PRM40
A LOOK AT PREVIOUS AND CURRENT METHODS USED TO COLLECT PATIENT-REPORTED OUTCOMES INFORMATION

OBJECTIVES: Patient reported outcomes (PROs) have become an important component of many clinical studies. The use of ePRO as a data collection method can alleviate the potential burden experienced by patients and/or sites. The purpose of this survey study was to capture current PRO data collection trends and summarize these findings side-by-side with results from a previous PRO data collection survey.

METHODS: Industry professionals were invited to complete a web-based survey fielded in late 2011 and early 2012. This survey included questions on professional demographics, experience using PROs (and ePROs) by study type and experience with ePRO technologies. Responses were analyzed descriptively.

RESULTS: To date, 54 industry professionals completed the 2011-2012 survey. Fifty nine percent of respondents reported pharmaceutical company involvement, 20% medical device, and 19% other (6%). While 49% of respondents in the 2010 survey had previous PRO study experience, 60% of respondents in the current survey had previous PRO experience. The proportion of respondents with prior ePRO experience, however, was similar across the two surveys (51% in 2010 and 54% in 2011-2012). Hand-held device (tablet, PDA) was the most common ePRO technology (71% in 2011, 64% in 2010), followed by interactive voice response (47% in 2011, 60% in 2010), and interactive web-response (29% in 2011, 51% in 2010). Among those with prior ePRO experience in 2011 and 2010, respectively, 59% and 86% strongly agreed/agreed they would use ePRO in future studies. Among those who never used ePROs, 58% in 2011 and 50% in 2010 indicated they would likely use ePROs in future studies.

CONCLUSIONS: Results from this survey suggest that ePRO use continues to gain momentum in industry professionals. These findings were based on a limited sample size. Future surveys should be administered to allow future trends in ePRO use to be observed over time.

PRM41
TRANSLATION AND CULTURAL ADAPTATION OF THE LANGUAGE DEVELOPMENT SURVEY

OBJECTIVES: The Language Development Survey (LDS) assesses children’s word comprehension and vocabulary and provides an accurate picture of a child’s developing language when completed by a parent or guardian. The LDS contains a list of 310 basic words. The person completing the questionnaire is asked to circle those words their child says spontaneously. Translations already existed in over ten languages. Twelve additional languages were translated, including Esperanto and Italian. A direct translation of some source words was not possible as some items are unknown in the target country, e.g. pretzel, spaghetti, Sesame St. Therefore, it was necessary to find a conceptually equivalent source word.

METHODS: Two approaches were adopted. An initial translatability assessment was carried out to identify problematic words. Equivalent source words were suggested. Further problematic words were identified during the translation and cognitive-debriefing process. Alternatives were suggested either during the translation stages or by the parent/guardian during the cognitive debriefing interview. Final wording was agreed on through discussion with the lead in-country translator and instrument developers.

RESULTS: A number of cultural adaptations were made. For all Indian languages, ‘cracker’ was translated as ‘papadom’ (a thin, crisp Indian cracker) and ‘cracker’ was translated as ‘papadom’ (a thin, crisp Indian cracker). When making cultural adaptations, care was taken to ensure that the translation was not altered in ways that would affect its meaning. For example, the word ‘pretzel’ was translated as ‘pretzel’ because it is a well-known term in the target country. The word ‘spaghetti’ was translated as ‘spaghetti’ because it is also well-known in the target country. However, the word ‘Sesame St’ was replaced with ‘Tom and Jerry’ in French. ‘saucisse’ was suggested as an alternative for ‘hot dog’ during cognitive debriefing. CONCLUSIONS: When translating a patient-reported outcome (PRO) the aim is to produce a translation that is conceptually equivalent to the source text. In some cases, cultural adaptation is essential. Translation of the LDS is an excellent example of this methodology and the translations are now available for use in multi-national studies.

PRM42
PATIENT PREFERENCES FOR REMINDERS IN CLINICAL TRIALS: IMPROVING PATIENT COMPLIANCE

OBJECTIVES: This session intends to identify patients’ preferences for receiving reminder messages when participating in clinical trials. Results of a survey on patient experiences will be reported. Recommendations for developing reminder strategies will be provided with the intention of enhancing the patient’s experience and compliance.

