Author's response to reviews

Title: The cost-effectiveness of tracking newborns with bilateral hearing impairment in Bavaria: A decision-analytic model

Authors:

Astrid Langer (langer@bwl.lmu.de)
Inken Brockow (Inken.Brockow@lgl.bayern.de)
Uta Nennstiel-Ratzel (Uta.Nennstiel-Ratzel@lgl.bayern.de)
Petra Menn (petra.menn@helmholtz-muenchen.de)

Version: 3 Date: 23 August 2012

Author's response to reviews: see over
Resubmission

Dear Editors,

Thank you for consideration of our manuscript for publication in your journal. We have reviewed the manuscript according to the reviewers’ and associate editor’s comments. According to the associate editor’s comments, tables and figures were adapted to the journal style, and the language was revised by a native English speaker. On the following pages, you can find a point-by-point response to the reviewers’ comments.

Best regards,

Astrid Langer, Inken Brockow, Uta Nennstiel-Ratzel and Petra Menn
Reviewer's report
Title: The cost-effectiveness of tracking newborns with bilateral hearing impairment in Bavaria: A decision-analytic model
Version: 2
Date: 18 May 2012
Reviewer: Scott D. Grosse

Reviewer's report:
The authors make the case in this paper that centralized tracking and follow-up of infants who fail hearing screening during the newborn period is not only essential to maximize the detection of bilateral congenital hearing impairment but is also cost-effective. I am in full agreement with this conclusion. A previous article in German by this group of investigators demonstrated the importance of centralized tracking for detection of children with hearing loss in Bavaria in reducing loss to follow-up. This paper is a useful complement to the previous publication. First, it shows that such tracking is cost-effective as well. Second, it will bring this point to a wider, English-language readership. This is the first paper I am aware of that has modeled the economic impact of centralized tracking of hearing screening results in terms of cost per case detected. This paper will be an important contribution to the UNHS literature.

Thank you for your comment.

Major compulsory revisions
1. Some clarification is required. The paper indicates that this analysis is based on the experience of tracking of hearing screening results in Bavaria. However, there are discrepancies with published reports of hearing screening in Bavaria, specifically both the pilot project conducted in the northern part of Bavaria from 2003 to 2008 and described in reference 7 and the state-wide screening programme which was implemented in 2009 and described in reference 6. Due to data availability, we used data from the pilot project from 2003 - 2008 and data from the state-wide screening programme implemented in 2009. In Table 2, the source for the probabilities used in the model is documented.

Those discrepancies need to be explicitly addressed, not necessarily in terms of changes to the model but in documenting limitations and implications for the interpretation of results. We completely agree with you. In the ‘Discussion’ section, it is stated: ‘Second, some of the data used are taken from a Bavarian pilot project. Therefore, these data may differ from data compiled since the nationwide implementation of newborn hearing screening in 2009 and data from other newborn hearing screening programmes in Germany, which may have implications for the generalizability of results. However, the issue of generalizability was addressed in sensitivity analyses’ (p. 13).

2. The first discrepancy is that the screening protocol described does not appear to correspond to either reference. In particular, referrals in Bavaria occur if an infant fails in either ear, not just those with bilateral refers. We modelled the referral of infants who failed testing in both ears, as is standard practice in Germany.

In the pilot project, there was a 3-step process in which infants were first screened in the birthing centre or paediatric clinic using TEOAE and, if that test failed were retested using AABR. If that test failed, infants underwent a confirmatory test by a physician, followed by a full diagnostic audiologic work-up. The report of the pilot screening project makes clear that
infants who failed the test in either ear were referred for further testing. In 2008, 2.2% of infants tested were referred, among which <30%, or 0.6% failed bilaterally.

The pilot project consisted of a 3-stage process as you described above. Due to a shortage of pediatric audiologists in Bavaria after the state-wide implementation of newborn hearing screening in 2009, not all children with positive test results after discharge from hospital are referred for a full diagnostic audiological work-up, but undergo up to two confirmatory tests, as is common practice in Bavaria and presumably in Germany. The structure of our model is based on this practice.

The state-wide screening protocol involves a single TEOAE test in both ears, with a failure in either ear leading to a second test using AABR prior to discharge. Infants who fail both tests in either ear are then referred for further testing. In 2010, 4.5% of all infants screening in Bavaria were referred for further testing, of which one-third, or 1.5%, failed the hearing screen in both ears. That discrepancy is not critical because tracking was limited to bilateral refers. Consequently, the cost of further testing for unilateral refers would presumably be the same in both arms and therefore not affect the estimates of incremental costs or yield.

