

questionnaire to ascertain whether the research question is still relevant and if there have been any developments to the evidence since publication of the HTA. The input from these stakeholders is collated and taken to HTW's Assessment Group to decide whether or not the HTA needs updating. If the Assessment Group decides that re-assessment of a HTA is warranted, HTW perform an updated literature search to inform the re-assessment.

**Results.** The HTA re-assessment SOP developed by HTW was approved by the organization's Assessment Group. At the time of writing this abstract, HTW sent questionnaires to stakeholders of three HTAs which had HTW guidance published three years ago, and were therefore due routine consideration for re-assessment as detailed in our SOP. HTW also received a request from a clinician for a more recent HTA to be considered for re-assessment as they believed the evidence-base had changed since original publication. These questionnaires have been collated and will be taken to an upcoming Assessment Group to decide whether HTW should proceed with the re-assessments.

**Conclusions.** HTW has developed a consistent process for HTA re-assessment, which ensures that HTAs done by HTW remain current and relevant to best serve the population and health and care providers in Wales. By utilizing expertise from HTA stakeholders and HTW Assessment Group members, an informed decision can be made as to whether a HTA warrants re-assessment after three years following publication or sooner if requested.

## PP92 Should Missing Data Be Multiply Imputed Prior To Longitudinal Linear Mixed-Model Analyses In Trial-Based Economic Evaluations?

Ângela J Ben ([angelajben@gmail.com](mailto:angelajben@gmail.com)),  
Johanna M van Dongen, Mohamed El Alili,  
Martijn W Heymans, Jos W R Twisk,  
Janet L MacNeil-Vroomen, Maartje de Wit,  
Susan E M van Dijk, Teddy Oosterhuis and  
Judith E Bosmans

**Introduction.** For the analysis of clinical effects, multiple imputation (MI) of missing data was shown to be unnecessary when using longitudinal linear mixed-models (LLM). It remains unclear whether this also applies to cost estimates from trial-based economic evaluations, that are generally right-skewed. Therefore, this study aimed to assess whether MI is required prior to LLM when analyzing longitudinal cost-effectiveness data.

**Methods.** Two-thousand complete datasets were simulated containing five time points. Incomplete datasets were generated with 10 percent, 25 percent, and 50 percent missing data in costs and effects, assuming a Missing At Random (MAR) mechanism. Statistical performance of six different methodological strategies was compared in terms of empirical bias (EB), root-mean-squared error (RMSE), and

coverage rate (CR). Six strategies were compared: (i) LLM (LLM), (ii) MI prior to LLM (MI-LLM), (iii) mean imputation prior to LLM (M-LLM), (iv) complete-case analysis prior to seemingly unrelated regression (CCA-SUR), (v) MI prior to SUR (MI-SUR), and (vi) mean imputation prior to SUR (M-SUR). To evaluate the impact on the probability of cost-effectiveness at different willingness-to-pay [WTPs] thresholds, cost-effectiveness analyses were performed by applying the six strategies to two empirical datasets with 9% and 50% of missing data, respectively.

**Results.** For costs and effects, LLM, MI-LLM, and MI-SUR performed better than M-LLM, CCA-SUR, and M-SUR, as indicated by smaller EBs and RMSEs, as well as CRs closer to the nominal levels of 0.95. However, even though LLM, MI-LLM, and MI-SUR performed equally well for effects, MI-LLM and MI-SUR were found to perform better than LLM for costs at 10 percent and 25 percent missing data. At 50 percent missing data, all strategies resulted in relatively high EBs and RMSEs for costs. In both empirical datasets, LLM, MI-LLM, and MI-SUR all resulted in similar probabilities of cost-effectiveness at different WTPs.

**Conclusions.** When opting for using LLM for analyzing trial-based economic evaluation data, researchers are advised to multiply impute missing values first. Otherwise, MI-SUR may also be used.

## PP93 The Impact Of Using Different EQ-5D Scoring Methods On Cost-Utility Outcomes: A Simulation Study

Ângela J Ben ([angelajben@gmail.com](mailto:angelajben@gmail.com)),  
Johanna M van Dongen, Aureliano P Finch,  
Mohamed El Alili and Judith E Bosmans

**Introduction.** Patients' EQ-5D health states are preferably valued using country-specific value sets. If value sets are not available, crosswalks may be used to estimate utility values. However, up until now the impact of using crosswalks instead of value sets on cost-utility outcomes remains unclear.

**Methods.** Trial-based cost-utility data were simulated for four conditions (depression, low back pain, osteoarthritis, and cancer), three levels of disease severity (mild, moderate, and severe), and three treatment effect sizes (small, medium, and large), resulting in 36 scenarios. For all scenarios, utility values were estimated using four scoring methods (EQ-5D-3L value set, EQ-5D-5L value set, 3L-to-5L crosswalk, and 5L-to-3L crosswalk) for three countries (the Netherlands, the United States, and Japan). Mean utility values, quality-adjusted life years (QALYs), incremental QALYs, and cost-utility outcomes (incremental cost-effectiveness ratios [ICER], probabilities of cost-effectiveness at willingness-to-pay [WTP] thresholds) were compared between value sets and crosswalks.

**Results.** Differences between value sets and crosswalks ranged from -0.33 to 0.13 for mean utility values, from -0.18 to 0.13 for QALYs, and from -0.01 to 0.08 for incremental QALYs. Because of the small differences in incremental QALYs, ICERs between scoring methods were considerably different. For small effect sizes, at a WTP of EUR