol (stresshormone), Aldosterone (salthormone) and Androgen (malehormones). In CAH patients the adrenal glands produce too little cortisol and aldosterone, and too much androgen. Newborns with this condition are at high risk for saltwasting and dehydration. At birth, girls will, to different measures, show male-like genitals. Early detection can prevent dehydration in the newborn and further genital deforming. In the Netherlands 15-20 children are born annually with this condition.

CH

CH stands for Congenital Hypothyroidism. It is a non-hereditary group of conditions in which the thyroid gland does not produce enough of the hormone thyroxine (T4). CH is usually permanent. Lack of T4 at a young age has a negative influence on cerebral development with a risk of permanent mental and physical inhibitions. If CH is diagnosed early and treatment started, the development of the body and brain should be normal and mental retardation can be avoided. In the Netherlands approximately 70 to 90 children are born annually with this condition.

PKU

PKU, phenylketonuria, affects the ability to properly use protein. When a normal person eats foods containing protein, enzymes break down the protein into separate amino acids as building blocks for body growth and repair. In PKU, one of the enzymes does not function properly, the one needed to convert phenylalanine into another amino acid, tyrosine. As a result, phenylalanine accumulates in the blood and other parts of the body. The excess phenylalanine prevents the brain from growing and developing normally. It also causes other problems such as skin rash, excessive restlessness, irritable behavior and a musty body odor. If the diet is started early enough and closely followed, the child's development will be normal in almost all cases. In the Netherlands 10 to 15 children are born with PKU annually.

BIOT

Biotinidase (BIOT) deficiency is a hereditary metabolism disorder where the body does not produce enough biotin (vitamin H). This deficiency can cause skin problems, epileptic seizures, boldness, slow development and muscular problems. Once symptoms occur, they are usually irreversible even with a therapeutic diet. In the Netherlands approximately 2 children are born with BIOT each year.

GAL

Galactosemia is a hereditary disorder where galactose (part of milksugar and lactose) is insufficiently broken down. Lactose is in breast milk and many other food products for newborns. If not detected early, GAL can cause severe jaundice, infection, stare and death. Even with treatment slow development, speech disorders and reduced fertility can still occur. In the Netherlands approximately six children are born with GAL annually.

GA-I

Glutaryl-acidura type I (GA-I) is a hereditary metabolism disorder that causes an insufficient breakdown of the amino acids lysine and tryptophan. If not treated, brain damage can occur in the newborn. In the Netherlands about 1 child is born with this condition each year.

HMG-CoA-lyase deficiency

HMG is a hereditary metabolism disorder where the amino acid leucine is not broken down properly causing a shortage of energy. Untreated HMG can cause vomiting, weakness, loss of consciousness, neurological impairment and reduced development. This condition is extremely rare, with about 100 patients worldwide. The specific incidence in the Netherlands is unknown.

HCS (MCD)

Multiple CoA carboxylase deficiency (HCSD/MCD) is a hereditary metabolism disorder where protein from food can not be converted into usable substances. This can lead to dehydration, loss of consciousness, skin problems and boldness. The exact incidence in the Netherlands is unknown but is thought to be one in 87.000 children.

HCY

Homocistinuria is a hereditary condition where the body is unable to break down the amino acid homocystine. Untreated HCY can cause eye problems, damage to blood vessels, thrombosis, pulmonary embolisms and serious development problems. In the Netherlands one to two children are born with this condition annually.

IVA

Isovaleric academia is a hereditary metabolism disorder where the body is unable to break down the amino acid leucine. Untreated IVA can cause vomiting, loss of consciousness, reduced development and death. In the Netherlands approximately three children are born with this condition each year.

LCHADD

Long chain 3-hydroxyacyl-CoA dehydrogenase deficiency (LCHADD) is a hereditary disorder where long chain fatty acids can not be used as an energy source. Problems can occur when a newborn eats few food or no food at all, e.g. with fever, sleeping the night without a feeding, vomiting or diarrhea. LCHADD can cause sleepiness, loss of consciousness and (hart)muscle problems. The incidence of LCHADD in the Netherlands is unknown.

MSUD

Maple Syrup Urine Disease (MSUD is a hereditary disorder where the body is unable to fully break down the amino acids leucine, isoleucine and valine. The newborns smell sweet. If treatment is not started immediately vomiting, loss of consciousness, developmental delays and death can occur. Approximately one child with MSUD is born in the Netherlands each year.

MCADD

Medium chain acyl-CoA dehydrogenase deficiency (MCADD) is a hereditary disorder where medium chain fatty acids can not be used as an energy source. Problems can occur when a newborn eats few food or no food at all, e.g. with fever, sleeping the night without a feeding, vomiting or diarrhea. MCADD can cause sleepiness and loss of consciousness. About 15 to 17 children in the Netherlands are born with this disorder annually.

3-MCC

3-methylcrotonyl CoA carboxylase (3MCC) is a hereditary metabolism disorder where certain proteins and the amino-acid leucine are not fully broken down. Untreated it can cause seizures, developmental delays and loss of consciousness. The exact incidence in the Netherlands is unknown, but is probably 1 in 50.000 newborns.

SCD (thalassemia)

Sickle cell disease is a hereditary hemoglobine disorder. Low oxygen density can cause red blood cell defects causing small hair vessels to clog. This results in severe bone aches and organ infarctions. There is also an elevated risk of infections because the spleen does not function properly. Elevated break down of blood can also cause anemia. In the Netherlands 40 to 60 children are born with (a variation of) this disorder each year.

TYR-I

Tyrosinemia type I (TYR-I) is a hereditary disorder that causes problems with the break down of the amino acid tyrosine. Untreated TYR-I can cause liver failure, kidney problems, nerve conditions, liver cancer and death. In the Netherlands approximately two children are born with this condition annually.

VLCHADD

Very long chain acyl-CoA dehydrogenase (VLCADD) is a hereditary disorder where very long chain fatty acids can not be used as an energy source. Problems can occur when a newborn eats few food or no food at all, e.g. with fever, sleeping the night without a feeding, vomiting or diarrhea. VCHADD can cause sleepiness, loss of consciousness and (hart)muscle problems. The incidence of VLCHADD in the Netherlands is unknown.

CF

Cystic Fibrosis (CF) is a hereditary disorder where the mucous in the body becomes thicker and tougher than is normally the case. These changes cause problems in the airways and in the gastrointestinal tract. Early treatment may help reduce these problems. CF is one of the most widespread hereditary disorders in the Netherlands with 50 children found each year.