We agree with you. In our model, we used the data from the pilot project (0.6% bilateral refers).

3. Rates of diagnosed bilateral congenital hearing impairment in Bavaria appear substantially higher than was assumed in the decision analytic model. The decision model projects that of 100,000 infants screened, 51 cases of bilateral hearing impairment will be found with tracking and 21 without tracking. It is useful to compare these results with the actual experience of the hearing screening programme in Bavaria. Reference 6 reports that in the first year of operation, 2009, there were 108 (102!) cases of confirmed bilateral hearing loss detected out of 58,500 infants screened, for a rate of 1.85 (1.74!) per 1000. That is 3 times higher than the rate of 0.51 per 1000 with tracking assumed in the decision analytic model. Moreover, reference 6 reported that 50.7% of the children with bilateral hearing impairment were followed up only because of the tracking centre. That implies that 53 (50!) cases of bilateral hearing impairment would have been diagnosed in the absence of the tracking programme, for a rate of 0.9 (0.85!) per 1000, which is more than 4 times higher than the rate projected in the decision analytic model.

During the pilot project, both refer rates and diagnosis rates were substantially lower, and only 0.7 per 1000 children screened were diagnosed with and received therapy for bilateral hearing loss. That implies that the sensitivity of the pilot screening protocol was less than half as high as the protocol used subsequently in the state-wide screening programme based on one year of data; confirmatory test results for 2010 were not yet available for all children when the study was published.

The refer rate was lower owing to a better education for the personnel during the pilot project. The diagnosis rate was lower, as children at risk are more likely to be born in other parts of Bavaria (Munich, Erlangen, and Nuremberg). These cities were not part of the pilot project. The prevalence was, therefore, lower.

Furthermore, in Germany, the quality of a newborn hearing screening programme is considered high, if coverage rate is over 95% and referral rate after the initial hearing screening test is under 4%, which was fulfilled in the pilot project (95.5% coverage rate and 2.2% referral rate, of which 0.6% are bilateral refers), but only partly in the state-wide screening programme in 2010 (96.3% coverage rate and 4.5% referral rate, of which 1.5% are bilateral refers).

Nonetheless, the actual rate of detection of hearing loss in the pilot screening project was still almost 40% higher than that projected in the decision analysis on the basis of a hypothetical
screening protocol. If actual data on cases of hearing loss detected by screening and tracking had been used, the cost-effectiveness ratio would presumably have been lower.

In the pilot project, it is reported that from 2003 to 2008, there were 51 cases of confirmed bilateral hearing loss detected out of 73,332 infants screened, resulting in a rate of 0.70 per 1000. That is indeed higher than the rate of 0.51 per 1000 with tracking in the model. In the pilot project, 52% of the children with bilateral hearing impairment were followed up only because of the tracking centre. Therefore, 27 cases of bilateral hearing impairment would have been diagnosed in the absence of the tracking programme, resulting in a rate of 0.5 per 1000 compared with 0.31 per 1000 in the decision-analytic model. That may imply that, if actual data on cases of hearing loss detected by screening and tracking had been used, the incremental cost-effectiveness ratio would probably have been lower. We added this paragraph to the ‘Discussion’ section (p. 12).

Minor essential revisions
1. The statement that the protocol was adapted from reference 19 should be modified to make clear that Kezirian et al. modeled the referral of infants who failed testing in either ear, as is standard practice in the United States.
We deleted this sentence in order to guard against misunderstandings. It now reads: ‘As these newborns are not identified at an early stage as a consequence of screening, they are not counted as part of the yield of the four-stage test procedure [19]’ (p. 7-8).

Discretionary revisions
1. The authors may wish to discuss the economic implications of extending centralized tracking to include unilateral referrals, some of which may result in the diagnosis of bilateral hearing impairment.
‘The cost-effectiveness of the intervention in different patient groups (uni- and/or bilateral hearing impairment) was not assessed because the target population was newborns with bilateral hearing impairment only, as is standard practice in Germany [4]. Thus, the analysis is rather conservative. If centralized tracking was extended to include unilateral referrals, some of which may result in the diagnosis of bilateral hearing impairment, the incremental cost-effectiveness ratio would presumably be lower. A recent study found that, children with unilateral hearing loss showed worse language skills compared with their siblings with normal hearing [27]. However, more research is needed to clarify this issue.’ (p. 13).
Reviewer's report
Title: The cost-effectiveness of tracking newborns with bilateral hearing impairment in Bavaria: A decision-analytic model
Version: 2
Date: 7 June 2012
Reviewer: Andreas Gerber

Reviewer's report:
The cost-effectiveness of tracking newborns with bilateral hearing impairments in Bavaria.

