Methods: The cohort comprised 4784 patients receiving one of two knee replacement devices at 37 sites. Mean LOS was 3.0 days (SD = 1.3; range = 1 to 27). The number of patients ranged from 3 to 579 per site, and the percent in one device group from 10% to 90%. We conducted simulations using the existing design to determine the empirical Type I error ($\alpha = .05$) and precision of the effect estimates. We evaluated linear-Gaussian (LG), log-Poisson (LP), log-truncated-Poisson (LTP) and log-quasi-Poisson (LQP) link-error families due to under-dispersion. We used GLMM, GEE and FE to account for correlation within site.

Results: With regard to the LG link-error family, the Type I error was unbiased for GLMM (.046) and FE (.05), but slightly anti-conservative for GEE for model based (.069) and empirical (.081) standard errors (SEs). With regard to the LP family, the Type I error was very conservative for GLMM (.0055) and FE (.0095), while GEE was slightly anti-conservative (.07) for model SEs and empirical SEs (.082). For the LTP family, the Type I error was conservative for GLMM (.018) and FE (.012) and for the LPQ family, the Type I error was unbiased for FE (.055). With regard to precision of the estimates, the GLMM models resulted in the lowest mean squared error (MSE) and the FE models the highest MSE.

Conclusions: Overall the GLMM models and FE models performed best with the GLMM models showing the lowest MSE. GEE performed best with the model based SEs but was slightly anticonservative.

232. Detecting Suicidal Outcomes: A Power Analysis

Richard S. Swain, Lockwood G. Taylor and Andrew D. Mosholder

Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, United States

Background: Suicidal outcomes, including ideation, attempt, and completed suicide, are an important drug safety issue, as evidence by boxed warnings for some antidepressants and other medications. However, suicidal outcomes are difficult to ascertain in observational study.

Objectives: To estimate the ability of clinical trials to detect association with suicidal outcomes.

Methods: Stone et al. (2009) performed a meta-analysis of the association between antidepressants and suicidal outcomes which encompassed 372 clinical trials including 99,231 participants with 15,505 patient-years of follow-up finding increased risk of suicidal outcomes in patients aged 25 and younger and a protective effect among patients aged 65 and older. We used data from Stone et al., assuming average study sample size, follow-up, and event rates from the high-risk subgroup (under age 25), to perform power and sample size calculations to estimate the ability of trials included in the meta-analysis to detect statistically significant associations between antidepressants and suicidal outcomes.

Results: An average trial included in Stone et al. had only 5.1% power to detect an incidence rate ratio (IRR) of 2.0 for suicidal behavior and would have required an IRR of 10.9 to detect a signal with 90% power. A trial would have needed 1,304 person years of follow-up, compared to an average follow-up of approximately 42 person years per study observed by Stone et al., to detect an IRR of 2.0 with 90% power. The detection of smaller effects would require exponentially more follow-up.

Conclusions: Our calculations demonstrate that even in an at-risk psychiatric population many clinical trials are not powered to measure associations with suicidal outcomes. Trials including subjects from broader age groups would have less power to detect increased risk of suicidal outcomes among younger patients and less power to detect the strong protective effect observed among patients over age 65. Our findings underline the importance of developing new methods to measure suicidal outcomes in pharmacoepidemiological studies.

233. Methods of Defining the Noninferiority Margin: A Systematic Review

Turki A. I. Althunian[1], Anthonius de Boer[1], Olaf H. Klungel[1], Widya N. Insani[1] and Rolf H. H. Groenwold[2]

[1] *Division of Pharmacoepidemiology and Clinical Pharmacology, Graduate School of Life Sciences, Utrecht University, Utrecht, Netherlands;* [2] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands*

Background: Noninferiority trials are conducted to investigate whether the efficacy of the test drug is not worse than the active comparator based on a pre-

defined noninferiority margin. There is no consensus on a preferred method for defining the noninferiority margin, and previous studies showed that the rationale for its choice is often not reported.

Objectives: To investigate how the noninferiority margin is defined in the published literature, and whether its reporting has changed over time.

Methods: A systematic PubMed search was conducted for all published randomized, double-blind, noninferiority trials from January 1, 1966, to February 6, 2015. The time trend of defining the margin since the first published noninferiority trial was analyzed using Poisson regression analysis. The impact of the 2010 US Food and Drug Administration (FDA) draft guidance for noninferiority clinical trials on the choice of the margin, and the impact of the extension of the Consolidated Standards of Reporting Trials (CONSORT) Statement on the reporting of noninferiority margins were analyzed using generalized estimating equations to account for the clustering of the margins within articles.

Results: We included 275 articles, which account for 283 trials and 328 noninferiority margins. The rationale for the choice of the margin was not reported for 191 margins (58.2%). The under-reporting of the rationale for the choice of the margin has not improved neither since the first published noninferiority trial in this review in 2000 (P = 0.86), nor since the publication of the extension of the CONSORT Statement in 2006 (P = 0.96). The other 137 margins were mainly defined based on the historical evidence of the active comparator (n = 56) or subjectively based on expert opinions (n = 46). There was a 3.5-fold increase in the use of the fixed-margin method, the recommended method by the FDA to define the margin, after the publication of the FDA draft guidance (from 2.6% to 9%, P = 0.04).

Conclusions: Margins in noninferiority trials are poorly reported and this has not improved over recent years. Authors, reviewers, and editors need to take notice of reporting this critical information to allow for better judgment of noninferiority trials.

234. Real Time Aggregate Clinical Safety Monitoring Methodology-Evaluation of Multiple Quantitative Methods

Syed S. Islam, Ran Gao and Mondira Bhattacharya

Pharmacovigilance and Patient Safety, Abbvie, North Chicago, IL, United States

Background: To date, there is no consensus on how to assess and interpret periodically the aggregate safety events in individual and combined studies during clinical development.

Objectives: Develop and test statistical inference methodology to assist the safety assessment teams evaluate and interpret safety data from both blinded and unblinded randomized clinical trials.

Methods: Blinded periodic aggregate analysis was done on a hypothetical clinical trial dataset using sequential probability ratio test (SPRT) and Bayesian critical value approach considering both binomial and Poisson models. We unblinded the same trial data and used Bayesian method to evaluate the probability of exceeding pre-specified risk difference or risk ratio.

Results: Safety boundary using Wald SPRT and Bayesian methods were created for blinded analyses using both Binomial and Poisson models. Unblinded analyses were done using Bayesian Method.

Conclusions: This presentation summarizes a number of methodologies that can be applied to aggregate safety data which will enhance early identification of signals in clinical trials.

235. Improving Short-Term Mortality Prediction Using Timing of Acute Comorbidities During Lookback

Henry T. Zhang[1], Leah J. McGrath[2], Alan R. Ellis[3], Richard Wyss[4], Jennifer L. Lund[1] and Til Stürmer[1]

[1]*Epidemiology, University of North Carolina at Chapel Hill, Chapel HillNC, United States;* [2]*RTI Health Solutions, Research Triangle Park, NC, United States;* [3]*Social Work, North Carolina State University, Raleigh, NC, United States;* [4]*Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States*

Background: A defined lookback period is often used to assess comorbidities for confounding control. Hazard functions for the outcome of interest can vary after an acute comorbid event, therefore individuals will be at different points on that curve as follow-up starts. Predictive models may need to account for the timing of events during lookback but little attention has been given to this issue.