METHODS: An internet survey was administered to patients (in December 2010) who participated in at least one clinical trial with patient diaries in the past two years. The survey focused on patients’ perceived experiences and preferences with patient diaries/ePRO, and how patient diary methods could be improved. This session will focus on the questions relating to patient reminders—specifically relating to preferences for how reminders were sent/received, activities for which reminders were found most useful, frequency of reminders, etc.

RESULTS: Reminder methods patients preferred most were email and text messages. Two areas patients preferred to be reminded about were any action required of them as part of their clinical trial participation and when and how often they wanted to receive reminders, the most frequent responses included whenever there was new information related to their trial participation and once per day. Patients also provided suggestions for reminders in future trials.

CONCLUSIONS: The patient population was tuned in to what they are being told and how often they want in terms of reminders, these perspectives should be accounted for to enhance the patient journey and compliance. If industry implements the reminder strategy wrong (for instance, annoying patients by reminding them too frequently), that may actually impact compliance negatively. Remind patients when necessary/not too often. Remind patients in ways they will be able to best receive/notice them. Appropriate use of reminders drives compliance and incorporating patient preferences will not only improve compliance rates, but will also enhance the patient’s experience.

RESEARCH ON METHODS – Statistical Methods

PRM43
ROBUSTNESS OF CONFIDENCE INTERVALS FOR RARE EVENTS

OBJECTIVES: Accurately estimating the upper bounds of confidence intervals for rare events such as hospitalization or death is an important activity in safety studies and post-market research. Confidence intervals, however, can be subject to considerable variation based upon the overall sample size and total number of observed events. This has led to a challenging convention that a minimum of 2 or 3 events are needed for computing meaningful confidence intervals. The objective of this study was to quantify the variation of the upper bound of confidence intervals for a binomial proportion in the setting of rare events. METHODS: Clogg-Pearson confidence intervals were constructed for sample sizes ranging from 50 to 1000, and numbers of events from 0 to 5. The robustness of the confidence interval was evaluated by calculating additional confidence intervals assuming: 1) one observed event than in the original sample and, 2) that the proportion of events is equal to the upper bound of the confidence interval for the original sample. RESULTS: With sample sizes of 50, 100, 200, 500 and 1000, the upper bound of the confidence intervals were 0.1371, 0.0740, 0.3570, 0.1446 and 0.7205, respectively, with 2 observed events in the original sample; 0.3570, 0.1446, 0.7205, and 26.40%, 13.94%, 7.16%, 2.91% and 1.47%, respectively; and, 26.40%, 13.94%, 7.16%, 2.91% and 1.47%, respectively, when the proportion of events was equal to the upper bound of the confidence interval for the original sample with 2 events. CONCLUSIONS: The upper bounds of confidence intervals for rare events vary greatly with sample sizes and the numbers of events observed when the sample size is small. A minimum of 500 subjects is optimal for constructing confidence intervals for rare events, even if 2 events or less are observed.

PRM44
POST-RANDOMIZATION IN RETROSPECTIVE ANALYSIS USING THE GENERALIZED MULTINOMIAL LOGIT FOR PROPENSITY SCORE GENERATION

OBJECTIVES: To develop and test a three-way propensity score matching algorithm to provide pseudo-randomization of subjects into three groups to allow for comparable groups in a retrospective study. METHODS: Logistic regression using the generalized multinomial logit linking function was used to calculate estimates of the propensity score: the probability of having received three putatively inter-changeable drugs from demographic (Race, Gender, Age) and comorbidities (Charlson Comorbidities Index) in a large, retrospective database. The most costly drug changeable drugs from demographic (Race, Gender, Age) and comorbidities (Charlson Comorbidities Index) in a large, retrospective database. The most costly drug was used as the reference group, and the probability of each treatment group having received the reference drug was retained as the propensity score. In the initial forward stepwise process, 23,912, 4,789, and 2,308 matched triplets were constructed for patients who had the propensity score: the probability of having received three putatively inter-changeable drugs from demographic (Race, Gender, Age) and comorbidities (Charlson Comorbidities Index) in a large, retrospective database. The most costly drug changeable drugs from demographic (Race, Gender, Age) and comorbidities (Charlson Comorbidities Index) in a large, retrospective database. The most costly drug was used as the reference group, and the probability of each treatment group having received the reference drug was retained as the propensity score. In the initial forward stepwise process, 23,912, 4,789, and 2,308 matched triplets were constructed for patients who had the propensity score. Random subsets of 1/4 and 1/10 the original sample were constructed for the purpose assessing multi-group propensity score matching (PSM) effectiveness in constructing comparable groups via pseudo-randomization with varying sample sizes (10,000, 20,000, 50,000). CONCLUSIONS: PSM was used as the reference group, and the probability of each treatment group having received the reference drug was retained as the propensity score. In the initial forward stepwise process, 23,912, 4,789, and 2,308 matched triplets were constructed for patients who had the propensity score. Random subsets of 1/4 and 1/10 the original sample were constructed for the purpose assessing multi-group propensity score matching (PSM) effectiveness in constructing comparable groups via pseudo-randomization with varying sample sizes (10,000, 20,000, 50,000).
variable. CONCLUSIONS: Logistic regression using a generalized multinomial logit link appears to provide a good propensity score from which pseudo-randomization into three groups can be performed in a retrospective sample.