Dear authors,

thanks for your health economic evaluation of a screening and tracking program for newborns to detect bilateral hearing impairment.

Overall I must say that I do not see a basis to perform a health economic evaluation of the problem you address. The data as you know are very weak, see Wolff et al. 2010 and IQWiG 2007 – with reference to these publications I would like to make you aware that they are the same product which I cannot gather from the way you quote them. Wolff p. 131 right column: “study quality was generally poor …. Only one treatment study showed “minor deficiencies”. Therefore, we cannot currently say anything about the effect of early vs. late treatment, henceforth nothing on the effect that a tracking system would have on the reduction of impairments children with bilateral hearing impairments would suffer from. Further, as you can see from p. 134 left column the authors conclude: “… this can only be regarded as indication that the expressive and receptive language abilities …. are better in children treated earlier.”

Thank you for your comment. We only partly agree with you. For instance, the report published by IQWIG in 2007 is five years old and, thus, does not reflect current evidence. Furthermore, your quotation above is rather selective. In the abstract of the systematic review by Wolff et al., you can read: ‘Results: The studies comparing screening versus no screening showed an improvement of speech development of children in the screening group compared with the group without screening. Early treatment was associated with better language development in comparison to children with later treatment. Conclusions: The authors concluded that there is a lack of high-quality evidence regarding all elements of newborn hearing screening. Early identification and early treatment of children with hearing impairments may be associated with advantages in language development. Other patient-relevant parameters, such as social aspects, quality of life, and educational development, have not been adequately investigated1. We are in full agreement with these results and conclusions. The findings by Wolff et al. are supported by a recent review2. To address your concerns, we changed the text to the following: ‘A systematic review of newborn hearing screening by Wolff et al. [2] found that screening may be associated with better language development than no screening. Furthermore, this systematic review revealed that early treatment may result in better language development than later treatment. These findings are supported by a recent review [3]’ (p. 4).

In the context of health economic evaluation of a screening program I would like to also mention the concept of linked evidence meaning that a screening in itself that is not linked to an effect in treatment would not be considered a patient-relevant outcome according to Social Code Book in Germany in particular and according to current standards of evidence-based medicine in general, see publications by Holger Schuenemann from Hamilton University on

the concept of linked evidence). Therefore, from my personal point of view and from the point of view of evidence-based medicine I think it is too early to perform health economic evaluations without the basis of solid evidence. This is a common problem of screening interventions, but we should be aware of it and not make conclusions on an invalid data basis. I think we cannot publish papers any more that do health economic analyses without knowing anything about the long-term outcomes, as you yourself point out in p. 7 of your manuscript.

We understand your concerns, but do not share them. First, health economic modeling is ‘an analytic methodology that accounts for events over time and across populations, that is based on data drawn from primary and/or secondary sources and whose purpose is to estimate the effects of an intervention on valued health consequences and costs’. The aim of health economic modelling is to generate expected values for the clinical and economic effects of [therapeutic] alternatives. Furthermore, decision analysis is considered a useful tool for difficult decisions with significant uncertainty. Second, we chose a short time horizon for our model to avoid the use of ‘invalid’ data. Third, there is growing evidence supporting the benefit of early treatment, and universal newborn hearing screening has been adopted in many countries to date.

There is also growing awareness of the importance of tracking. For instance, in a recent publication on newborn screening and outcomes by CDC (Centers for Disease Control and Prevention), the authors concluded: ‘Given all these challenges and opportunities, screening itself clearly is not enough. It is critical to avoid complacency in assuming that every newborn who is screened will receive optimal service and care. Short-term follow-up and management of children with disorders and long-term follow-up activities within the entire newborn screening system are central to realizing the promise of newborn screening’. We are in full agreement with this conclusion. Furthermore, an evaluation of a Dutch newborn hearing screening programme found that ‘both participation in diagnostic testing after a positive screen result and the timing of the diagnostic testing can still be improved.’ Our intention was to show the overall importance of short-term follow-up for the effectiveness and cost-effectiveness of newborn hearing screening programmes. For instance, a recent study showed that higher levels of maternal education were linked to earlier confirmation of hearing loss. A previous publication by the Bavarian Food and Health Safety authority found that with early treatment the concep

---


within 3 months and treatment within 6 months\textsuperscript{12}. A recent study from Japan showed that early intervention had a strong influence on language development\textsuperscript{13}. Furthermore, the authors pointed out that it should be ensured that early detection results in early treatment. As Kemper and Downs have recently stated: ‘The key challenge is assuring that screening is consistently administered with good follow-up and that those identified with hearing impairment receive effective intervention’\textsuperscript{14}.

