ECONOMIC EVALUATION OF NEWBORN HEARING SCREENING PROCEDURES*

Carmen Herrero and Juan D. Moreno-Ternero**

WP-AD 2002-06

Correspondence to C. Herrero: Departamento de Fundamentos del Análisis Económico. Universidad de Alicante. 03071 Alicante. Spain. Phone: (34)965903618. Fax (34)965903685. E-mail: carmen.herrero@ua.es

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
First Edition June 2002.
Depósito Legal: V-2066-2001

---

* This paper originates from an observation made by Manuela López-Azorín and Ignacio Ortuño-Ortín, who suggested the interest of analyzing this problem. We are also grateful to Han Bleichrodt, Mike Drummond, Dunia López-Pintado, Lars Osterdal, Jaume Puig and the audience at the Department of Public Health, University of Alicante, for helpful comments and suggestions. Part of this research was conducted while the second author was visiting the Institute of Economics at Copenhagen University, thanks to a Marie Curie Fellowship. Financial support from the Ministerio de Ciencia y Tecnología, under project BEC2001-0535, and from the Generalitat Valenciana under project GV01-371, is gratefully acknowledged.

** C. Herrero: University of Alicante and Ivie. J.D. Moreno-Ternero: University of Alicante.

ECONOMIC EVALUATION OF NEWBORN HEARING SCREENING PROCEDURES

Carmen Herrero and Juan D. Moreno-Ternero

A B S T R A C T

In the past few years, there has been a growing interest in the medical literature about the study of the different existing screening procedures to detect hearing impairment in infants and young children. However, concerning their economic evaluation, there are some important aspects, such as indirect costs, which are not considered by that literature. Here, we present an economic evaluation of these screening procedures, using utility theory, to measure benefits of a health care program, i.e. a cost-utility analysis. The analysis is presented from different points of view, depending on the cost we would like to compute. If we only consider direct costs, then targeted procedures, based on high risk criteria are preferred. On the other hand, if indirect costs, such as special education, and disability allowances were computed, then cost-utility analysis would advocate for the implementation of universal screening procedures.

Keywords: Economic Evaluation; Cost Utility Analysis; Cost-Sensitivity Ratios; Newborn Hearing Screening.

1 Introduction

Recently, the *Consellería de Sanitat de la Generalitat Valenciana*, the health care authority in the *Comunidad Valenciana*, a Spanish region, has edited a protocol for the early detection of hearing impairment in newborns at this region [8]. The protocol states the purpose, in the long run, of the implementation of a universal newborn hearing screening ("UNHS" hereafter) program to detect the impairments mentioned above. As a first step, they decided to implement the program to all infants who were born with a risk factor in the main hospitals. When the screening protocol is only applied to those infants who were born under a risk factor the program is said to be targeted newborn hearing screening ("TNHS" hereafter). The election among the two possibilities (universal or targeted) was a source of debate and it generated a wide literature about it, in the last decade. However, this literature mainly focussed on the medical aspects of the problem and it ignored considerably the economic viewpoint. The objective of this paper is to make a cost-utility analysis by using the QALY index to measure benefits in health.

1.1 Background

At the beginning of 1993, the National Institutes of Health Consensus Development Conference on Early Identification of Hearing Impairment, addressed the advantages of early (consequences of late) identification of hearing impairment [32]. Moreover, it concluded that universal screening should be implemented for all infants within the first 3 months of life.

The hearing impairment satisfy all the medical requirements to impose a prevention program, by terms of a universal screening. On the one hand, it is a serious handicap impairment, whose absence of early diagnostic will cause problems on language acquisition [30], [42]. Significant hearing loss interferes with the development of speech perception abilities needed for later language learning, e.g., meaningful language at the word, phrase, and sentence levels. These impairments in communication skills can lead to poor academic performance (especially reading), and ultimately, to limitations in career opportunities [18]. Moreover, it has a considerable prevalence, comparing with other illnesses for which there exist universal newborn screening programs.[1] Finally, there are reliable and sure screening methods, like *Otoacoustic Emissions (OAE)* and *Auditory Brainstem Responses (ABR)*, with high levels of sensitivity and specificity, and the most important thing, there exists treatment (e.g., cochlear implants).

At the moment of the conference there was not any UNHS implemented. There were only some programs based on the high risk criteria.[2] The conference pointed out two main disadvantages of these criteria. The principal one is that approximately 50 percent of newborns with congenital hearing deficits are not in the high risk groups and are missed by the screen. Moreover, children who

[1] For instance, in Navarra, a Spanish region, there exists a universal newborn screening program to detect phenylcetunoria, whose prevalence is of $\frac{1}{3500}$ [18].

[2] For a complete list of risk factors of hearing impairment, see, for instance, [1] and [29].

are not born in larger hospitals may not be routinely identified as being at risk. Related to this, we have the second main disadvantage. The average age of identification in the United States remained closed to 3 years in 1993, which coincides precisely with the end of the most important period for language and speech development.

Since the conclusions of the conference were published, it seemed difficult to argument against the recommendation of the UNHS from a medical point of view. This line of argument was followed by a large number of papers (see, for instance [1], [3], [10], [13], [26], [36], [37], [41], and [42]). However, one year later, Bess and Paradise [4], published a paper which argued strongly against the UNHS. There were also subsequent papers which criticized UNHS (see, for instance, [34]).

In the year 2000, the Joint Committee on Infant Hearing, published the principles and guidelines for early detection and intervention programs [22]. In this statement, based on pilot experiences which took place in some states of the USA, (e.g. [3], [26], [37], [40]) they pointed out again the necessity of a UNHS. As a matter of choice, the committee preferred a 2-stage procedure, in which the first stage is the OAE method followed by ABR, as a second stage for those infants who failed the first one.

The purpose of our work is to make a cost-utility analysis of the two criteria. The two main sources of the data are two previous works. On the one hand, we have Kemper and Downs's paper [23], which collects information about strategies, cost, and prevalence, reflected in previous works, and it presents a cost-effectiveness analysis of both criteria, the universal one and the targeted one. This analysis, neither makes use of utility theory to measure the benefits of each alternative, nor takes into account the indirect costs, like special education and disability allowances. On the other hand, Kezirian et al. [24] also present a cost-effectiveness analysis, but they do about four different universal screening techniques which are currently in use in most of USA hospitals, which actually carry out a universal procedure for early hearing impairment detection. Furthermore, they fix a different threshold for hearing impairment. In fact, Kezirian et al., consider as a hearing impairment, bilateral or unilateral hearing loss above 30 dB, while Kemper and Downs do for ≥ 40 dB hearing loss bilaterally. As a consequence, we will have different prevalence data, depending on the threshold we would like to consider.

Finally, we use data about a similar experience which took place for the first time in Spain at the end of the 90's [18]. As in Kemper and Downs [23], the threshold for the existence of hearing impairment under this third protocol was fixed in 40 dB.

The rest of the paper is organized as follows. First, we analyze some properties of the newborn hearing screening techniques mentioned above, and the subsequent protocols, obtained as a result of their combinations. Secondly, we focus on the different kind of costs associated to a congenital hearing impairment, and on how to measure health benefits in terms of utility. Finally, we present the results and conclusions.

2 Screening strategies

Screening is traditionally defined as testing a population of asymptomatic individuals to identify unrecognized early disease or precursors of disease. The screening procedure itself does not diagnose illness. Those who test positive are sent on for further evaluation by a subsequent diagnostic test or procedure to determine whether they do in fact have the disease. An implicit assumption underlying the concept of screening is that early detection, before the development of symptoms, will lead to a more favorable prognosis because the treatment begun before the disease becomes clinically manifest and therefore, it will be more effective than later treatment. As we mentioned above, this is an assumption clearly satisfied in the case of screening for hearing impairment.

The validity of a test is defined as the ability of a test to distinguish between who has a disease and who has not. Validity has two components: *sensitivity* and *specificity*. The *sensitivity* of the test is defined as the ability of the test to identify correctly those who *have* the disease. The *specificity* of the test is defined as the ability of the test to identify correctly those who *do not have* the disease (see Table 1 for further details). Obviously it would be desirable to have a screening test that was both highly sensitive and highly specific. Usually that is not possible, and there is generally a trade-off between the sensitivity and specificity of a given screening test.

