making, reduced patient benefits and less efficient use of public resources. Within individual agencies, prioritization is also a business function that must balance the need to plan for and manage the allocation of resources with the need to provide expeditious advice to decision makers and adapt quickly to changing circumstances. This research describes a transparent and responsive framework for selecting health technology assessment (HTA) reports, by assessing the potential incremental effectiveness and ability of alternative technologies. The framework feeds into an in-depth expert group discussion, which also considers operational issues such as the extent of the advice required to inform the decision, data availability and costs associated with the assessment. We also describe a software tool that helps carry out the process, as well as measures for quality assurance and ongoing performance evaluation.

PRM240

GOAL ATTAINMENT SCALING – A USEFUL INDIVIDUALIZED CLINICAL OUTCOME MEASURE

Jones M, Khara N, Fichna P, Gandhi P

Bridgeman Medical, London, UK

Goal Attainment Scales (GAS) capture outcomes relevant to individual patients and provide “real-world” outcome measurement. This abstract will describe the background to GAS, an operationalization, advantages and limitations. Traditional outcome measures assess a standardized set of questions regardless of their relevance to each patient. GAS overcomes these weaknesses because it is an individualised assessment of patient goals. GAS is unique in that it is patient-derived, minimising the potential for important technologies to be missed and providing a useful resource for HTA agencies facing similar issues. Topics are identified through a mix of routine horizon scanning, a formally convened advisory group consisting of the major decision makers from within the publicly funded health system and informal business intelligence gathering. Screening is carried out to eliminate technologies that are clearly unsuitable and provisionally grade all remaining candidates according to three principal criteria: 1) clinical impact (potential to affect a stated outcome and potentially incremental effect and availability of alternatives); 2) economic impact (incremental costs and potential disruptive effect on how services are currently organised) and 3) policy impact (link to decision-making and factors) which is personal to the patient and a key feature of the national health care agenda). The screening process feeds into an in-depth expert group discussion, which also considers operational issues such as the extent of the advice required to inform the decision, data availability and costs associated with the assessment. We also describe a software tool that helps carry out the process, as well as measures for quality assurance and ongoing performance evaluation.

PRM241

AVOIDING AND IDENTIFYING ERRORS AND OTHER THREATS TO THE CREDIBILITY OF HEALTH ECONOMIC MODELS

Zapletalova C, Scott J, Bruce A

University of Sheffield, Sheffield, UK

Health economic models have become the primary vehicle for undertaking economic evaluation and are used in various health care jurisdictions across the world to inform decisions about the use of new and existing health technologies. Models are required because a single source of evidence, such as a randomised controlled trial, is rarely sufficient to provide all relevant information about the expected costs and health consequences of all competing decision alternatives. Whilst models are used to synthesise all relevant evidence, they also contain assumptions, abstractions and simplifications. By their very nature, all models are therefore “wrong.” Whilst the presence of imperfect evidence provides the impetus for developing models, it is also the reason why we can never fully validate them. As such, the interpretation of estimates of the cost-effectiveness of health technologies requires careful judgement about the degree of confidence that can be placed in the models from which they are drawn. The presence of a single error or inappropriate judgement within a model may lead to inappropriate decisions, an inefficient allocation of health care resources and ultimately suboptimal outcomes for patients. This study tests out a taxonomy of threats to the credibility of health economic models. The taxonomy segregates threats to model credibility into three broad categories (1) unequivocal errors, (2) violations and (3) matters of judgement, and maps these across the main elements of the model development process. These three categories of threats to model credibility are defined as a framework of evidence criteria for judging correctness, the degree of force with which such criteria can be applied, and the means by which potential threats can be handled. A range of suggested processes and techniques for avoiding and identifying these threats is put forward with the intention of prospectively increasing the credibility of any given model.