PRM45 NETWORK META-ANALYSIS OF STUDIES WITH OUTCOMES AT MULTIPLE TIME POINTS USING FRACTIONAL POLYNOMIALS
Veitch da Silva MC1, Jansen JP2
1Mapi Consultancy, Boston, MA, USA, 2Mapi Consultancy / Tufts University School of Medicine, Boston, MA, USA

OBJECTIVES: Network meta-analysis of randomized controlled trials (RCTs) are often based on one effect measure per study. However, many studies have data available at multiple time points. Furthermore, not all studies might have measured the outcomes at the same time points. As an alternative to network meta-analysis based on the results at one time point, a network meta-analysis method is presented that allows for the simultaneous analysis of outcomes at multiple time points.

METHODS: The development of outcomes over time of interventions compared in a RCT are modeled with fractional polynomials, and the difference between the parameters of these polynomials within a trial are synthesized across studies with a Bayesian network meta-analysis. RESULTS: The proposed models are illustrated with an analysis of RCTs evaluating interventions for osteoarthritis of the knee. Fixed and random effects first and second order fractional polynomials were evaluated. CONCLUSIONS: Network meta-analysis with models where the treatment effect is represented with several parameters using fractional polynomials can be used to simultaneously analyze results at multiple follow-up times that are not consistent across studies.

PRM46 CONTROLLING FOR MULTIPlicITY IN Pursuit OF A Pro-BASED Label WHEN multiple PROs ARE ASSESSED
Cole JC
Covance Market Access Services, Inc., San Diego, CA, USA

OBJECTIVES: The FDA's final Guidance for industry on patient reported outcome (PRO) use in support of labeling claims was issued in December, 2009. In their Guidance, the FDA noted that a study's endpoint model must consider the hierarchy of multiple endpoints, including how PROs used for a label claim fit into this hierarchy. Whereas most studies implement a basic sequential gatekeeping process to articulate their hierarchy, this may place some potential labels at risk. Researchers should be knowledgeable of the various ways falsifiability error is influenced and how best to control for it with an informed multiplicity plan as part of their endpoint model.

METHODS: Outcomes from previously published literature were examined for the influence of various falsifiability error issues and related multiplicity controls, including analytic issues, gatekeeping, and precision alpha control (vs. Bonferroni or Hochberg).

RESULTS: In a study with one clinical and three PRO outcomes, A Bonferroni correction resulted in just one significant result. A gatekeep between primary and secondary outcomes resulted in two significant results. The proposed Gamma model, despite being more theoretically justifiable, is of simpler linear model seems to estimate the attributable costs sufficiently well.

CONCLUSIONS: The simplest linear model seems to estimate the attributable costs sufficiently well. The proposed Gamma model, despite being more theoretically justifiable, is of questionable benefit. Further investigation is needed for refining the Gamma model and selecting appropriate measures of model assessment and comparison.

PRM48 DIAGNOSTIC TOOLS FOR THE ASSESSMENT OF THE UNDERLYING MODEL ASSUMPTIONS IN THE STUDY OF HEALTH CARE COSTS
Jansen JP
Thomson Reuters, Boyd, MD, USA

OBJECTIVES: It is a common practice to use a log link and assume a gamma distribution when performing regressions of health care costs as an outcome on a set of potential predictors. In many circumstances, this approach is reasonable and performs well; however, do circumstances exist where these assumed model characteristics are untenable? If so, do simple diagnostic procedures exist that can assess the assumptions? A simulation study is performed to explore if the assumptions of a gamma model are justified as an outcome?