Screening is often carried out in stages, as for example in the issue we have concerned. In such a case, a less expensive, less invasive, or less uncomfortable test is carried out first, and those who screen positive on this test are recalled for further testing with a more expensive or more invasive test, which may have greater sensitivity and specificity. It is hoped that bringing back for further testing those who screen positive will reduce the problem of *false positives.*

On the question of early hearing impairment detection, there are two strategies that have become available. *Otoacoustic emissions (OAE)* measures sounds that are generated by the cochlea in response to acoustic stimulation [38]. They are thought to represent a reflection of sound waves when sounds are presented to normal ears, and they are not detected in ears affected by the large majority of types of hearing loss in newborns. *Auditory Brainstem response (ABR)* presents sounds to the ear and detects nervous system activity in specific locations of the hearing pathway. Complete ABR testing remains the gold standard for determination of hearing loss, but there is a shorter screening version (S-ABR) that is less expensive and quicker [24]. Automated devices for measurement of OAE and ABR are available, and both of them have reported high sensitivity and specificity levels.

Let us note that there is a lack of reliable data for sensitivity and specificity of screening programs in the literature, specially in two-stages programs. To solve this problem, notice that with the prevalence of congenital hearing loss being relatively low, the specificity of the screening test is well approximated by the share of infants who fail the test. About the sensitivity of a two-stage program, it can be well approximated by the product of the sensitivity at each stage.

In 1999, the *American Academy of Pediatry* [1], pointed out that a UNHS should be the long run objective. Now, it also prescribed that the used methodology should have a false positives rate below 3%, and a null false negatives rate. As a matter of fact, they postponed to choose a specific method, while there were not a method satisfying the above requirements, or at least, specifically better than the other ones.

3 Protocols

The model of decision analysis about the newborn hearing screening is a 2-stage protocol, similar to the one recommended by the National Institutes of Health (NIH) [32] in 1993 (Figures 1 and 2). Universal 2-stage screening protocol, consists of testing all newborns with OAE, as a first stage, following by an ABR test for those who failed the first stage. The second stage used to be carried out before the newborns would leave the hospital due to their accessibility. If such is not the case, we face the problem of having a *return rate* (of babies whose parents decide to come back for the second stage) about 79% [24]. This procedure agrees with what we mentioned above, about the 2-stage screening protocols. A less expensive test is carried out first, and those who screen positive on this test, are recalled for further testing with a more expensive, and efficient, test.

We present two slightly different versions of this universal 2-stage screening, by changing some aspects of the OAE and ABR strategies. On the one hand, we consider automated transient-evoked otoacoustic emissions (TEOAE) as a first stage, followed when indicated by automated auditory brainstem response (ABR) testing [23]. We will refer this protocol as UNHS1 (Figure 1). On the other hand, we consider otoacoustic emissions (OAE) as a first stage,[3] followed when indicated by a shorter screening version of automated auditory brain response testing (S-ABR) [24]. We will refer this protocol as UNHS2 (Figure 2). Targeted screening reserves the two stage screening process for those infants at risk for congenital hearing loss. Therefore, as before, we have two alternative targeted screening procedures, which will be called TNHS1 (Figure 1) and TNHS2 (Figure 2). Every protocol concludes with a diagnostic test for those who failed both stages. The diagnostic test is usually carried out after the baby leaves the hospital, in order to ensure maxima levels of sensitivity and specificity. Thus, we face again the problem of a low return rate.

The analysis is based on a fixed and defined cohort of newborn children in Spain. Some data from similar experiences in Canada and USA will be applied. We assume that all newborns are screened by the 2-stage protocol and only 79% of those who were classified as positives return for the diagnostic test [24]. We also assume that each of these latter ones receive all necessary follow-up testing.

[3] There are two tests that are based on OAE: transient-evoked otoacoustic emissions (TEOAE) and distortion product otoacoustic emissions (DPOAE). For this second protocol, the otoacoustic emissions tests were treated as a single entity because, following Kezirian et al. [24], their cost and validity have been similar to date.

Finally, there is a UNHS currently in practice in another Spanish region, Navarra. In this case, the implemented protocol has three stages. The first stage consists on an OAE test to every newborn at the third day of life, before leaving the nursery. For those who failed it, there will be a second OAE at the fifteenth day of life. Finally, the third stage involves those neonates who failed the second stage to return at the third month, in order to receive a new OAE test. Those who did not pass the third OAE test will go through a diagnostic confirmation by terms of an ABR test. We will refer to this protocol as UNHS3 and to its corresponding one, only referred to those infants who were born with a high risk factor, as TNHS3 (Figure 3).

Table 2 summarizes the data of each screening protocol, including sensitivity, specificity and costs. We can observe that UNHS2 presents the best level of sensitivity, while TNHS1 is the one with the lowest cost.[4]

4 Costs

Cost data are difficult to obtain and standardize. For example, the cost of using an automatic screening device depends on the time over which the device will be depreciated and the estimation of the number of children who will be screened [23]. We therefore evaluate costs based on reports in the literature. Furthermore, there exists controversy in the literature with the cost of detecting high-risk factors in newborns. The range of values provided by it, goes from \$0.50 to \$15 [38], [41]. As we will see later, this cost will play a crucial role in the cost-utility analysis from the perspective of a hospital (when indirect costs are not evaluated).

Table 2 only shows direct costs of each screening test, namely OAE, ABR, diagnostic and the cost of identifying high risk factors. These costs are enough to present an analysis from the point of view of a hospital. By contrast, if we want to present an analysis from other points of view, then we need to compute, not only direct costs, but also indirect ones, such as follow-up and treatment, special education and disability allowances. It seems plausible that if these costs were evaluated, then the conclusions will differ from the case in which they were not evaluated.

A key recommendation of the recent Panel on Cost-Effectiveness in Health and Medicine was to carry the Cost-Effectiveness analyses ("CEA" hereafter) from a societal perspective, including all the costs and consequences of health interventions [39]. Nonetheless, they also recommend that morbidity costs (productivity gains and losses resulting from interventions that reduce morbidity) should not be included among the costs in CEA. The Panel also says that true social costs of lost productivity would be captured only by transfer payments from disability insurance, and it recommends to include all health care costs [39]. Meltzer and Johannesson [28], criticize these two recommendations of the

[4] For the purpose of a sensitivity analysis, confidence intervals, for each screening method-specific parameter values, were given in the table. In the case that the literature did not provide them, the mean value was decreased and increased by 25%.

Panel. On the one hand, they point out that excluding morbidity costs is likely to lead to underestimating the CEA of interventions that enhance productivity by decreasing morbidity [28]. On the other hand, if all health care costs were included, it could lead to a double-counting, provided that people could incorporate personal financial consequences into QALY weights [28]. Thus, according to them, personal health care costs should not be included.[5]

Under our personal opinion that previous criticism to the Panel recommendations only works if health costs are borne by the individual. If health costs are bone by the Health Authority, they should be included. In our case, and in the Spanish (or in general the European) perspective, where all costs are borne by the Public Authority: health care (implants and treatments), education and disability allowances, according to the Panel recommendations, we should consider all sort of expenses (disability allowances can be understood as disability insurance). And this case does not contradict the position in [28]. It would be different, according to the Panel recommendations, if some of the costs were borne privately, as it happens in the US. In such a case, all follow up costs should be excluded.

As a consequence, we will present the results from different viewpoints, depending on the costs we wish to consider. We will observe that results from a "societal perspective" may be altered depending on the costs we introduce. That is an additional evidence in the current debate introduced by Meltzer and Johannesson, about the inconsistencies in the Panel on Cost-Effectiveness in Health and Medicine.

Related to this, we will assume that every newborn with hearing impairment receives a treatment, but they do in different points of time. There are two possibilities. First, since early detection leads to a more effective treatment, as we mentioned in Section 2, we assume that an infant who receives the treatment exactly after being detected by a screening procedure, will need neither special education nor disability allowances. On the other hand, if the impairment was detected without a screening program, then he would receive the treatment after the maximum possible time to avoid some irremediable consequences, like a misleading cognitive development, or speech perception. Consequently, he would need special education and disability allowances.

As a result, for each of the available newborn hearing screening programs, there will be four different costs (per newborn) associated to it, concerning the screening itself, the follow-up and treatment, the special education and the disability allowances. More precisely, given a particular newborn hearing screening program s, we will denote c_s^1 to its direct cost, c_s^2 to the treatment costs associated to it, c_s^3 to its resulting special educational costs and c_s^4 to the associated costs concerning disability allowances.

