Reviewer’s report

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

Version: 3 Date: 9 September 2012

Reviewer: Scott Grosse

Reviewer’s report:

The authors were responsive in adding text on page 12 which acknowledges that the model under-predicted the results of the pilot screening program in rates of diagnoses by 30-40%, 0.51 vs. 0.70 per 1000 with tracking and 0.31 vs. 0.50 per 1000 without tracking. The authors followed my previous suggestion in stating that the ICER would probably have been lower if actual data on numbers of cases detected from the pilot screening program had been used. However, the incremental number of children detected as a result of tracking is the same, 0.20 per 1000. Because the costs of follow-up in practice were higher than modeled the ICER may be a bit higher rather than lower.

It would be desirable for the authors to explain briefly in the Discussion why their model under-predicts the numbers of cases detected relative to both the pilot and statewide screening programmes. The model assumes that the probability of a positive diagnosis following a bilateral refer is independent of the probability of a bilateral refer. If a screening test is applied carefully, there should be fewer false positives and the probability that a child who tests positive is a true positive will be higher. The authors in their response indicate that the personnel conducting screening in the pilot screening study were more thorough, which explained the lower referral rate. The findings of the two Bavarian screening programmes support the logical argument that a higher screening referral rate for a given population because of less rigorous screening is associated with a lower positive predictive value. The higher bilateral refer rate in the statewide programme, 1.5-2.1% vs. 0.6% in the pilot programme, was associated with a lower probability of diagnosis conditional on a bilateral refer, 7-8% vs. 12% (my calculations). Consequently, the diagnosis probabilities in the statewide programme were 30-40% lower than in the pilot programme. By multiplying diagnosis probabilities from the statewide screening programme by bilateral refer rates from the pilot programme, the model understated the numbers of cases detected relative to the experience of both programmes.

The authors stated that a strength of their model was that they modeled different probabilities of passing repeat tests or being diagnosed based on whether it was a second, third, or fourth test. However, that required the merging of data from two different programmes which had different referral rates and rates of diagnosis conditional on referral. The authors could have compensated by adjusting or standardizing the observed diagnosis probabilities from the statewide program on the basis of the overall diagnosis probability in the pilot
programme, which would have yielded overall estimates consistent with the pilot programme. However, such an adjustment does not appear to make much difference to the incremental numbers of cases detected due to centralized tracking or to the ICER. Therefore, the results presented in the paper appear to be a reasonable approximation of the cost-effectiveness of centralized tracking for screening done in a way consistent with the pilot screening programme.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I have nothing new to add since the previous review.