PRM242

IMPACTS OF EPRO DATA COLLECTION MODE SELECTION ON PATIENT INCLUSION

Hofbauer T, Ross J, Wade M, Rothrock T

Alma Clinical Technologies, Souderstown, PA, USA

OBJECTIVES: The Electronic Patient Reported Outcome (ePRO) data collection mode selection has a significant impact on patient accrual and a product’s real world effectiveness. Understanding causes for HTE is required to inform decisions about the use of new and existing health technologies. Models are required because a single source of evidence, such as a randomised controlled trial, is rarely sufficient to provide all relevant information about the expected costs and health consequences of all competing decision alternatives. Whilst models are used to synthesise all relevant evidence, they also contain assumptions, abstractions and simplifications. By their very nature, all models are therefore “wrong.” Whilst the presence of imperfect evidence provides the impetus for developing models, it is also the reason why we can never fully validate them. As such, the interpretation of estimates of the cost-effectiveness of health technologies requires careful judgement about the degree of confidence that can be placed in the models from which they are drawn. The presence of a single error or inappropriate judgement within a model may lead to inappropriate decisions, an inefficient allocation of health care resources and ultimately suboptimal outcomes for patients. This study tests out a taxonomy of threats to the credibility of health economic models. The taxonomy segregates threats to model credibility into three broad categories (1) unequivocal errors, (2) violations and (3) matters of judgement, and maps these across the main elements of the model development process. These three categories of threats to model credibility are defined as a framework of evidence criteria for judging correctness, the degree of force with which such criteria can be applied, and the means by which potential threats can be handled. A range of suggested processes and techniques for avoiding and identifying these threats is put forward with the intention of prospectively increasing the credibility of any given model.

PRM243

CLINICAL OUTCOME ASSESSMENT (COA) INSTRUMENT SCORING: THE VALIDITY AND PRECISION OF UNWEIGHTED SUMMARY SCORES VERSUS IRT WEIGHTED SCORES, AND THE ADDED VALUE OF IRT STANDARD ERRORS

Coop CD, Lenderking WR

1Adelphi Values, Boston, MA, USA, 2Evidera, Lexington, MA, USA

COA development experts in recent years have given thought to the psychometric evaluation of instruments and their ability to detect meaningful differences between treatments. The scoring of the instruments, however, has received less attention, with various approaches sometimes suggested without a clear preference or justification. The score is ultimately used for evaluating patient outcomes and treatment efficacy and is what requires validation, so this seems like a significant omission. We examine the traditionally accepted unweighted summary score approach and compare it to the more complex IRT weighted scoring to evaluate if the gain in precision justifies the increased scoring complexity. Precision may differ depending on whether the score is close to the mean of the population or closer to the extreme ends of the distribution. Simulated data are used for this comparison to evaluate if the precision of the scores differs depending on the location of the score and the common conceptions of validity and the consequences of this. Additionally, we recognize that the reliability of a scale is likely to be variable across the range of its scores. With that in mind, we consider an approach to comparing means across groups that minimizes the standard error of each individual IRT score into the model. By using the IRT standard errors, we can adjust for the different levels of uncertainty associated with ranges of scores along the scale, ultimately providing us greater confidence in the group comparison results.

PRM244

EVALUATION OF ESTIMATORS OF TREATMENT EFFECT IN OBSERVATIONAL STUDIES

Pafes D1, Lipkovich I1, Kadziola Z2

1Eli Lilly and Company, Indianapolis, IN, USA, 2Eli Lilly Regional Operations GmbH, Vienna, Austria

There is increasing scrutiny of pharmaceuticals on their value proposition as well as their real world effectiveness once they are commercially available. There are many challenges in producing valid and reliable estimates of real world effectiveness. A major challenge is assessing a product’s effectiveness relative to what patients may respond differently to a treatment (i.e., identifying groups of patients who might benefit from treatment) by sub-grouping individuals (e.g., sub-group identification methods). Assessing HTE is critical to understanding differences that may exist between the efficacy observed in randomized clinical trials and the potential real world effectiveness. Understanding causes for HTE is required for correct attribution of any observed difference between efficacy and effectiveness to the product versus other sources (e.g., patient behavior). Not recognizing and accounting for HTE will confound assessment of a product’s performance, which ultimately affects its acceptance and use by payers, physicians, and patients. Failure to define and incorporate subgroups is a frequent criticism of systematic evidence reviews and comparative effectiveness research reports. However, the analytical methods for identifying subgroups that are both challenging due to many known statistical issues (e.g., limited statistical power, multiplicity adjustments) Real world data exacerbates the analytical challenges due in part to biases (e.g., selection bias) and issues (e.g., data quality) inherent in the data. We will present results from a simulation experiment that compared and validated several subgroup methods developed to address these data and analytical issues. We simulated 22 permutations of subgroups with known identification criteria and treatment effects to determine the performance of the methods.