Now, in an economic evaluation of alternative health care programs, costs must be computed with respect to a reference 'status quo'. In other words,

[5] Meltzer and Johannesson position [28] would be also to compute some other costs, mortality costs. Now, since the health programs that are being analyzed here (newborn hearing screening) do not reduce mortality rates (they only reduce morbidity) those costs do not appear in our case.

the different costs mentioned above must refer, indeed, to the incremental costs associated to the change from the 'status quo' to the new situation, after implementing the program. In the case of our work, we suppose that the 'status quo' is the absence of any early detection program of hearing impairment, but not the absence of any health policy concerning such an impairment, like treatments, special education and disability allowances. As we mentioned above, we assumed that every impaired newborn receives the same treatment, no matter what screening procedure we decided to implement. The only difference is in the precise point of time in which they receive it, which may alter considerably his future consequences. Related to this, we assumed that those individuals whose impairment was not eradicated after the treatment (every impaired newborn, in the case of the 'status quo') receive special education and disability allowances. As a consequence, the overall cost of treatments is reflected in the 'status quo' we are considering, and therefore, the incremental treatment costs of each screening program is null (i.e. $c_s^2 = 0$).[6] Furthermore, it is straightforward to see that $c_s^3 = -(p \cdot se_s \cdot ce)$ and $c_s^4 = -(p \cdot se_s \cdot \rho)$, where se_s denotes the sensitivity of s, ce the incremental cost of special education, ρ the disability allowances, and p the prevalence of hearing loss.[7] Depending on the viewpoint we consider to make our analysis, then c_s, the total costs associated to s, will be either c_s^1 (Hospital viewpoint), or $c_s^1 + c_s^2$ (Health Care Authority viewpoint), or $c_s^1 + c_s^2 + c_s^3$ (Health Care and Education Authorities viewpoint), or $c_s^1 + c_s^2 + c_s^3 + c_s^4$ (Public Sector viewpoint).[8]

In [33] the educational costs for severely hearing-impaired and profoundly hearing-impaired children are evaluated. From this study, savings in educational costs that would result from enabling the profoundly hearing-impaired to function as severely hearing-impaired were determined. More precisely, the discounted difference in education costs associated with a profoundly hearing-impaired child as compared with a severely hearing-impaired over compulsory school years (ages 4-16) was $42850. The improvements in the hearing impairment have been found in children who received a cochlear implant at 4 years and wore hearing aids after it. The same article shows that the discounted value of a pediatric cochlear implant costs over 70 years of life is $109010.[9] We will assume that both amounts are estimations of the mean savings in special education and treatment costs, respectively.

There is another important aspect we may wish to treat in the analysis of

[6] Since some individuals receive the same treatment earlier than others, there might be differences in their costs, caused by the inflation. Notwithstanding, since we assume the same horizon of life to every newborn, there might be differences about the cost of maintenance, as well. For instance, if a newborn receives a cochlear implant two years before than other, then the implant will be less costly, but he will incur in two more years of maintenance costs. We will assume that both differences neutralize each other, which ensures $c_s^2 = 0$.

[7] The negative sign in c_s^3 and c_s^4, refers to the fact that an early detection program produces savings in special education costs and disability allowances, with respect to the status quo.

[8] As one can easily infer from our assumptions, and subsequent discussion, the analysis from a hospital viewpoint coincides with the analysis from a health care authority viewpoint.

[9] The first three years involve the assessment, implantation and rehabilitation, while the remain ones refer to the maintenance.

newborn hearing screening programs. Actually, there are many countries which give disability allowances to people belonging to the deaf community. These subsidies could even be such as those for the disabled people. Obviously, a newborn hearing screening procedure would lead to save part of these public subsidies.

In order to get concrete numbers from the last point, we used the data from the European Union Home Panel, relative to Spain in the year 1994. In that panel, we can observe that the mean of disability allowances is \$4063 (\$1=180 ptas.), which could be an estimation of the net earnings obtained by a deaf individual in a year. If we compute this amount, discounted throughout the working years, we have an estimation of \$343709.

Table 3 summarizes all of these concepts, about the total costs of each procedure, with respect to the 'status quo'. Note that all costs were rounded to the nearest \$.01.

5 Measuring benefits in terms of utility

The importance of economic evaluation of health care programs raises the issue of how to measure, value and incorporate changes in quality of life into the economic evaluation. Cost Utility Analysis ("*CUA*" hereafter) is the most common type of analysis used in health care decision making and it measures the benefits of a program in utility terms. More precisely, it measures changes in the quality of life with the *Quality Adjusted Life Years* ("*QALYs*" hereafter) measure.

The evaluation of health care programs involves both technical and value judgements. The value judgements concern mainly the trade-off between the two important outcomes of such programs: quality of life gained and quantity of life gained. QALYs offer one way of incorporating these two benefits of health care programs into a single index measure: QALYs gained [5]. On the basis of the QALY index, decisions concerning the allocation of resources in the health care sector can be made. The program that should be implemented is the one that offers the largest number of QALYs per dollar or, what is equivalent, the one that has the lowest costs per QALY gained [11]. For an extensive survey of the advantages and disadvantages of this measure see, for instance, [31].

Once we decide to measure benefits of a health care program by terms of utility, with the QALY index, we need to decide how to assign quality weights to each year provided by the program. Measuring preferences for health outcomes is a very time consuming and complex task. A recent alternative that is being widely used is to avoid the measurement task by using one of the pre-scored multi-attribute health status classification systems that exist [12]. Moreover, Weinstein et al. [39], recommend, for a reference case, the use of a generic health measure, which can be applied to a wide set of conditions and diseases. Examples of such measures are the *EuroQol (EQ-5D)* and the *Health Utilities Index* (*HUI*). Both measures assign to each patient a particular health status, among a set of available ones. To do this, the authors assume that the health can be characterized by a set of scores assigned to a finite number of aspects of

the health status. These aspects are referred as dimensions or attributes. In the case of the EQ-5D, there are five dimensions and three levels for each dimension: no problems, some problems and extreme problems. In the case of the HUI, the number of dimensions, and even the number of levels for each attribute can vary (See the Appendix for the details). The EQ-5D was implemented in the 90's and it is often used in the evaluation of health programs, specially in Europe. On the other hand, the HUI measure fits better for pediatric purposes, due to the fact that it is valid for patients over 2 years, while the EQ-5D is only valid over 16 years.

In order to get the total benefits (QALYs) of a particular program, we divide the whole population of newborns in four groups, depending on their prevalence and their test results, i.e., true positives, false positives, false negatives and true negatives. Each of these groups will be characterized by a different health status, provided by each of the above measures we decide to use. The total QALYs gained, associated to a particular program, will be the difference between the aggregation (weighted by the cardinality of each group) of the QALYs associated to each of these four groups, and the number of QALYs in the same population, without implementing any program.

Table 4 shows the estimation of the benefits provided by each of the four protocols we are considering, with each of the two generic measures mentioned above. This was done by reviewing the literature and by asking some experts in the issue (See the Appendix for the details).

Once we have measured the benefits and the costs of each alternative, decisions about the implementation of a screening procedure can be made. If there exists a procedure which is the less expensive and it provides the highest benefits (measured in QALYs gained) then that one should be implemented. In general, decisions are not so trivial and there is a trade-off between the program that minimizes the costs and the one which gives the highest number of QALYs. In such a case, CUA refers to the *Cost-Utility* ratios ("*CU ratios*" hereafter), i.e. the cost per QALY gained. The preferred screening procedure would be the one which shows the lowest CU ratio. In other words, CUA advocates for implementing that screening procedure with the lowest cost per unit of utility gained, or what is equivalent, the program that offers the highest number of QALYs per dollar.

Notwithstanding, let us see that, under an additional assumption, CUA can be reduced to a simpler analysis in which only the costs and the sensitiveness, of each program, matter. The assumption is that the utility of a false positive individual is the same than that of a true negative individual. The reason to do this is the following. Since false positives are correctly identified in a short period of time (undoubtedly, before the treatment starts), the possible utility differences between such an individual and a true negative one, cannot be captured by any dimension of each of the generic measures considered here. Notice that the only difference between two individuals without congenital hearing impairment, but with different test results, lie in the hypothetical depression caused by a wrong positive test. But, it is plausible to think that newborns do not usually get depressed.

Notice that, concerning the measurement of the benefits of each alternative, we only compute the health benefits of the newborns, without paying attention to the parents' utility. One could argue that with this assumption we would be bypassing an important negative externality, noting that the hypothetical anxiety or depression provoked by a false positive test, could be translated to the patient's parents. Nevertheless, recent papers have shown that the effect of these cases on the parental anxiety is not so important (see, for instance, [7] and [35]). Furthermore, it is not less true that if the hypothetical parental anxiety would be considered as a negative externality, then we should also take into account a positive externality caused by the care that the patient obtains from her parents, which could help considerably to her treatment. Without any doubt, this would increase exponentially the difficult task of computing health benefits, and as a result, we will restrict our attention to the patients' utility.