METHODS: Application of residual analyses available in common statistical software packages (e.g., SAS) afford practitioners the ability to graphically and analytically evaluate whether the choice of a link is appropriate in a given cost model regression scenario. These same tools can also assist with the assessment of the model fit. RESULTS: The study will illustrate the evidence of where the choice of a generalized linear model with a log link and an assumed gamma distribution are defensible and where these assumptions are not met and may lead to errors in subsequent inference. CONCLUSIONS: With the use of these readily available diagnostic procedures found in common software packages, it is possible to easily evaluate whether underlying model assumptions are tenable and if the choice of a simpler, more common approach may actually demonstrate higher fidelity to its underlying model assumptions than the commonly used generalized linear model with a gamma distribution and a log link.

PRM49 HOW TO PRESENT THE PROBABILITY OF BEING THE BEST TREATMENT IN THE CONTEXT OF A BAYESIAN NETWORK META-ANALYSIS OF PARAMETRIC SURVIVAL CURVES?
Cope S1, Veitch da Silva MC2, Jansen JP2
1Mapi Consultancy, Boston, MA, USA, 2Mapi Consultancy / Tufts University School of Medicine, Boston, MA, USA

OBJECTIVES: Increasingly, network meta-analysis (NMA) of published survival data are based on parametric survival curves as opposed to reported hazard ratios to avoid relying on the proportional hazards assumption, which may not be valid. One approach to a Bayesian approach to NMA is that the probability of being the best treatment out of all those compared can be calculated. This directly supports decision-making. However, in the context of survival analysis multiple options are available. METHODS: Based on a case study in oncology, the probability that each treatment is best as a function of survival proportions over time, as well as expected survival over time are presented. RESULTS: Based on the following underlying assumptions: 1) the hazard over time, 2) the cumulative hazard over time, 3) the survival proportions over time, 4) the expected survival over time, 5) the expected survival at maximum follow-up, 6) expected survival when all patients have died, and 7) median survival. RESULTS: Since the NMA of survival curves results in changing hazard and survival estimates over time for the compared interventions, calculations of the probability that a certain treatment is best varies with the different alternatives. With methods 1-4 the probability that a certain treatment is best will vary as a function of follow-up, which provides relevant information. With methods 5-7 only one probability of being the best is obtained for each treatment, which is easier to understand. Method 1 does not directly relate to the survival proportion, which makes it not very intuitive. Method 7 discards a lot of information. CONCLUSIONS: Different approaches to present the probability of each treatment being the most efficacious treatment for finding the one obtained with a NMA of survival curves have pros and cons. The probability that a certain treatment is best as a function of survival proportions over time, as well as expected survival over time seem the most useful and intuitive.

PRM50 META-REGRESSION MODELS TO ADDRESS HETEROGENEITY AND INCONSISTENCY IN NETWORK META-ANALYSIS OF SURVIVAL OUTCOMES
Cope S1, Jansen JP2
1Mapi Consultancy / Tufts University School of Medicine, Boston, MA, USA

OBJECTIVES: As an alternative to network meta-analysis (NMA) of survival data based on the single constant hazard ratio (HR), NMA with a multi-dimensional treatment effect were introduced recently. With these models the HR is modeled as a function of time, and violations of the transitivity assumption are less likely. Bias is still present, however, if there are systematic differences in effect modifiers across comparisons. The objective of this paper is to extend multidimensional NMA models for survival data with treatment-by-covariate interactions to adjust for confounding bias. METHODS: By means of an example network of randomized controlled trials evaluating different interventions for melanoma, three different approaches for the analysis of overall survival (OS) are compared: 1) NMA assuming a constant HR between treatment and control group for each study; 2) a two-dimensional NMA model assuming survival outcomes are described by a Weibull function; and 3) an extension of method 2 with treatment-by-covariate interactions to adjust for systematic differences across studies. RESULTS: Methods 2 and 3 are more reliable than method 1 as long as the data to be compared are independent of the data with the model constant HR (approach 1). Adding treatment-by-covariate interactions for the scale parameter of the two-dimensional NMA models reduced inconsistency. CONCLUSIONS: Adding treatment-by-covariate interactions to multi-dimensional NMA models for published survival curves is worthwhile to explain systematic differences across studies and reduce inconsistencies. An additional advantage is that heterogeneity in survival data can be addressed.