The Probability of a Successful Probabilistic Sensitivity Analysis

Tereza Lanitis, MSc, Senior Research Associate, Modeling & Simulation, Evidera
Noemi Muszbek, MSc, Senior Research Scientist, Modeling & Simulation, Evidera
Eszter Tichy, MSc, Research Associate, Modeling & Simulation, Evidera

INTRODUCTION

Probabilistic sensitivity analysis (PSA) is increasingly viewed as a required step in conducting economic evaluations[1] and a formal requirement from agencies such as the National Institute of Health and Care Excellence (NICE).[2] Research on the appropriate ways to structure and conduct PSA and to present results has been prominent in health economics in the last decade[3,4] with a best practices guideline published in 2012 by the joint International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and Society for Medical Decision Making (SMDM) Task Force.[5]

Among other recommendations, the ISPOR-SMDM Task Force and current and previous NICE guidelines recommend that all parameters subject to uncertainty be included in PSA; that the selection of probability distributions be based on sound statistical principles and data, avoiding arbitrary measures; that possible correlations among parameters be considered; and that structural uncertainty should be assessed.[2,5,6] Despite the consistency of these requirements with earlier recommendations,[7] both the implementation and the presentation of PSA in NICE technology appraisals (TAs) have received criticism.[8]

We recently conducted a review on the methods used in all completed, full (excluding patient access submissions), NICE single TAs published in 2013-2014.[9] The aim was to review the most recent approaches adopted for conducting PSA in NICE submissions, assessing whether they conform with the guidelines, if methods have improved since previous criticisms and how PSA ultimately influences decision making.

METHODS

Final appraisal documentations (FADs), evidence review groups (ERG) reports and, where available, manufacturer submissions were reviewed. Data extraction tables were designed to capture:

- The basic characteristics of the TAs
- The methods employed by manufacturers and the ERGs
- Ranges of parameters incorporated in the PSA

- Choices of probability distributions
- Sources of variation
- Assessment of structural uncertainty
- Reporting of limitations
- Overall reporting
- Influence on the ultimate decision

The PSA methods adopted were compared against the NICE reference case from the 2013 NICE guidance.[2] Data extracted by one reviewer was checked by an additional reviewer. (For further detail, please see Lanitis et al. 2014.[9])

RESULTS AND DISCUSSION

Thirty-one TAs were identified, of which 13 were excluded. Excluded TAs were: terminated (4 TAs); multiple TAs (3); revised submissions, including patient access scheme submissions (4 TAs); and lack of publicly available documentation (2). One TA was excluded for two reasons (multiple TA and lack of documents), resulting in 18 TAs included in the review.

Our findings were consistent with an earlier review that criticized the methodology and reporting used in NICE TAs prior to the 2008 and 2013 methodology guidelines.[8] We found that PSAs were heavily criticized by ERGs with at least one methodological issue reported in 84% of cases. Despite these criticisms, PSA results were considered more informative than the deterministic results in 27% of TAs. PSA results were mentioned and reviewed by the committee in almost all FAD reports (84%). However, although potentially discussed in the TA committee meetings, PSA results were only mentioned in the FADs as part of the decision in three TAs (16%).[9]

The main issues that arose from the review were the questions around the choice of distributions; the variation of input parameters; not taking into account the correlations and dependencies between the parameters; the lack of representation of structural uncertainty within PSA; and the appropriate presentation of results.

CHOICE OF DISTRIBUTION

Most TAs did not report in sufficient detail the methods used to populate the PSA and the rationale for the choice of distribution for each parameter and the variation surrounding it.[9] The choice of distribution used for parameters if justified was usually based on conventions, with no additional justification provided.

It is important for the analyst to understand the limitations of the distributions employed in comparison to the nature of the parameter varied. For example, while the gamma and lognormal distributions are bounded by 0, the upper interval of the distribution can go above 1, thus may be inappropriate when the parameter is a risk or probability and thus should be between 0 and 1. In many cases, use of a normal, gamma or lognormal distribution may still remain within the bounds of 0-1 depending on the mean and standard error; however, it is important to test this to ensure the simulated parameter falls within plausible bounds. Usually, use of the beta distribution is recommended for probabilities, as it is a conjugate of the binomial distribution.[3,7] A beta distribution can be parameterized through use of the mean and standard error; however, if the latter is not available, it can be parameterized by using the shape parameter (alpha) as the number of events observed for the preferred outcome (e.g., number of patients experiencing a given outcome) and the scale parameter (beta) as the number of failures of the outcome observed (e.g., number of patients that did not experience a given outcome).