Given a newborn hearing screening procedure s, let us define its *cost-sensitivity ratio* ("*CS ratio*" hereafter), as the ratio between its total cost c_s, and its sensitivity se_s. In other words, if R_s denotes the CS ratio, then $R_s = \frac{c_s}{se_s}$. Notice that c_s depends on the viewpoint from which we are making the analysis. The proof of the following result is shown in the Appendix.[10]

Claim 1 *Under the assumptions mentioned above, the screening procedure that should be implemented, according with a CUA, is the one that shows the lowest CS ratio.*

As a result, and in the case of non-trivial decisions, we simply need to calculate the CS ratios of each alternative to select the best option. This would imply that the recommended program by a CUA would be independent on the generic utility measure we would decide to choose. Nonetheless, the claim only provides ordinal preferences among the set of possible alternative programs. Consequently, in order to get cardinal preferences among them, we will also show the CU ratios, provided by each of the generic measures mentioned above. Furthermore, these CU ratios will also show the necessary cost to gain a QALY in each program, which is an interesting information by itself.

In the QALY model, there exists the possibility of introducing time preferences, by terms of a (constant or variable) discount rate.[11] If such is the case, obviously the CU ratios would be greater than in the general case. The cost per QALY gained would increase as long as the discount rate also does. Obviously, a discount rate would be an uncertain parameter in the model, and a sensitivity

[10] Strictly speaking, the proof of the claim requires that the utility associated to a generic false positive individual would be greater than the one associated to a false negative individual. We did not mentioned explicitly this assumption before, but obviously we did implicitly. Otherwise, it would be unnecessary to implement any early detection program.

[11] When a discount rate r is considered, the number of QALYs of a health stream with a life horizon of T years would be

$$Q = \sum_{t=1}^{T} \left(\frac{1}{1+r} \right)^{t-1} U(x_t).$$

analysis (the main method by which analysts have allowed for uncertainty in economic evaluations [12]) for it, would be required. However, the above claim tells us that the ordinal preferences among the set of available programs do not depend on the discount rate either, which would solve the uncertainty about the election of it. (See the proof of the Claim in the Appendix)

6 Results

As we mentioned above, there are different results in the CUA, depending on the viewpoint we wish to consider or, what is equivalent, depending on the costs we would like to compute.

6.1 CUA from the hospital point of view

In this first case, we simply compute the direct costs of the health care system, namely the cost of each screening procedure and the subsequent definitive diagnostic test. In such a situation, targeted procedures are less expensive programs but the universal procedures are the ones which produce the highest benefits. Therefore, we need to look at the CS-ratios to decide.

Table 5 shows the medical cost-utility (MCU) ratios, and the CS ratios from a hospital viewpoint.[12] It follows from the data in the table, that the targeted procedures are the ones which give better ratios. In particular, TNHS2 is the one which presents a better CS ratio. Therefore, according to a CUA from a hospital viewpoint, TNHS2 should be the one implemented, among the programs considered here. More precisely, from a hospital viewpoint, we would have the following order of preferences among the set of available programs:

$$TNHS2 \succ TNHS3 \succ TNHS1 \succ UNHS1 \succ UNHS3 \succ UNHS2,$$

where $\succ$ means "strictly preferred to".[13]

As we mentioned in Section 5, this conclusion is independent on the generic utility measure chosen and on the discount rate. The sensitivity analysis tells us that the uncertainty with respect to the cost of detecting a high risk factor on a newborn has substantial influence, because of its wide confidence interval. In particular, if it would increase from its mean estimation \$1, to \$6.17, then the universal procedure UNHS1 would lead to a lower CS ratio.[14] The other parameter that has a considerable influence is the sensitivity. Since the confidence intervals of it, for some procedures, are also wide enough, it might alter the preferences within the targeted or universal procedures. However, the main conclusion, namely targeted procedures are preferred to universal ones, is still true for any sensitivity values, within their confidence intervals.

[12] The life horizon of a newborn considered here was 74 years.

[13] Notice that in any case the procedures corresponding to the Spanish protocol (TNHS3 and UNHS3) lie in between the other two possibilities.

[14] Note that the confidence interval for the cost of detecting a high risk factor on a newborn is [\$0.5, \$15] (see [23]). Thus, such an increment, from \$1 to \$6.17, would not be so strange.

For the sake of completeness, we show in Table 5 the MCU ratios from a hospital viewpoint provided by the EQ-5D and HUI measures, without discounting first, and with a discount rate of 0.05. The difference between them is that all of the ratios, provided by the EuroQol measure, are higher than the corresponding ones obtained with the HUI. Thus, it is more expensive to gain a QALY, if we measure benefits with the EQ-5D measure. This is due to the fact that the HUI measure provides higher gains of utility (See the Appendix for the details).

6.2 CUA from the Health Care Department point of view

Now, not only the direct costs of a newborn hearing screening procedure, but also the subsequent follow up and treatments, will be financed from the Health Care Institution budget.[15] Therefore, from this new viewpoint we should also consider these latter costs. However, as we mentioned above, we assume that the treatment costs are the same to each screening method s, and therefore the total costs of each one increase in the same proportion. Moreover, these are also costs included in the 'status quo'. In other words, even in the case that no early detection program exists, every infant with congenital impairment receives a treatment. As a consequence, the costs associated to each of the available screening programs, with respect to the 'status quo' would coincide with the direct costs, and therefore, the CUA from the Health Care Authority point of view, would be exactly the CUA from the Hospital viewpoint.

6.3 CUA from a "Societal Perspective"

Finally, we present another viewpoint in which we also include the remaining costs associated to a newborn hearing screening procedure, i.e. special educational costs and disability allowances.

Under this perspective, the early detection programs would not only provide health benefits with respect to the 'status quo', but also savings in special education costs and disability allowances. More precisely, as we mentioned in Section 4, each screening program s would save an amount (per newborn) of $p \cdot se_s \cdot (ce + \rho)$ ($p \cdot se_s \cdot ce$, when disability allowances are not computed), where se_s denotes the sensitivity of s, ce the incremental cost of special education, ρ the disability allowances, and p the prevalence of hearing loss.

We can observe from Table 3 that UNHS2 is the program with the higher savings, even in the case that we do not include disability allowances. Moreover, notice that the order of programs according to the savings they produce, is exactly the same one that according to the benefits (QALYs gained) they produce, as Table 4 shows.[16] More precisely

$$UNHS2 \succ UNHS3 \succ UNHS1 \succ TNHS2 \succ TNHS3 \succ TNHS1,$$

[15] Recall that we assume a "European Perspective" in which all health costs are borne by the Health Authority.

[16] In both cases, the ranking coincides with the ranking of sensitivity levels. As the proof of the Claim shows, this is independent on the employed utility measure and discount rate.

where $\succ$ means "strictly preferred to". Notice that this is a totally different order to the one provided in Section 6.1. The sensitivity analysis says that such a ranking is still valid when we vary the mean estimations of ce or ρ, within their confidence intervals.[17]

As we pointed out in Section 4, and since the analysis presented here is based on a fixed and defined cohort of newborn children in Spain, where all of these costs are borne by the Public Authority, computing educational costs and disability allowances does not contradict the position in [28], about the inconsistencies in the Panel on Cost-Effectiveness in Health and Medicine. Notwithstanding, if some costs were borne privately, the situation changes. In such a case, all follow up costs should be excluded. As a result, the unique cost to be computed would be the direct cost of each screening program. This framework would coincide with the CUA from a hospital viewpoint, presented in Section 6.1, in which targeted screening programs would be preferred.

7 Discussion

We have presented here some ideas about the current debate concerning the adequacy of an implementation of a newborn hearing screening program. Several comments can be made about it, depending on the point of view from which we wish to make the analysis. In other words, depending on the costs we wish to compute, from each alternative.

