size of an RTBCEA that balances its cost and the value of perfect information that would remain after its completion.

PRM211

DISPROPORTIONALITY MEASURES USED IN SIGNAL DETECTION: AN ASSESSMENT ON PHARMACOVIGILANCE ADVERSE EVENT REPORTING SYSTEM DATA

Quattrini G[1], Zambon A[1], Simoni L[2], Fiori G[3]

[1]University of Milano-Bicocca, Milan, Italy, [2]MEDIDATA SRL, MODENA, Italy, [3]Medineos Observational Research, Modena, Italy

OBJECTIVES: Quantitative analysis of spontaneous adverse drug reaction reports is increasingly used in drug safety research. Signals are detected by disproportionality measures (DM). Different types of DMs are available: a debate is ongoing on which performs better. The aim was to evaluate the sensitivity on identification of signals with known safety profiles of incretin drugs. **METHODS:** Adverse Events (AE) reported to FDA Adverse Event Reporting System (FAERS) between 2005 and 2014 were included. To evaluate the impact of warning actions, two separated analysis were conducted: restricted to the time before and after a regulatory action. We selected 20 AEs for each drug, half as "positive controls" (with a known causal association from literature) and half as "negative controls" (with no evidence in literature). Proportional Reporting Ratio (PRR), Reporting Odds Ratio (ROR), Bayesian Confidence Propagation Neural Network Analysis (BCPNN) and Gamma-Poisson Shrinker (GPS) were calculated on 120 combinations of AE-drugs. Correlation between sensitivity and number of AEs per year reported for each drug was calculated to evaluate the effect of number of reports on sensitivity. **RESULTS:** The number of reports analyzed in 2014 was 1,934,607. After warning action, PRR showed a sensitivity of 0.42 (0.29; 0.55), ROR 0.55 (0.42; 0.68), BCPNN 0.53 (0.40; 0.66) and GPS 0.23 (0.13; 0.36). Analog findings were observed before warning actions. The concordance of signals identification was good for all pairwise comparison between DMs (>0.56). The correlation varied among 0.49 (for PRR) and 0.82 (for ROR and BCPNN) after warning action. **CONCLUSIONS:** The sensitivity of measures was low (<0.6), without impact of warning actions. ROR and BCPNN showed the most elevated values of sensitivity not allowing to determine a clear superiority of neither frequentist nor bayesian DMs. As expected, the positive correlation suggests the presence of a strong impact on sensitivity of higher number of AEs reported.

PRM212

CORRECTING FOR SWITCHING TO SECOND LINE TREATMENT IN THE SURVIVAL ANALYSIS: AN EXAMPLE OF THE USE OF INVERSE PROBABILITY OF CENSORING WEIGHTED ANALYSIS

Sanchez L, Luaces P, Crombet T

Center of Molecular Immunology, Havana, Cuba

OBJECTIVES: Randomized clinical trials of biological products are commonly analyzed with an intent-to-treat (ITT) approach, whereby patients are analyzed in their assigned treatment group regardless of actual treatment received. The ensuing switch to second line treatment disturbs randomization, compromising the utility of clinical data. The main objective of the study is to discuss how the statistical procedure of inverse probability of censoring weighted (IPCW) analysis may be used in this situation **METHODS:** The first step in the IPCW analysis is to predict the probability of switch on the basis of each patient's baseline characteristics, such as age, sex, race, the time from diagnosis to randomization, and biological markers by fitting a logistic regression model. Finally, Overall Survival (OS) is analyzed with the censored data set and observations weighted by the inverse of the predicted probability of censoring. The method was illustrated for one clinical trial evaluating the effect of one monoclonal antibody combined with gemcitabine in patients with advanced pancreatic cancer. **RESULTS:** A total of 192 patients were randomized (average age 63.6 ±10 years; 60% male; 69% ECOG PS 0), Of 96 patients enrolled in the nimotuzumab arm (OSAG), 40 patients (41.7%) switch to second line, while in the Placebo arm 41(42.7%) switch to the second line. The hazard ratio and 95% CI for OS was 0.83 (95% CI, 0.62 to 1.12) for ITT, was 0.75 (95% CI, 0.51 to 1.12) for the censored analysis, and for IPCW was 0.82 (95% CI, 0.67 to 0.98). **CONCLUSIONS:** The switch to second line treatment affects efficacy results of the ITT analysis of the nimotuzumab plus Gemcitabine therapy in pancreatic cancer. Additional IPCW analysis indicates that the benefit of the molecular antibody, nimotuzumab, is greater than that reflected by the ITT estimate.