Aside from this basis comment I have some more general comments. This paper seems to be drafted with a very specific audience in mind, making it difficult to comprehend. Substantial knowledge of Germany seems to be necessary to fully appreciate. It is, for instance, never made clear what the Joint Federal Commission is and how it is relevant. Indeed, our paper is primarily a contribution to the universal newborn hearing screening literature. However, we revised the article, where necessary, to bring it to a wider readership. For instance, we replaced the ‘Joint Federal Commission’ by Germany, as naming of this institution is not central to understand our paper and may cause confusion especially with regard to non-German readers.

Furthermore, the presentation is a bit problematic as it is difficult to follow, in particular when the reader does not have any background in hearing impairment screening. It is not clear why several test stages are needed if the AABR test as a 100% test accuracy (p. 10). From the 9 studies on reliability and validity of methods that were included in the IQWiG 2007 report S-OAE and A-ABR not sufficiently evaluated. So please explain where you get the data from.

You are absolutely right. The full diagnostic audiologic work-up performed by pediatric audiologists to confirm or exclude hearing impairment consisted of an impedance test of the middle ear and an AABR test. We deleted the former sentence. The source of data is presented in Table 2 (please see also the answer to your next comment).

Specific comments:
The authors might want to describe the pilot, the number of included children from Upper Palatine etc. in more detail for the non-German native speakers that would want to read this article. A short description of the timeline etc. would do. As I know that reviewers ask for additions, but never suggest where to cut I would suggest to shorten the section on the literature review of other health economic evaluations and report them in a separate paper or just refer to some of them that are specifically on tracking in the discussion.

We used data from a pilot project on newborn hearing screening conducted in the northern part of Bavaria (Upper Palatinate and Upper Franconia) from 2003 to 2008 (see reference 7) and data from the state-wide screening programme in Bavaria which was implemented in 2009 (see reference 6). More information is provided in the following table, which was added to the main text (see p. 21).

<table>
<thead>
<tr>
<th>Overview of the newborn screening programmes in Bavaria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Newborns screened</strong></td>
</tr>
<tr>
<td>Newborns screened</td>
</tr>
<tr>
<td>Coverage rate</td>
</tr>
<tr>
<td>Refer rate (bilateral)</td>
</tr>
<tr>
<td>Tracking rate</td>
</tr>
</tbody>
</table>

\textsuperscript{12} Brockow I, Kummer P, Liebl B, Nennstiel-Ratzel U: [Universal newborn hearing screening (UNHS): is it possible to successfully implement it nationwide?]. Gesundheitswesen 2011, 73:477-482
Cases of bilateral hearing impairment detected

<table>
<thead>
<tr>
<th></th>
<th>51</th>
<th>102</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to diagnosis in months (mean/median)</td>
<td>5.4/4.5</td>
<td>6.0/5.5</td>
<td>5.1/4.2</td>
</tr>
<tr>
<td>Time to treatment in months (mean/median)</td>
<td>5.7/5.1</td>
<td>6.8/6.2</td>
<td>5.5/4.8</td>
</tr>
</tbody>
</table>

The authors use the term “failing” a test to indicate the test was positive (see Figure 1 footnote). In the abstract, however, they speak of “positive test results” and do not use “failing”. A positive test could be false or a true positive. I miss a discussion whether the higher detection rate with tracking also affects the ratio of true to false negative and how this affects cost-effectiveness.

In the long-term, detection rates are the same. Tracking does not affect diagnostic accuracy, but, of course, the time of detection and confirmation, which affects cost-effectiveness, as is shown in our model.

It is not clear to me how the confidence intervals for the cost estimates in Table 2 were derived. In particular as the EBM system seems to be a fixed fee aside from some minor regional differences throughout Germany.

We agree with you that the EBM system is a fixed fee scale, as is the case for the DKG-NT system. We deleted the respective confidence intervals.

Table 4: Why would you need an epidemiologist if the tracking program became standard? Therefore I do not understand why you should include these costs?