As one could expect from the intuition, only evaluating the direct costs, a targeted screening, based on high risk factors detection, would be preferred to a universal one. This conclusion is still true when treatment (cochlear implants) costs are included (Health Care Authority viewpoint). As far as we are concerned about indirect costs, namely special education and future disability allowances, then the conclusion reverses, i.e. a universal procedure would be preferred. These conclusions support the reflections made by Meltzer and Johannesson [27] and [28], about the inconsistencies in the Panel on Cost-Effectiveness in Health and Medicine. Those reflections suggest that accounting for future costs in medical cost-effectiveness analysis may substantially alter both the absolute and relative cost-effectiveness analysis of medical interventions. This is indeed the case of our work. While an evaluation of direct costs advocates for a targeted screening procedure, an estimation of indirect costs, in the case that they were borne by the Public Authority, as it happens in most of the European countries, advocates for a universal screening procedure.

Furthermore, the support of a UNHS holds just considering the indirect costs of a newborn hearing screening program, associated to special education. As far as we introduce other kind of indirect costs, like disability allowances, the support is even stronger. Obviously, this would not be the case in countries,

[17] Furthermore, the conclusion would still be valid when disability allowances were not computed, i.e. $\rho = 0$, and $ce > 32137.5$. In particular, such a condition is fulfilled for the mean value of ce considered here. Since the confidence interval for ce is $[18374, 67326]$ [33], disability allowances need to be computed, in order to ensure the same ranking, for lower values of ce.

where some of these indirect costs are borne by individuals.

Notice that the support for the UNHS, as a result of a cost-utility analysis evaluating indirect costs, reinforces the support of it, that was also presented in some previous works, from different perspectives. For instance, cost-effectiveness analyses (c.f. [23], [24] and [25]) or analyses based on the study of the resulting income distribution of the population (c.f. [17] and [21]).

The results have been obtained by evaluating health benefits in terms of utility, i.e. with the QALY index. More precisely, we used two generic health measures, such as the EuroQol (EQ-5D) and the Health Utilities Index (HUI), to assign quality weights to each year provided by each alternative health care program. Under the plausible assumption of our model, i.e. the absence of utility differences between false positives and true negatives, the ranking among the set of available procedures does not depend neither on the generic measure, nor on the discount rate we wish to consider, which diminishes considerably the 'uncertainty effect' of our problem. Since the time for confirming a false negative is not high enough, the hypothetical anxiety caused by it, is not manifested in the newborn yet. Furthermore, due to the fact that this confirmation must be done before starting any treatment, none of the other dimensions which comprise each of the generic utility measures considered here will vary, between a true negative and a false positive newborn. Therefore, the EQ-5D and the HUI will assign the same quality weights to each of them, as our assumption asserts. The main impact of false positives concerns to the hypothetical anxiety or depression on their parents, but we pointed out that recent works (c.f. [7] and [35]) have diminished the importance of this effect.

Notwithstanding, depending on the employed measure we would like to compute, the amounts provided by each cost-utility ratio, i.e. the cost per QALY gained, and the relationship between the health benefits and the savings of each program, in the societal perspective, may vary. In other words, we have consensus about ordinal preferences among the set of available programs, but a lack of it, if we turn to cardinal preferences among them. Furthermore, we should point out that we considered here the Spanish tariffs estimated for the EQ-5D that are available [2], but that is not the case of the HUI system. The only available scoring formula for the last version of the HUI system, is based on the estimations made from a sample in Canada [15], and therefore, we used those weights, which are the only available ones at the moment. As it can be observed from Table 5, both measures provide significantly different amounts, concerning the ratios. Furthermore, there are also EQ-5D tariffs, estimated from an England survey [9], and these weights also lead us to different ratios (See the Appendix for the details). As a consequence, if we would like to get enough consensus about the cardinal preferences among the set of available programs, the above generic utility measures do not seem to be very appropriate. One possible reason for this, is that they do no fit exactly to our problem, specially due to the fact that some dimensions of each measure (like anxiety or self-care) do not have full sense when they refer to newborns. Thus, a possible line of argument for future research could be to construct a specific utility measure for this problem. Nevertheless, if we accept the assumptions mentioned above, the

conclusions shown in our model, about the ordinal preferences among them, are unobjectionable, even with the use of a more specific utility measure.

We have presented three different protocols in the targeted and universal procedures respectively, as a consequence of the three different sources of data employed here (c.f. [18], [23], and [24]). On the one hand, we presented two different protocols following the NIH recommendation [32], which are currently in use in some hospitals of USA. The difference between both protocols can be due to the fact that Kemper and Downs [23] base on the hypothetical protocol recommended by the NIH [32], while Kezirian et al. [24] do it on a slight different version of this protocol, which is currently implemented in most of the USA hospitals. Furthermore, since Kemper and Downs [23] fix the threshold for serious hearing loss in $\geq$ 40 dB hearing loss bilaterally, while Kezirian et al. [24] do it in $\geq$ 30 dB hearing loss unilaterally, we have different prevalence of hearing loss from both papers. On the other hand, we presented a different protocol which is currently in practice in some hospitals of a Spanish region. This third protocol seems to be in the middle of the above ones, by terms of preferences, but in any case the main conclusion, namely any UNHS is preferred to any TNHS, still holds here, when indirect costs are computed.

Finally, about the treatment costs, we only included here the costs concerning a cochlear implant, because it is the only treatment whose data and effectiveness is well established in the literature. Further research about other alternatives, such as hearing aids, which is currently in use when moderate hearing impairment exists, seems to be necessary, in order to generalize this analysis. Moreover, further medical research and consensus about the protocols used here, seems to be necessary too, in order to confirm crucial data for their economic evaluation, such as sensitivity.

Let us also note that we have supposed that all treatments are successful in true positives newborns. In other words, once they receive the treatment, they do not need neither special education, nor disability allowances. However, this is not the case of the false negatives. These latter individuals receive a treatment but, since it was not an early treatment, they will also need special education, and future disability allowances. Both of these assumptions could have been relaxed in the following way. We could have supposed that everybody receives the treatment, but there is a higher probability of success when this treatment is received after an early detection, by terms of a screening procedure (See [21], for a similar study). Theoretically, this could have been done, but since there is no clear evidence about these probabilities of success in the literature, the uncertainty parameters of the model would have been increased and therefore, it would turn it less tractable from a practical viewpoint.

To conclude, as one can infer from our model, and the subsequent sensitivity analyses, the sensitivity (obviously, apart from the costs) becomes the crucial data of each alternative. This is particularly meaningful when the impact of false positives is diminished as it can be inferred from the literature, in the case of our problem. Under our opinion, this should induce specialists to refine protocols in order to look for better levels of sensitivity, even at the cost of presenting worse levels of specificity. Obviously, that should be done taking into

account the trade-off between cost and sensitivity which is reflected in what we called cost-sensitivity ratios. However, this may be a hard task for specialists, and therefore it should be done by health policy makers, as part of an ensuing economic evaluation.

8 Appendix

In this section, we present an explanation of the utility weights, given by each measure of benefits, EQ-5D and HUI, to each of the resulting groups after the implementation of a particular screening procedure. Furthermore, we give the Proof of the Claim.

8.1 EuroQol measure (EQ-5D)

The EQ-5D descriptive system, developed by the EuroQol Group, classifies patients into 243 health states plus the states death and unconscious. Once the patient has been classified into such a health state, the researcher can assign a relevant value. These values are based on previous research where usually not all the values of the possible health states have been measured.

One of the methodological starting points of the EuroQol Group was the assumption that health can be characterized by a set of scores applied to five aspects of health status. The EQ-5D refers to these aspects as dimensions. In particular, the EQ-5D refers to five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression.[18] The dimensions each comprise three levels: no problems, some or moderate problems, and extreme problems or unable to. Within the EQ-5D classification system, every individual health state can be described by a row vector $\overrightarrow{x} = (x_1, ..., x_5)$, in which the element x_i represents the score on dimension i. Thus, x_1 is the score on mobility, x_2 is the score on self-care, etc. The score on a dimension is '1' if it is the highest level and '3' if it is the lowest. For example, a health state such as: some problems in walking about, no problems with self-care, no problems with performing usual activities, moderate pain and moderate anxiety could be represented by the row vector $(2, 1, 1, 2, 2)$, usually abbreviated to 21122. This notation is used throughout this work. The states death and unconscious cannot be described within the classification system.

In our problem, the EQ-5D measure provides the following quality weights, which were obtained by reviewing the literature and by asking some experts in the issue. First of all, as we mentioned above, we assigned the same utility weights to those individuals without congenital impairment, independently of their test results. Let us now consider the utility gains of those impaired newborns who were correctly identified after a screening procedure (true positives) compared with those who were not (false negatives).