PRM213

THE USE AND ACCEPTANCE OF META-ANALYSIS OF SURVIVAL OUTCOMES TO SUPPORT HEALTH TECHNOLOGY ASSESSMENT SUBMISSIONS IN THE UK

Batson S, Hudson P, Webb N, Greenall G

Abacus International, Oxfordshire, UK

OBJECTIVES: Meta-analysis of survival data is most commonly performed by using the individual summary statistic hazard ratio (HR) from each study as an appropriate measure of effect. The aim of this study was to (i) assess the literature reporting on the use of meta-analysis of parametric survival curves, an alternative novel method for the evidence synthesis of survival data; (ii) assess technology appraisals (TAs) submitted to the National Institute for Health and Care Excellence (NICE) to determine whether this novel method has been accepted within UK Health Technology Assessment (HTA). **METHODS:** Embase, Medline and the Cochrane Library were searched to identify publications reporting on novel statistical methods. The NICE website was interrogated to identify oncology TAs, the associated Evidence Review Group (ERG) and final appraisal document (FAD) as published between 2011 and 2014 which reported novel statistical methods. **RESULTS:** Four publications reported on the use of meta-analysis of parametric survival curves. Of the most recent 60 NICE TAs, a single TA included the use of meta-analysis of parametric survival curves. **CONCLUSIONS:** Meta-analysis of survival curves has been developed to address limitations which arise where the proportional hazards assumption does not hold for survival curves; however, to date the method has not been validated by independent statisticians and currently there is no guidance from NICE regarding inclusion of this methodology within HTA. Conclusions on the acceptance of this methodology by NICE cannot be made due to limited examples.

PRM214

META-ANALYSIS TO SUPPORT TECHNOLOGY SUBMISSIONS TO HEALTH TECHNOLOGY ASSESSMENT AUTHORITIES: CRITICISMS BY NICE AND EVIDENCE REVIEW GROUPS IN THE UK

Batson S, Webb N, Greenall G

Abacus International, Oxfordshire, UK

OBJECTIVES: The quality of evidence used in manufacturers' submissions to health technology assessment (HTA) bodies is an important factor for the success of technology appraisals (TA). Indirect comparisons (IC) and network meta-analyses (NMA) are used in health policy decisions via the clinical effectiveness evidence in HTA submissions to the National Institute for Health and Care Excellence (NICE). The aim of this study was to: (i) assess the use of ICs and NMAs in HTA submissions to NICE; (ii) identify criticisms of ICs and NMAs in TAs generated by NICE and the Evidence Review Group (ERG); (iii) provide key insights and recommendations to minimise criticism of an IC or NMA in future HTA submissions. **METHODS:** The NICE website was interrogated to identify both TAs and the associated ERG/final appraisal document reports published from January 2013 to June 2015 in any therapeutic setting. **RESULTS:** A large proportion of the TAs analysed included ICs or NMAs. Common criticisms were related to the identification of data and the study selection for inclusion, study heterogeneity and the inadequate reporting of methods and analyses. **CONCLUSIONS:** The majority of criticisms of evidence synthesis submitted to NICE were related to issues around the primary evidence included in the analyses rather than the statistical methods of the analyses. To avoid many of the criticisms identified in this study a transparent approach to the reporting of the ICs and NMAs (and systematic review) is recommended.

