International Journal of Technology Assessment in Health Care, 20:4 (2004), 488–492.

Policy relevance of Bayesian statistics overestimated?

Gert Jan van der Wilt
University Medical Centre Nijmegen

Maroeska Rovers
University Medical Centre Utrecht

Huub Straatman
Sjoukje van der Bij
Paul van den Broek
Gerhard Zielhuis
University Medical Centre Nijmegen

Objectives: The observed posterior probability distributions regarding the benefits of surgery for otitis media with effusion (OME) with expected probability distributions, using Bayes' theorem are compared.
Methods: Postal questionnaires were used to assess prior and posterior probability distributions among ear-nose-throat (ENT) surgeons in the Netherlands.
Results: In their prior probability estimates, ENT surgeons were quite optimistic with respect to the effectiveness of tube insertion in the treatment of OME. The trial showed no meaningful benefit of tubes on hearing and language development. Posterior probabilities calculated on the basis of prior probability estimates and trial results differed widely from those, elicited empirically 1 year after completion of the trial and dissemination of the results.
Conclusions: ENT surgeons did not adjust their opinion about the benefits of surgical treatment of glue ears to the extent that they should have done according to Bayes' theorem. Users of the results of Bayesian analyses, notably policy-makers, should realize that Bayes' theorem is prescriptive and not necessarily descriptively correct. Health policy decisions should not be based on the untested assumption that health-care professionals use new evidence to adjust their subjective beliefs in a Bayesian manner.

Keywords: Bayesian statistics, Policy relevance, Otitis media with effusion, Surgery

The Bayesian approach to statistical analysis offers a formal algorithm for revising subjective probabilities in the light of newly available evidence (1). Recent reports have recommended that this approach should be used more frequently in health technology assessment (HTA) studies (2;4). The Bayesian approach has also been criticized, primarily for the role it attributes to prior probabilities, which may be quite arbitrary and devoid of empirical content. However, this can easily be accommodated by using multiple priors, including a skeptical one, or by actually estimating the priors in a sample of professionals from the target group. Much less attention has been devoted to the question of whether posterior probabilities, calculated on the basis of Bayes' theorem, correspond with subjective probabilities actually held by health professionals. For users of HTA reports, notably health policy-makers, this is relevant information when designing and implementing new policies. The aim of our

The authors thank all respondents for their cooperation. The study was supported by a grant from the Investigative Medicine Fund of the National Health Insurance Board of the Netherlands. (Ontwikkelingsgeneeskunde, Ziekenfondsraad)

https://doi.org/10.1017/S0266462304001400 Published online by Cambridge University Press

study was to empirically assess to what extent the Bayesian algorithm is descriptively correct in the case of subjective beliefs among ear-nose-throat (ENT) surgeons regarding the benefits of surgery in children with otitis media with effusion (OME). To this end, we conducted a randomized controlled trial (RCT), which was preceded by elicitation of prior beliefs among ENT surgeons and which was then followed by elicitation of posterior beliefs among ENT surgeons, 1 year after completion of the trial and dissemination of its results. This method allowed us to compare observed (elicited) with expected (calculated) posterior beliefs.

MATERIALS AND METHODS

Elicitation of the Prior Probability Distribution

A postal questionnaire was sent to a random sample of seventy-five registered ENT surgeons in the Netherlands. A reminder was sent out after 3 weeks; all questionnaires were processed anonymously. Respondents were asked to estimate the probability of uneventful, complete recovery of hearing level and of normal language development after a period of 12 months in male children with bilateral OME with the following characteristics: age, 16 months; failed the Ewing hearing test on three consecutive occasions; referred to otorhinolaryngologist at the age of 12 months, during which visit at both sides a flat, type B tympanogram, signs of fluid in the middle ear cavity (otoscopy), and a hearing loss of 30 dB were established. No further action was taken, and the child was scheduled for a return visit. Respondents were asked to rate their probability estimates on a visual analog scale ranging from 0 to 100 percent, both in the case of expectant management and in the case of bilateral tube insertion.

The Trial

Details of the trial have been described elsewhere (3). Briefly, children with bilateral OME persisting for 4 to 6 months were randomly allocated to either watchful waiting or bilateral tube insertion. Follow-up was 1 year and included assessment of hearing level, tympanometry, otoscopy, and assessment of language development. (Reynell and Schlichting test) Children who failed a routine hearing test at three consecutive occasions were referred to one of the thirteen participating ENT outpatient clinics for diagnosis and follow-up. Those who had otoscopically and tympanometrically confirmed bilateral OME, persisting for at least 4 months, and for whom informed consent was obtained were randomly allocated to either watchful waiting (WW group, n = 94) or surgical management (tube insertion, VT group, n = 93). Approval was obtained from Ethical Review Boards of all thirteen participating centers.