We consider that there are no differences between them from the birth to the beginning of the treatment, not far away of the sixth month of life, as the NIH recommend [32]. We estimated three years and a half, for the subsequent treatment of a true positive newborn, including the possibility of a cochlear implant at the fourth year of life, hearing aids, and the deaf community language's learning, which will help the children in the communication with her parents first and with the rest of the society on a second step. Since the hearing impairment

[18] See [6] or [12], among others, for further details about the EQ-5D measure.

detection point of time is estimated in the mean age of 3 years [32], we assume that there is a utility loss, from the sixth month to the third year, caused by the hypothetical pain of the treatment. Namely, a health status of 11121 for the true positive individual and 11111 for the false negative one. Finally, after the fourth year, and once the treatment has been finished, we assume that the true positive individual reaches the perfect health status, while the false negative one reaches only 11212. This will remain constant throughout the remain 70 years of life.[19] To summarize, and considering that the total number of true positives provided by a screening program s is $p \cdot se_s$, where p denotes the prevalence of hearing loss and se_s denotes its sensitivity, the utility gains would be

$$p \cdot se_s \cdot \left(\frac{5}{2} \left(U(11121) - U(11111) \right) + 70 \left(U(11111) - U(11212) \right) \right) = 17.4305 \cdot p \cdot se_s.$$

The utility weights were obtained from the Spanish tariffs estimated for the EQ-5D, that are available [2]. There are also EQ-5D tariffs, estimated from an England survey [9], which would give us a significantly different amount of QALYs gained ($12.65 \cdot p \cdot se_s$). As a consequence, this would produce significantly different ratios. About the discounted number of QALYs, with a discount rate of 0.05, the utility gains would be $3.79 \cdot p \cdot se_s$ ($2.66 \cdot p \cdot se_s$ in the case of the English weights).

8.2 Health Utilities Index measure (HUI)

There are three HUI systems: HUI1, HUI2, and HUI3. Each system consists of a health status classification system and one or more scoring formulae. As in the EQ-5D, in all cases the scores are preference-based, interval-scaled, and on the 0 (dead) to 1 (perfect health) value scale. Scores are derived from preferences of members of the general public. The systems were developed over time, each building in part on the previous one. The HUI1 system only considered four attributes, and it was originally developed for the evaluation of outcomes, including indirect ones, of neonatal intensive care, but was later used more broadly (See [11], for further details about it, and [31] for an application of it to our problem). The HUI system was further extended, specially for paediatric applications, resulting in the HUI2 system. The HUI2 system comprises six attributes and preferences for its scoring function were measured on a random sample in Canada (See [12], for further details about it, and [31] for an application of it to our problem). Finally, the HUI3 classification system was based closely on that of the HUI2. The sensory attribute of HUI2 was expanded in HUI3 into the three attributes: vision, hearing and speech. The remain changes were made to increase the structural independence of the attributes, and therefore the estimation of the scoring system was simplified. As with the HUI2 system, preferences for the HUI3 were measured on a random sample in Canada (See [12], [14] or [15] for further details about it). To summarize, within the HUI3

[19] Recall, that we assume a life horizon of 74 years.

classification system, every individual health state can be described by a row vector $\overrightarrow{x} = (x_1, ..., x_8)$, in which the element x_i represents the score on dimension i. More precisely, x_1 is the score on vision, x_2 is the score on hearing, x_3 is the score on speech, x_4 is the score on ambulation, x_5 is the score on dexterity, x_6 is the score on emotion, x_7 is the score on cognition and x_8 is the score on pain. The dimensions each can comprise a different number of levels. It can be observed that, since there is a specific attribute which concerns hearing status, this measure fits better to our problem.

In our problem, the HUI3 measure provides the following quality weights. We proceed in the same manner than in the case of the EQ-5D, but with the new scoring function. As a result, the utility gains would be

$$p \cdot se_s \cdot \left(\frac{5}{2} \left(U(11111112) - U(11111111) \right) + 70 \left(U(13211211) - U(15311321) \right) \right) = 26.71 \cdot p \cdot se_s.$$

About the discounted number of QALYs, with a discount rate of 0.05, the utility gains would be $6.28 \cdot p \cdot se_s$.

8.3 Proof of the Claim

Given a screening procedure s, let us denote its costs by c_s, and by Q_s the utility gains (in QALYs) after implementing it. Thus, its CU-ratio, i.e. the cost per QALY gained that offers, is $\frac{c_s}{Q_s}$. For the sake of completeness, let us denote by se_s the sensitivity of s, and by p the prevalence of hearing loss. Now, under our assumption, there are not utility differences between those newborns without congenital impairment, even when they got different results in the screening test. As a consequence, the utility gains provided by a screening program refer to those newborns, whose congenital impairment was detected after the screening test, and therefore, they were referred for the subsequent treatment (i.e. the true positives). If u_1 denotes the utility of a newborn's health profile when its impairment was detected thanks to the screening, and u_2 when it was not, then the utility gains are the following:[20]

$$Q_s = p \cdot se_s \cdot (u_1 - u_2),$$

where $p \cdot se_s$ is the proportion of true positives after implementing the screening procedure s. Notice that u_i depend on the generic utility measure, and on the discount rate chosen. As a result, the CU-ratio related to s could be expressed as follows:

$$\frac{c_s}{Q_s} = k \cdot R_s,$$

where $k = k(u, r, p) = \frac{1}{p \cdot (u_1 - u_2)} > 0$, and $R_s = \frac{c_s}{se_s}$ is what we called the CS-ratio. Note that k is a positive value which depends on the prevalence of

[20] Assume $u_1 > u_2$.

hearing loss, and the chosen discount rate and utility measure. However, k is not screening method-specific. In other words, k is constant for every screening procedure s. As a result, the program which offers the lowest CU-ratio, i.e. the one that CUA would recommend for implementing, is the program which offers the lowest CS-ratio. ■

9 Tables

Status (S) / Test (T)	Positive (S^+)	Negative (S^-)	Total
Positive (T^+)	a	b	$a+b$
Negative (T^-)	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d=q$

Sensitivity (se)	$se = P(T^+\|S^+) = \frac{a}{a+c}$
Specificity (sp)	$sp = P(T^-\|S^-) = \frac{d}{b+d}$
Prevalence (p)	$p = \frac{a+c}{q}$
True positives (a)	$a = p \cdot se \cdot q$
False positives (b)	$b = (1-p) \cdot (1-sp) \cdot q$
False negatives (c)	$c = p \cdot (1-se) \cdot q$
True negatives (d)	$d = (1-p) \cdot sp \cdot q$

Table 1: Congenital impairment status

VARIABLE		MEAN	RANGE	REF.
UNHS1	Sensitivity	0.784	$0.528 - 1$	[23]
	Specificity	0.996	$0.75 - 1$	
	Cost ($)	10.05	$6.5 - 15.2$	
UNHS2	Sensitivity	0.9025	$0.81 - 0.96$	[24]
	Specificity	0.95	$0.93 - 0.98$	
	Cost ($)	13.91	$11.75 - 20.38$	
UNHS3	Sensitivity	0.84	$0.58 - 0.94$	[18]
	Specificity	0.995	$0.75 - 1$	
	Cost ($)	11.68	$10.14 - 16.05$	
TNHS1	Sensitivity	0.463	$0.264 - 0.64$	[23]
	Specificity	0.999	$0.75 - 1$	
	Cost ($)	1.59	$0.83 - 15.77$	
TNHS2	Sensitivity	0.532	$0.405 - 0.615$	[23] [24]
	Specificity	0.998	$0.75 - 1$	
	Cost ($)	1.65	$1.08 - 16.02$	
TNHS3	Sensitivity	0.4956	$0.29 - 0.60$	[18]
	Specificity	0.999	$0.75 - 1$	
	Cost ($)	1.57	$1 - 15.82$	

Table 2: Protocols.

PROC.	PREV.	c_s^1	$c_s^1 + c_s^3$	$c_s^1 + c_s^3 + c_s^4$
TNHS1	0.11%	1.59	−20.234	−195.284
TNHS2	0.11%	**1.65**	−23.426	−224.564
TNHS3	0.11%	1.57	−21.79	−209.167
UNHS1	0.11%	10.05	−26.904	−323.318
UNHS2	0.11%	13.91	**−28.629**	**−369.846**
UNHS3	0.11%	11.68	−27.913	−345.501
TNHS1	0.35%	1.59	−67.845	−624.829
TNHS2	0.35%	**1.65**	−78.137	−718.123
TNHS3	0.35%	1.57	−72.758	−668.955
UNHS1	0.35%	10.05	−107.53	−1050.67
UNHS2	0.35%	13.91	**−121.442**	**−1207.13**
UNHS3	0.35%	11.68	−114.299	−1124.8

Table 3: Incremental costs (per newborn) with respect to the 'status quo'.