As is indicated in the title, our objective was to evaluate the cost-effectiveness of tracking newborns with bilateral hearing impairment in Bavaria based on Bavarian data and a decision-analytic model, the structure of which also reflected Bavarian practice. As you know, tracking is not standard practice in Germany. Even if the tracking programme became standard, an epidemiologist would be needed, for instance, to maintain, improve and evaluate the quality of the programme.

P 13: What are the “better outcomes” of tracking?

As the term ‘outcomes’ may be misleading here, we changed the sentence as follows: ‘In comparison with no tracking, tracking resulted in more cases of bilateral hearing impairment detected and higher costs.’ (p. 11).

In the discussion of the review of existing models (p. 11) it is not discussed why the estimate differ so dramatically, e.g. 25,813 GBP of a child detected vs. 5,113USD. Furthermore, currency abbreviations are not explained.

As this is the first paper that addresses the economic impact of centralized tracking of hearing screening test results in terms of cost per case detected and due to your other comment, we deleted this paragraph. Moreover, a systematic review on the cost-effectiveness of universal screening programmes for newborns with bilateral hearing impairment has recently been published\(^\text{15}\).

The authors report only ICERs for their analysis and not ACERs. This would not only be insightful but also helps to make it comparable with the estimates reported in the review of existing models.

We added this information to Table 6.

The authors acknowledge four major limitations. In particular, the explanation of the first limitation mentioned (the conditional probability of failing) has to be expanded. The probability of failing the second test is different from the probability of failing the first test, since the second test is only performed in those children who failed the first test. The present model accounts for this conditional dependence by using a different probability of failing for each test. In contrast, previous models had to assume that the probability of failing a test remains constant for all tests. We added this paragraph to page 13.

Furthermore, the discussion of the limitations should also try assessing how these affect the cost-effectiveness, i.e. on what direction are the ICERs affected.

Third, only parameter and structural uncertainty was addressed via sensitivity analyses, whereas methodological uncertainty (for example, discount rate, long time horizon) was not addressed owing to a lack of long-term data. The cost-effectiveness of the intervention in different patient groups (uni- and/or bilateral hearing impairment) was not assessed because the target population was newborns with bilateral hearing impairment only, as is standard practice in Germany [4]. Thus, the analysis is rather conservative. If centralized tracking was extended to include unilateral refers, some of which may result in the diagnosis of bilateral hearing impairment, the incremental cost-effectiveness ratio would presumably be lower. A recent study found that, children with unilateral hearing loss showed worse language skills compared with their siblings with normal hearing [27]. However, more research is needed to clarify this issue. Fourth, owing to a lack of adequate data, the time horizon was limited to the newborn hearing screening programme (initial hearing screening test and subsequent hearing tests or diagnosis) and a scenario analysis was not conducted. Over a lifelong time horizon, centralized tracking within a newborn hearing screening programme may be even cost saving.’ (p. 13-14).

A further major limitation I see is in the set-up of the costs and probabilities for the model. If I understand it correctly, the probabilities for a given stage are a weighted average of different kinds of test with different corresponding failure rates. This mixture is correctly reflected when calculating the costs by ensuring the costs at each stage reflect this weighted average of the corresponding tests. This means the results are only valid if this weighted average is constant. For example, if at the 2nd stage the share of (cheaper) OAE tests increases the probabilities used are not valid anymore and it is unclear in what direction the effectiveness changes.

In principle, your comment is correct. However, we performed extensive sensitivity analyses to see how the results change with a variation of these probabilities. Overall, the results of our model were relatively robust. Furthermore, for instance, it is common practice that normal children receive OAE in the first stage and children at risk AABR. This ratio can change from year to year.

On p. 15: I do not understand the term “methodological uncertainty” in this context, please expand.

The definition of ‘methodological uncertainty’ is taken from Philips et al.: ‘Methodological uncertainty relates to whether particular analytic steps taken in the analysis are the most appropriate (for example, discount rate used)’16. For clarification, we added ‘for example, discount rate, long time horizon’ in brackets after the term ‘methodological uncertainty’ (p. 13).

Some editing could increase clarity tremendously, e.g. on p. 7: “the benefits […] are open”. Prior to submission, the manuscript was revised by a native speaker. Nevertheless, we will seek the assistance of another native speaker. For instance, it now states: ‘Moreover, evidence on the benefits of early detection of newborns with unilateral hearing impairment, e.g., in terms of language and speech development, is lacking [10]’ (p. 6).