Calculation of Prior and Posterior Probability Distribution

The prior distributions for the probability of complete recovery of hearing level and normal language development after a period of 12 months were assumed to follow beta distributions with means and ranges as observed among respondents to the first questionnaire. The results of the trial were then used to calculate the posterior probability distribution, using standard Bayesian theory for proportions with beta distributions.

Elicitation of the Posterior Probability Distribution

One year after completion of the trial, a second postal questionnaire was sent to a random sample of seventy-five registered ENT surgeons. The questionnaire was identical to the one that was used to elicit prior probability distributions. In addition, respondents were asked to indicate whether they were familiar with the results of the trial and whether they agreed with the major recommendations. One reminder was sent after 3 weeks; data were processed anonymously.

RESULTS

Prior Probability Distribution

A total of 52 questionnaires (response rate 69 percent) was returned and could be used for analysis. The distributions of estimated probabilities of complete recovery of hearing after 12 months, in the case of tube insertion and in the case of watchful waiting, are presented in Figure 1.

The results show that, at a time when our trial results were not yet available, ENT surgeons estimated the probability of spontaneous recovery of hearing after a period of 12 months in these children at ca. 50–60 percent. The range of these estimates was quite broad, indicating large variation in opinion among respondents. They were slightly more optimistic about spontaneous normalized language development (median ca. 75 percent, not shown).

Respondents were optimistic about complete recovery of hearing in these children at 12 months after tube insertion (median ca. 95 percent). Ranges were smaller, indicating a greater extent of agreement about the effectiveness of tubes among respondents. The same holds for estimates of normal language development.

Results of the RCT

Mean age at randomization was 19.5 months (se = 1.7) in the VT group and 19.4 months (se = 1.9) in the WW group. Mean hearing level at base line was 46.4 dB (VT group) and 43.3 dB (WW group). Mean Reynell scores (equivalent age – real age) at base line were –0.91 months (VT) and –0.31 months (WW), respectively. At 12-month follow-up, no differences were observed between both groups in hearing level or language development that reached conventional

https://doi.org/10.1017/S0266462304001400 Published online by Cambridge University Press

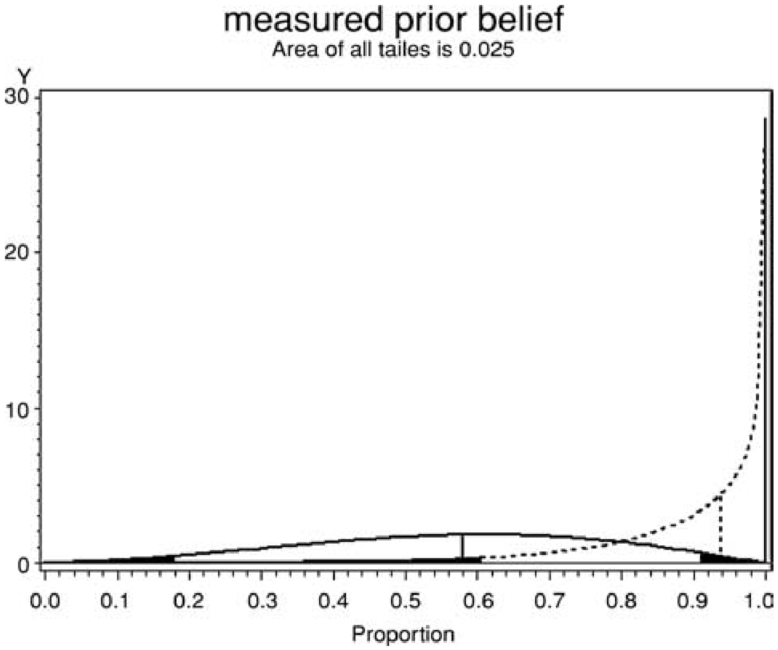

Figure 1. Distributions of estimated prior probabilities of complete recovery of hearing after 12 months in children with otitis media with effusion (OME), in case of watchful waiting (continuous line) and in case of tube insertion (dotted line).

levels of statistical significance (Table 1, Chi-squared test for proportions). The major conclusion of our trial was that in screen-detected children with OME, tube insertion had no demonstrable benefit as compared with watchful waiting.

Dissemination of the Trial Results

Results of the trial were published in national and international journals, and in a thesis that was distributed among all registered Dutch ENT surgeons. In addition, results were presented at regional and national meetings of ENT surgeons.