PROC.	PREV.	QALYs gained		($r = 0.05$)	
		EQ-5D	HUI3	EQ-5D	HUI3
TNHS1	0.11%	0.009	0.014	0.0019	0.0032
TNHS2	0.11%	0.010	0.016	0.0022	0.0037
TNHS3	0.11%	0.009	0.015	0.0021	0.0034
UNHS1	0.11%	0.015	0.023	0.0033	0.0054
UNHS2	0.11%	**0.017**	**0.027**	**0.0038**	**0.0062**
UNHS3	0.11%	0.016	0.025	0.0035	0.0058
TNHS1	0.35%	0.028	0.043	0.0061	0.0102
TNHS2	0.35%	0.032	0.050	0.0071	0.0117
TNHS3	0.35%	0.030	0.046	0.0066	0.0109
UNHS1	0.35%	0.048	0.073	0.0104	0.0172
UNHS2	0.35%	**0.055**	**0.084**	**0.0120**	**0.0198**
UNHS3	0.35%	0.051	0.079	0.0111	0.0185

Table 4: QALYs gained (per newborn).

PROC.	PREV.	CS-RATIOS	MCU-RATIOS		($r = 0.05$)	
			EQ-5D	HUI3	EQ-5D	HUI3
TNHS1	0.11%	3.434	179.107	116.882	823.547	497.203
TNHS2	0.11%	**3.101**	**161.759**	**105.562**	**743.78**	**449.045**
TNHS3	0.11%	3.168	165.221	107.821	759.697	458.655
UNHS1	0.11%	12.819	668.571	436.298	3074.13	1855.96
UNHS2	0.11%	15.413	803.854	524.582	3696.17	2231.5
UNHS3	0.11%	13.905	725.205	473.257	3334.54	2013.17
TNHS1	0.35%	3.434	56.2909	36.7345	258.829	156.264
TNHS2	0.35%	**3.101**	**50.8387**	**33.1765**	**233.759**	**141.128**
TNHS3	0.35%	3.168	51.9267	33.8865	238.762	144.149
UNHS1	0.35%	12.819	210.122	137.122	966.155	583.3
UNHS2	0.35%	15.413	252.64	164.869	1161.65	701.33
UNHS3	0.35%	13.905	227.922	148.738	1048	632.712

Table 5: Ratios from a hospital viewpoint.