A beta distribution may not be appropriate when the parameter modeled is a rate expressed, for example, as per 100 patient years as its natural bounds do not fall within the 0-1 range of the beta distribution. In such cases, the gamma and lognormal distribution can be considered as they are also bounded by a lower 0 limit. Caution should be exercised in utilizing the normal distribution for such parameters as estimates can go below 0. Limitations associated with the distributions should be evaluated according to each parameter varied. Several publications provide recommendations on the choice of distributions for each type of parameter.[3,5,7]

VARIATION OF INPUT PARAMETERS

In the reviewed TAs, the variation for the parameters was in most cases assumed and not informed by data, with 68% of TAs including at least one parameter where the standard error was assumed to be 10–30% of the mean, with 20% being the most common assumption.[10] In some TAs, the assumed variation was large and extensive, e.g., varying all parameters by 30%, while in others it was minimal and applied only to selected costs. No justification was reported for the size and extent of this variation.

Arbitrary variation of parameters, however, leads to arbitrary results and misrepresentation of the uncertainty. A scatter-plot or cost-effectiveness acceptability curve (CEAC) plotted assuming 20% variation in all parameters may over- or under-estimate uncertainty surrounding the decision. It does not, as intended, reflect the uncertainty of the results and the decision due to parameter uncertainty, but on arbitrary assumptions of uncertainty. Recently developed models tend to have a large number of parameters, and the assessment of uncertainty surrounding them is difficult. In most

cases, however, 95% confidence intervals, standard errors, minimum and maximum, patient numbers or patient-level data are available to inform estimates of variation. Where nothing is available, transparency is required from the analyst regarding the choice of variation, with explicit acknowledgement of the limitations of the analysis.

CORRELATIONS AND DEPENDENCIES AMONG INPUT PARAMETERS

Although guidelines suggest the incorporation of correlation and dependencies between parameters, only one of the 18 reviewed TAs considered this.[9] This is a major limitation in most PSAs as correlation and major dependencies between parameters exist in almost all models. One example is the progression-free survival and overall survival in oncology models. A patient can't progress after they have died, yet independent variation of the survival curves could lead to these curves crossing. In addition, these curves are often varied independently of the comparator curves incorporating the implicit assumption of no correlation between comparators. The assumption of no correlation in these cases can lead to misleading probabilistic results and the overestimation of uncertainty.

Similarly, various other input parameters in a model can be correlated. For example, independent variation of parameters could lead to assigning higher utility values to milder conditions than to more severe conditions in some simulations. Parameters can be correlated using the Cholesky decomposition[7] and methodologies have been proposed to address dependencies such as using z scores to maintain continuity between parameters.[10] The analyst should consider the presence of correlation or dependencies in the model and evaluate their potential influence on the results. If such aspects are not considered in the PSA, appropriate caveats and limitations need to be presented alongside results, including potential scenario analysis of the PSA to gain an understanding of where the true probabilistic estimates may lie.

PRESENTATION OF RESULTS

Several TAs reported mean incremental cost-effectiveness ratios (ICERs) and confidence intervals surrounding the ICER. However, as the ICER is the ratio of the incremental costs and the incremental health benefits, a negative ICER can suggest that the new health technology is less costly (has negative incremental cost) and more effective (has positive incremental health benefit) or it can suggest that the new health technology is more costly (has positive incremental cost) and less effective (has negative incremental health benefit). Similarly, the positive ICER can have opposing interpretations.

Due to this inherent complexity of the ICER having alternative interpretations when falling in different quadrants of the scatterplot,[7] the calculation of confidence or credible intervals around the ICER is not straightforward and there is no consensus on the appropriate methodology. Various methods have been proposed and challenged.[11] Due to these limitations, the scatter-plots and cost-effectiveness acceptability curves can be a more appropriate way of representing uncertainty around the ICER than confidence or credible intervals when observations fall in more than one quadrant.[5,12] It is important for the analyst to understand these limitations before presenting confidence or credible intervals.