Posterior Probability Distribution on the Basis of Prior Estimates, Trial Results, and Bayes' Theorem

Posterior probability distributions of complete recovery of hearing after 12 months, in the case of tube insertion and in the case of watchful waiting, calculated on the basis of prior probability distribution, trial results, and Bayes' theorem are presented in Figure 2. The outcome of the trial resulted primarily in a substantial downward adjustment of the probability of successful outcome after tube insertion and in a narrowing down of the range of the distribution.

Observed Posterior Probability Distribution

At the second elicitation procedure, conducted after the dissemination of the trial results, a total of 43 questionnaires (response rate, 57 percent) was returned and could be used for analysis. The distributions of estimated probabilities of complete recovery of hearing after 12 months, in the case of tube insertion and in the case of watchful waiting, are presented in Figure 3. The results show that, at a time when our trial had been completed and its results disseminated, ENT surgeons estimated the probability of spontaneous recovery of hearing after a period of 12 months in these

Table 1. Summary of Trial Results[a]

	Watchful waiting	Tube insertion
Number (proportion) of children with a hearing loss in either ear <35 dB at 12-month follow-up	49 of 81 (60.5%) (95% CI, 50–71%)	65 of 88 (73.9%) (95% CI, 64–82%)
Number (proportion) of children with a delay in language development (of <1 month) at 12-month follow-up:	59 of 86 (68.6%) (95% CI, 59–78%)	52 of 87 (59.1%) (95% CI, 48–69%)

[a] Effectiveness of surgical treatment with ventilation tubes versus watchful waiting in children with otitis media with effusion, in terms of reduction of children with hearing loss and delayed language development at 12-month follow-up.

https://doi.org/10.1017/S0266462304001400 Published online by Cambridge University Press

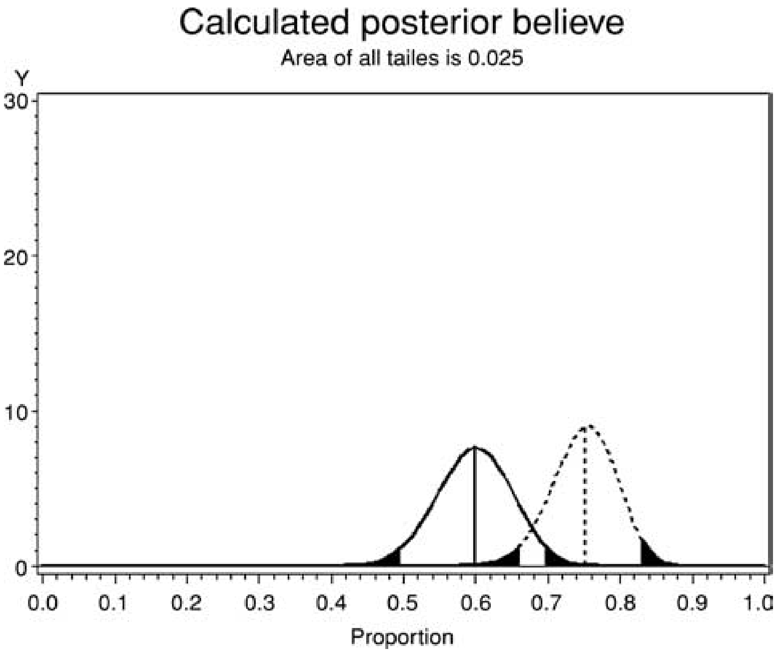

Figure 2. Distributions of calculated posterior probabilities of complete recovery of hearing after 12 months in children with otitis media with effusion (OME), in case of watchful waiting (continuous line) and in case of tube insertion (dotted line).

children slightly higher than before (65 percent). It is remarkable that the range of these estimates was still quite broad, indicating that the trial had little impact in terms of building consensus among respondents. As before, ENT surgeons were slightly more optimistic about spontaneous, undelayed language development (ca. 75 percent, not shown). In this posterior measurement, estimated probabilities of normal hearing and language development in these children at 12 months after tube insertion were very much in line with those measured before the trial. Despite the negative results of our trial, they were still quite optimistic about normal hearing and language development in these children at 12 months after tube insertion (estimated at ca. 95 percent). Again, ranges were still quite broad, indicating that our trial had little impact in terms of consensus building among respondents. Of the respondents, 93 percent indicated that they were aware

