tions were based in Asia (k=4), North America (k=3), Africa (k=3) and Europe (k=2). The target population of these publications were the general public (k=7), patients (k=4) or both (k=1). A slight majority of the 12 primary evidence publications (k=7) and a larger majority of the 5 secondary evidence publications (k=4) reported a SP bias on the results of the BG. Various parameters such as male gender, higher education and higher income levels were, in some instances, associated with higher WTP amounts. Other factors analyzed were patients vs. general population, and the location of the study. Association between these factors and the occurrence of starting point bias is examined and will be reported.

CONCLUSIONS: There is evidence in the literature of a SP bias on the results of BGs, without however a full consensus on the matter. Further research is warranted in order to evaluate the conditions under which such bias appears.

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 work at pharmaceutical companies, biotech (20%), medical device (9%), and 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, 58% 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 like to use ePROs in future studies.

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

TRANSLATION AND CULTURAL ADAPTATION OF THE LANGUAGE DEVELOPMENT SURVEY

Congdon W1, Wong D1, Sundenaraj R2

University of Vermont, Burlington, VA, USA

OBJECTIVES: The Language Development Survey (LDS) assesses children’s word combinations 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 words their child says spontaneously. Translations already existed in over ten languages when completed by a parent or guardian. The LDS contains a list of words their child says spontaneously. Translations already existed in over ten languages.

METHODS: A number of cultural adaptations were made. For all Indian languages, ‘cracker’ was translated as ‘papadom’ (a thin, crisp Indian cracker) and ‘pizza’ as translated as ‘dosa’ (a type of Indian pancake). This was decided before languages, ‘cracker’ was translated as ‘papadom’ (a thin, crisp Indian cracker) and ‘pizza’ as translated as ‘dosa’ (a type of Indian pancake). This was decided before languages.

CONCLUSIONS: Alternatives were suggested either during the translation stages or by the parable groups in a retrospective study.

OBJECTIVES: Accurately estimating the upper bounds of confidence intervals for rare events such as hospitalization or death is an important activity in safety studies as it is a routine research. Confidence intervals, however, for rare events are 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: Cope-per-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 more 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 13.71%, 7.04%, 3.57%, 1.44% and 0.72%, respectively, with 2 observed events in the original sample; 16.55%, 8.52%, 4.32%, 1.74% and 0.87%, respectively, (3 observed events); 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. Similar trends were seen when using other numbers of observed 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.

DEVELOPMENT SURVEY

Griffiths W1, Wong D1, Sundenaraj R2

Oxford Outcomes, Oxford, Oxfordshire, UK, 3Bryn Mawr College, Bryn Mawr, PA, USA

OBJECTIVES: When translating a patient-reported outcome (PRO) the translation and cultural adaptation of the language development survey was also a challenge. Translations already existed in over ten languages. Translators were interviewed to examine the translation process. Alternatives were suggested either during the translation stages or by the parable groups in a retrospective study.

RESULTS: A number of cultural adaptations were made. For all Indian languages, ‘cracker’ was translated as ‘papadom’ (a thin, crisp Indian cracker) and ‘pizza’ as translated as ‘dosa’ (a type of Indian pancake). This was decided before languages.

CONCLUSIONS: Alternatives were suggested either during the translation stages or by the parable groups in a retrospective study.

OBJECTIVES: Appropriate use of reminders drive compliance and incorporating patient preferences will not only improve compliance rates, but also enhance the patient’s experience

RESEARCH ON METHODS – Statistical Methods

PRM43

ROBUSTNESS OF CONFIDENCE INTERVALS FOR RARE EVENTS

Su T1, Mendelsohn A1, Kim J2, Gremmen R3

Quintiles Outcome Sciences, a Quintiles Company, Cambridge, MA, USA, 2Quintiles Global Consulting, London, Reading Berkshire, UK, 3Quintiles Global Consulting, Rockville, MD, USA

OBJECTIVES: Accurately estimating the upper bounds of confidence intervals for rare events such as hospitalization or death is an important activity in safety studies as it is a routine research. Confidence intervals, however, for rare events are 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: Cope-per-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 more 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 13.71%, 7.04%, 3.57%, 1.44% and 0.72%, respectively, with 2 observed events in the original sample; 16.55%, 8.52%, 4.32%, 1.74% and 0.87%, respectively, (3 observed events); 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. Similar trends were seen when using other numbers of observed 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

PSEUDORANDOMIZATION IN RETROSPECTIVE ANALYSIS USING THE GENERALIZED MULTINOMIAL LOGIT FOR PROPENSITY SCORE GENERATION

Wilson SM, Mayne T

Eviva Clinical Research, Minneapolis, MN, USA

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 (Char-son 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 randomization. Generalized multinomial logit 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 (Char-son 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 randomization.

RESULTS: 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. Two methods were used: randomization with using calipers ranging from 8 digits to one digit of propensity score. Assessment of among-group differences before and after PSM were conducted using Chi-square tests for categorical variables and GLM analysis, with difference scores and their confidence intervals for continuous variables. The sample sizes matched triplets were retained. There were no significant differences among matched triplets on any continuous variables.