ERGs often require reporting of a mean probabilistic ICER. In this case, the abovementioned limitations of the ICER need to be assessed as well. The mean probabilistic ICER can also differ from deterministic results and, if this is the case, it is important to understand the source of the deviation. Recording the values that each parameter takes in the individual simulations together with the results and analyzing the recorded data using regression techniques can prove to be a useful tool in understanding results and drivers of this discrepancy and potential non-linearities. A careful consideration of the number of simulations included in the PSA could also provide solutions. In the reviewed TAs, the median number of simulations used for the PSA was 1,000, varying between 1,000–10,000. However, only one TA provided a rationale for the number of simulations.[9] A formal test of convergence[13] can aid the choice in the appropriate number of simulations required.

CONCLUSION

Compared to the previously conducted review,[8] there seems to be insufficient improvement in conducting PSAs for TAs, with the majority of TAs still not conforming to best practices. Consequently, the interpretation of the probabilistic results is limited by the use of arbitrary variation, methodological

inaccuracies, insufficient reporting and various implicit assumptions as well as the omission of uncertainty in key parameters. As a result, there is a danger that the probabilistic results better represent the underlying assumptions of the analyst than the true impact of parameter uncertainty and can therefore be misleading when informing decision making.

There is considerable scope for improvement when conducting and interpreting PSAs, while the various aspects and challenges in methodology require further research and discussion. In addition, due to these various challenges, the analysts should fully and transparently report on the assumptions required and the limitations of the approaches taken so that they may be taken into account in the decision making.

For more information, please contact Tereza.Lanitis@evidera.com, Noemi.Muszbek@evidera.com or Eszter.Tichy@evidera.com.

References

[1] Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value Health.* 2012; 15(6):796-803.

[2] National Institute for Health and Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal. London, UK: National Institute for Health and Care Excellence; 2013. Available at: http://publications.nice.org.uk/pmg9. Accessed Sept. 29, 2014.

[3] Briggs AH, Ades AE, Price MJ. Probabilistic Sensitivity Analysis for Decision Trees with Multiple Branches: Use of the Dirichlet Distribution in a Bayesian Framework. *Med Decis Making.* 2003; 23(4):341-350.

[4] Ades AE, Claxton K, Sculpher M. Evidence Synthesis, Parameter Correlation and Probabilistic Sensitivity Analysis. *Health Econ.* 2006; 15(4):373-381.

[5] Briggs AH, Weinstein MC, Fenwick EA, et al. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health.* 2012; 15(6):835-842.

[6] National Institute for Health and Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal. London, UK: National Institute for Health and Clinical Excellence; 2008.

[7] Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation.* 2006, pp. 256. ISBN13: 9780198526629; ISBN10: 0198526628 Paperback. Oxford: OUP.

[8] Andronis L, Barton P, Bryan S. Sensitivity Analysis in Economic Evaluation: An Audit of NICE Current Practice and a Review of Its Use and Value in Decision-Making. *Health Technol Assess.* 2009; 13(29):iii, ix-xi, 1-61.

[9] Lanitis T, Muszbek N, Tichy E. Uncertainty in Uncertainty: A Review of Probabilistic Sensitivity Analysis Conducted in Health Technology Appraisals. Paper Presented at: Society of Medical Decision Making (SMDM), 15th Biennial European Meeting, 2014; Antwerp, Belgium.

[10] Schauer DP, Eckman MH. The Use of z Sscores in Probabilistic Sensitivity Analyses. *Med Decis Making.* 2014; 34(3):403-406.

[11] Wang H, Zhao H. A Study on Confidence Intervals for Incremental Cost-effectiveness Ratios. *Biom J*. 2008 Aug; 50(4):505-514.

[12] Al MJ. Cost-effectiveness Acceptability Curves Revisited. *Pharmacoeconomics*. 2013 Feb; 31(2):93-100.

[13] Robinson S. *Simulation: The Practice of Model Development and Use:* John Wiley & Sons, Ltd.; 2003.