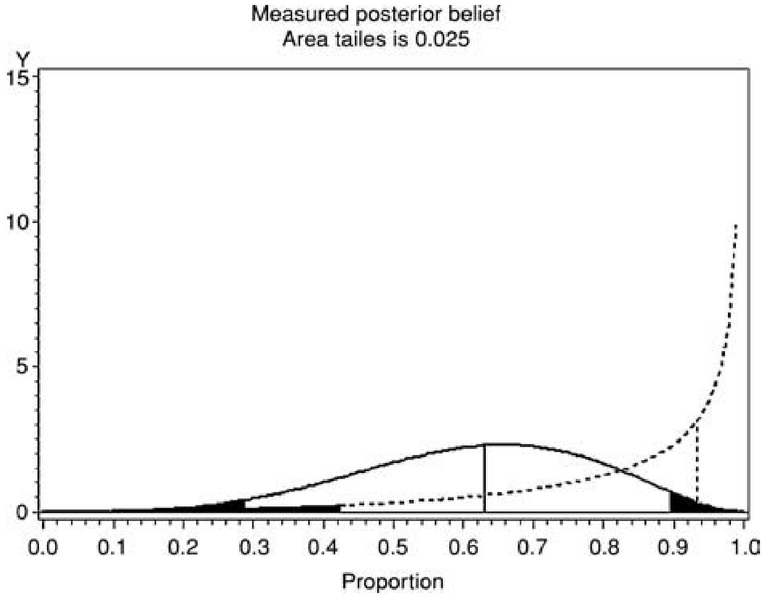

Figure 3. Distributions of observed posterior probability estimates of complete recovery of hearing after 12 months in children with otitis media with effusion (OME), in case of watchful waiting (continuous line) and in case of tube insertion (dotted line).

https://doi.org/10.1017/S0266462304001400 Published online by Cambridge University Press

of the results of our trial. Of these, 38 percent agreed completely with the recommendations, 45 percent agreed with part of the recommendations, and 10 percent did not agree with the recommendations.

DISCUSSION

The most striking result of our study was that a substantial discrepancy existed between observed posterior probability estimates and probability estimates expected on the basis of Bayes' algorithm. Clearly, the health-care professionals who participated in this study, did not revise their probability estimates in the way Bayes' theorem predicts. This finding despite that almost all respondents were aware of the trial results and indicated to agree (partially or completely) with its recommendations. As already indicated, it is important to realize that Bayes' theorem was never intended to be descriptively correct. Rather, its intention is prescriptive: it prescribes how subjective probability estimates should be revised when new, relevant data become available. Why, then, examine its descriptive adequacy? Not, we should emphasize, to prove the Bayesian approach in statistical analysis wrong. Logically, that would make no sense: a prescriptive statement cannot be shown to be incorrect by taking recourse to empirical observations (a version of the famous Humean naturalistic fallacy). Instead, our results should come as a warning to policy-makers. They should not uncritically assume that calculated posterior probabilities exist in the minds of health-care professionals who contemplate new evidence, nor should they design and implement policies on the basis of such an assumption. Our study shows that, at least in the case of subjective beliefs regarding the natural course of OME and the benefits of surgery, such assumption would be clearly unwarranted. Further studies would be needed to assess whether our findings apply to other contexts as well.

In conclusion, we would argue that the frequently posed question of the superiority of Bayesian vs. frequentists approaches in statistical analyses is misguided. For, it tends to overlook the prescriptive nature of the Bayesian approach, which would require an altogether different mode of argumentation. We believe that the Bayesian approach can be fruitfully applied to explore differences and commonalities among actors (in our case: general practitioners, ENT surgeons, policy-makers, and parents) in prior probability estimates. How likely do they consider hypotheses, claiming certain benefits from treatment? What factors, for example, patient characteristics, contexts, etc., seem to affect these probability estimates? Such information might be used in designing a trial and in choosing a strategy for dissemination of its results.

POLICY IMPLICATIONS

The implications for policy-makers is that they should not read the results of Bayesian analyses as evidence of how subjective probabilities, held by health-care professionals, are adjusted in the light of new evidence. To avoid such false readings, the prescriptive nature of the results of Bayesian analyses should be communicated unambiguously: *given prior subjective probabilities held by relevant health-care professionals, this is how these probabilities should be revised in the light of this new piece of new evidence*. Commissioners of health-care technology assessments can demand from researchers to use such phrasing when communicating results of Bayesian analyses.

REFERENCES

1. Berger JO. *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag; 1980.
2. Luce BR, Tina Shih Y-Ch, Claxton K. Bayesian approaches to technology assessment and decision making. *Int J Technol Assess Health Care*. 2001;17:1-122.
3. Rovers MM, Straatman H, Ingels K, et al. The effect of short-term ventilation tubes versus watchful waiting on hearing in young children with persistent otitis media with effusion: A randomized trial. *Ear Hear*. 2001;22:191-199.
4. Spiegelhalter DJ, Myles JP, Jones DR, et al. Bayesian methods in health technology assessment: A review. *Health Technol Assess*. 2000;4:1-130.

https://doi.org/10.1017/S0266462304001400 Published online by Cambridge University Press