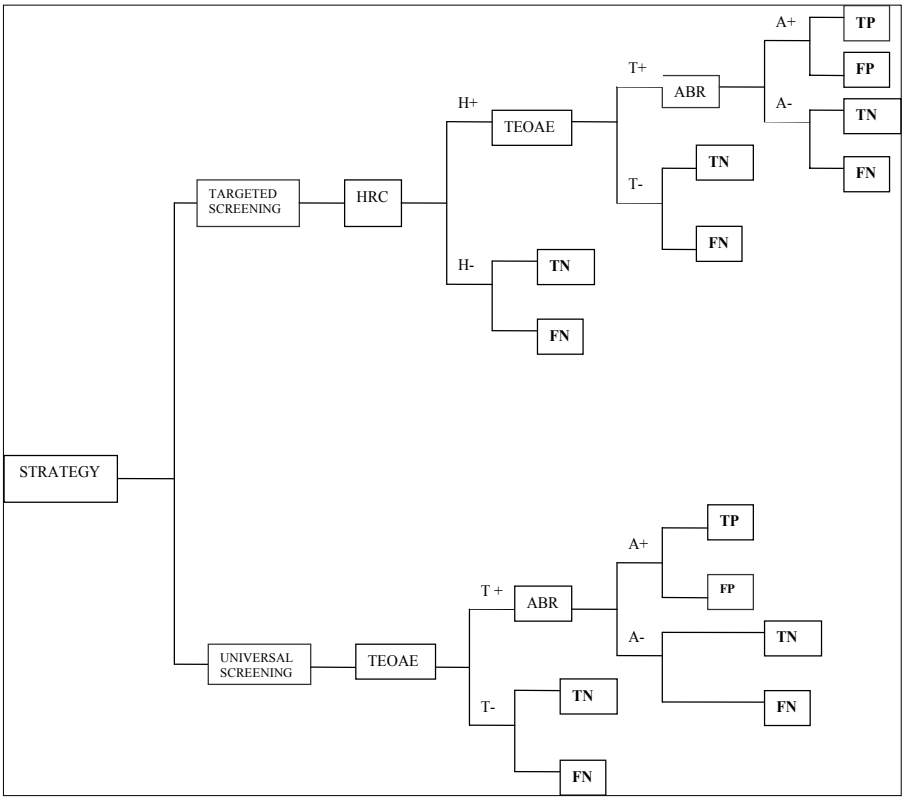

Figure 1: UNHS1 Vs. TNHS1

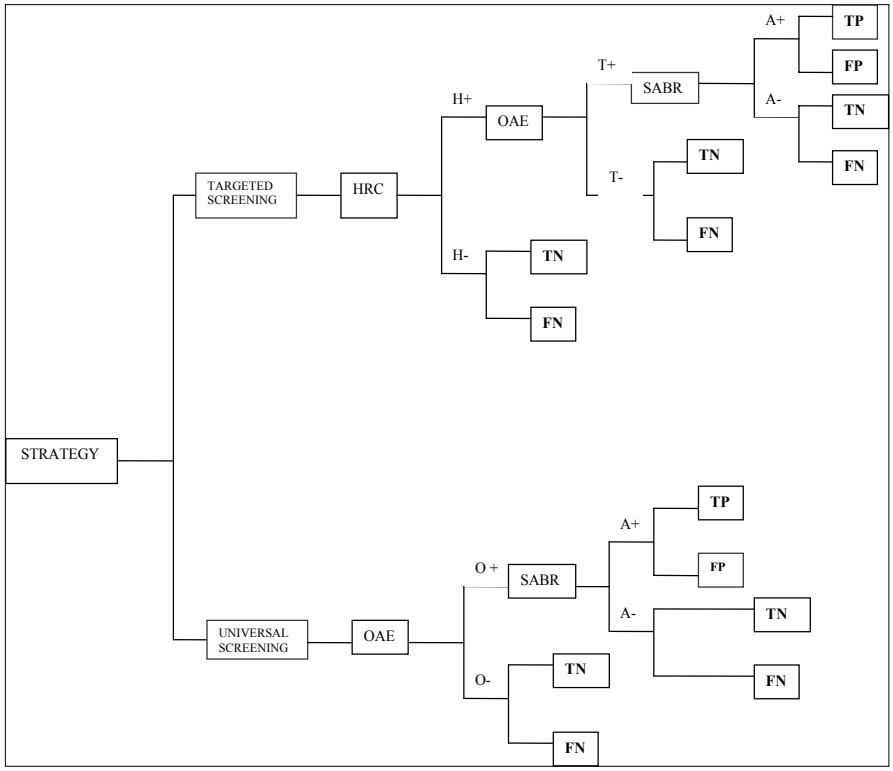

Figure 2: UNHS2 Vs. TNHS2

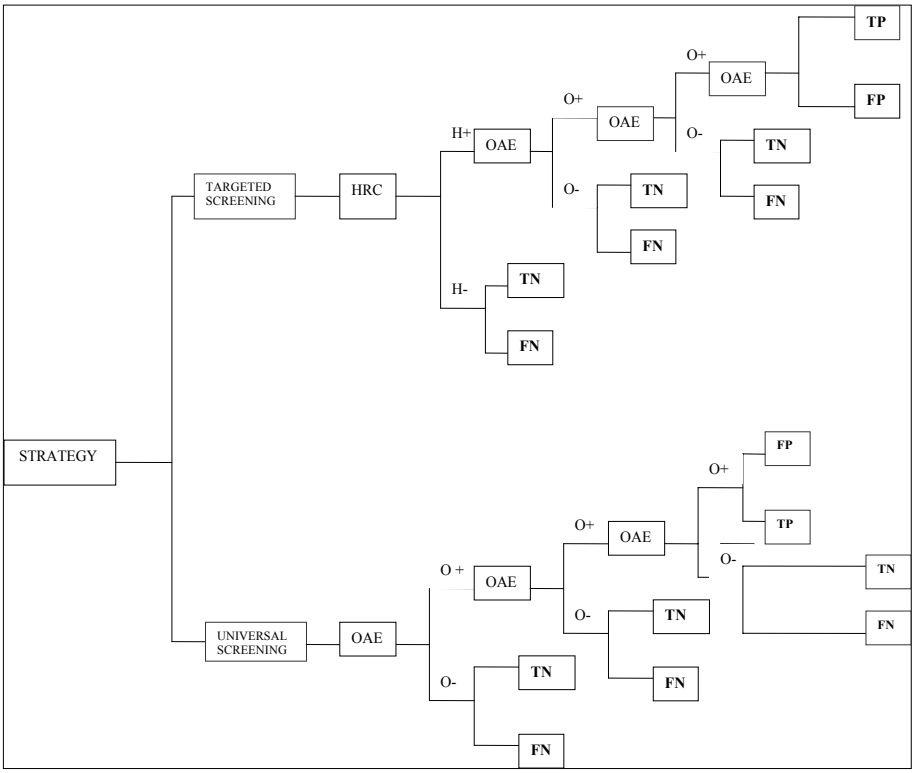

Figure 3: UNHS3 Vs. TNHS3

References

[1] American academy of pediatrics. *Task force on newborn and infant hearing. Newborn and infant hearing loss: detection and intervention.* Pediatrics. **103 (2)** 1999. 527-530.

[2] Badia X, Roset M, Monserrat S, Herdman M, & Segura A. *The Spanish version of EuroQol: a description and its applications. European Quality of Life Scale.* Med. Clin. **112** (Supl. 1) 1999. 79-86

[3] Barsky-Firsker L, Sun S. *Universal newborn hearing screening: a three year experience.* Pediatrics. **99 (6)** 1997.

[4] Bess FH, Paradise JL. *Universal screening for infant hearing impairment: not simple, not risk-free, not necessarily beneficial and not presently justified.* Pediatrics. **93 (2)** 1994. 330-334.

[5] Bleichrodt H. *Applications of utility theory in the economic evaluation of health care.* Thesis Erasmus University Rotterdam. Rotterdam. 1996.

[6] Busschbach J, McDonnell J, Essink-Bot ML, van Hout BA. *Estimating parametric relationships between health description and health valuation with an application to the EuroQol EQ-5D*, J. Health Econ. **18.** 1999. 551-571

[7] Clemens CJ, *The false positive in universal newborn hearing screening*, Pediatrics. **106 (1)** 2000. E7.

[8] Comisión para la detección precoz de la hipoacusia en recién nacidos de la Comunidad Valenciana, *Programa de detección precoz de hipoacusias,* Consellería de Sanidad. Generalitat Valenciana. 1999.

[9] Dolan P, *Modeling valuations for EuroQol health states,* Med. Care. **35.** 1997. 1095-1108.

[10] Downs MP, Yoshinaga-Itano C. *The efficacy of early identification and intervention for children with hearing impairment*, Pediatr. Clin. North. Am. **46 (1)** 1999. 79-87.

[11] Drummond M, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes.* O.U.P. 1989.

[12] Drummond M, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes. Second Edition* O.U.P. 1997.

[13] European consensus project on neonatal hearing screening. *European consensus statement on neonatal hearing screening.* Eur. J. Pediatr. **158** 1999. 95-96.

[14] Feeny, D., Furlong, W., Torrance, G.W., Goldsmith, C.H., Zhu, Z., Depauw, S., Denton, M., and Boyle, M. *Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system.* Med. Care forthcoming

[15] Furlong, W., Feeny, D., Torrance, G.W., Goldsmith, C.H., Depauw, S., Zhu, Z., Denton, M., and Boyle, M. *Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report.* Hamilton, Ontario, Canada: McMaster University, Centre for Health Economics and Policy Analysis. Working Paper 98-11, 1998.

[16] Gordis L. *Epidemiology.* W.B. Saunders Company. Philadelphia 1996.

[17] Grant R. *The case to fund universal newborn hearing screening in New York State.* Int. J. Pediatr. Otorhinolaryngol. **54** 2000. 79-80.

[18] Grupo técnico de trabajo sobre la detección precoz de hipoacusias neonatales en Navarra. *Programa universal de detección precoz de hipoacusias neonatales en Navarra.* Servicio de docencia, investigación y desarrollo sanitario. Departamento de salud. Govierno de Navarra. 1999.

[19] Health Utilities Inc. *Multi-Attribute Health Status Classification System: Health Utilities Index Mark 3 (HUI3).* Available online at www.healthutilities.com/hui3.htm

[20] Hennekens CH. *Epidemiology in medicine.* Little Brown and Company. Boston. 1987

[21] Herrero C. and Moreno-Ternero J.D. *Income-based decision making in the economic evaluation of health care programs.* Mimeo. University of Alicante. 2001

[22] Joint Committee on Infant Hearing. *Year 2000 position statement: principles and guidelines for early hearing detection and intervention programs.* Am. J. Audiol. **9** 2000. 9-29.

[23] Kemper AR, Downs SM. *A cost-effectiveness analysis of newborn hearing screening strategies*, Arch. Pediatr. Adolesc. Med. **154 (5)** 2000.

[24] Kezirian EJ, White KR, Yueh B, Sullivan S. *Cost and cost-effectiveness of universal screening for hearing loss in newborns*, Otolaryngol Head Neck Surg. **124(4)**. 359-67. 2001.

[25] Maxon AB, White KR, Vohr BR, Behrens TR. *Referral rates and cost efficiency in a universal hearing screening program using transient evoked otoacoustic emissions*. J. Am. Acad. Audiol. **6 (4)** 1995. 271-7.

[26] Mehl AL, Thomson V. *Newborn hearing screening: the great omission.* Pediatrics. **101 (1)** 1998.

[27] Meltzer D. *Accounting for future costs in medical cost-effectiveness analysis*, J. Health Econ. **16 (1).** 1997. 33-64.

[28] Meltzer D, Johannesson M. *Inconsistencies in the "societal perspective" on costs of the Panel on cost-effectiveness in health and medicine*, Med Decis Making. **19.** 1999. 371-377.

[29] Meyer C et al. *Neonatal screening for hearing disorders in infants at risk: incidence, risk factors and follow-up*, Pediatrics. **104 (4)** 1999. 900-904.

[30] Moeller MP. *Early intervention and language development in children who are deaf and hard of hearing*, Pediatrics **106 (3)** 2000.

[31] Moreno-Ternero J.D. *Métodos para la evaluación económica de programas de atención sanitaria.* Tesina Universidad de Alicante. Alicante. 2001.

[32] National Institutes of Health. *Consensus statement. Early identification of hearing impairment in infants and young children.* NIH Consensus Statement.1993. Mar 1-3**. 11,**1-24.

[33] O'Neill C, O'Donogue G, Archbold S, Normand C. *A cost-utility analysis of pediatric cochlear implantation.* Laryngoscope **110**. 2000. 156-160

[34] Paradise JL. *Universal newborn hearing screening should we leap before we look?* Pediatrics. **103 (3)** 1999. 670-672.

[35] Stuart A, Moretz M, Yang E. *An investigation of maternal stress after neonatal hearing screening.* Am. J. Audiol. **9** 2000. 1-7.

[36] Thomson V. *The Colorado newborn hearing screening project.* Am. J. Audiol. **6 (3)** 1997. 74-77.

[37] Vohr BR, Carty LM, Moore PE, Letourneau K. *The Rhode Island hearing assessment program: experience with statewide hearing screening.* J. Pediatr. **133 (3)** 1998. 353-357.

[38] Vohr BR, Maxon AB. *Screening infants for hearing impairment.* J. Pediatrics. **128 (5)** 1996. 710-714.

[39] Weinstein MC, Siegel JE, Gold MR, Kamlet MS. *Recommendations of the panel on cost-effectiveness in health and medicine.* JAMA **276** 1996. 1253-1258.

[40] Weithrather YP, Korth N, White KR, Downs D, Woods-Kershner N. *Cost analysis of TEOAE-based universal newborn hearing screening.* J. Commun. Disord. **30** 1997. 477-493.

[41] White KR, Maxon AB. *Universal screening for infant hearing impairment: simple, beneficial and presently justified.* Int. J. Pediatr. Otorhinolaryngol. **32 (3)** 1995. 201-211.

[42] Yoshinaga-Itano C, Sedey AL, Coulter DK, Mehl AL. *Language of early and later identified children with hearing loss.* Pediatrics. **102 (5)** 1998. 1161-71.