PRM215

ALTERNATIVES TO WINBUGS FOR NETWORK META–ANALYSIS

Stephenson M[1], Fleetwood K[2], Yellowlees A[2]

[1]University of Guelph, Guelph, ON, Canada, [2]Quantics Consulting Ltd, Edinburgh, UK

OBJECTIVES: Network meta-analysis (NMA) of clinical trial outcomes is usually based on Bayesian statistics and hence requires software for Monte Carlo Markov chain (MCMC) sampling. The most common choice of software for NMA is WinBUGS, in part because there is a large body of WinBUGS code for NMA in the literature; for example in the NICE Decision Support Unit (DSU) Technical Support Documents (TSD) on evidence synthesis. However, WinBUGS is slow, difficult to use, and better, more efficient, options may be available. This project aimed to identify and evaluate alternatives to WinBUGS. **METHODS:** We identified candidate alternatives for evaluation via journal articles and websites. We performed an initial examination against a set of criteria including (a) compatibility with Windows (b) speed (c) ease of use (d) publication quality graphics (in the system or ease of linking with an external program such as R) (e) ability to handle large datasets within MCMC software and (f) cost. We ranked the candidates and then performed a validation of the top-ranked choice by running a set of examples found in the NICE DSU TSDs to ensure matching of the results. **RESULTS:** We found nine potential alternatives to WinBUGS: OpenBUGS; JAGS; GeMTC; LaplacesDemon; Mamba; PyMC; SAS PROC MCMC; MCMCpack; Stan. Stan was the most promising and we tested it against a number of datasets used from the NICE guidance. **CONCLUSIONS:** We found Stan, an open source program for Bayesian statistical inference, to be the best option for NMA. Stan provides an excellent balance of model flexibility, allowing for manual user specification, and is easily integrated with R for producing publication quality graphics. We found it straightforward to learn because it is accompanied by an extensive user manual and provides helpful error messages. We recommend that NMA practitioners should consider Stan as an alternative to WinBUGS.

PRM216

USING MACHINE LEARNING TECHNIQUES TO CLASSIFY OECD COUNTRIES ACCORDING TO HEALTH EXPENDITURES

Cinaroglu S

Hacettepe University, Ankara, Turkey

OBJECTIVES: Machine learning techniques are used for analysis of large complex datasets. Classification is an important part of machine learning applications, it defines groups within population. There are many different methods which are compare results to determine the best classification. In this study we aim to use machine learning techniques to classify OECD countries according to their health expenditures. **METHODS:** Different algorithms can be use in machine learning techniques; C4.5 which is an extension version of ID.3 algorithm and CART algorithm are one of these most commonly use algorithms. Random Forest which constructs a lot of number of trees is one of another useful technique for solving both classification and regression problems. In this study we compare classification performances of different decision trees (C4.5, CART) and Random Forest which was generated by using 50 trees. We perform this prediction model for predicting OECD countries health expenditures for the year 2011. We use number of independent variables for this prediction. These are; life expectancy at birth, number of physicians, number of hospitals, hospital aggregates, alcohol consumption, GDP per capita, perceived health status and immunization. We use AUC results and ROC curve graph for performance comparison. **RESULTS:** As a result of this study it was seen that classification performances of machine learning techniques were good (AUC≥0.90) and Random Forest [50] classification performance results much higher [AUC=0.98] than CART (0.95) and C4.5 (0.90). Decision tree graphs shows that GDP per capita was a variable which has more information gain for predicting health expenditures. **CONCLUSIONS:** To conclude according to our knowledge this is the first study applied machine learning classification methods to health expenditure data. Future studies will compare classification performances of Random Forest using different types of health expenditure datasets, different predictor variables while increasing the number of trees in the forest.

PRM217
THE USE OF PROPENSITY SCORE MATCHING DOES NOT PROTECT AGAINST REGRESSION ARTIFACTS (REGRESSION TOWARDS THE MEAN)
Caron C[1], Wasser T[2], Eisenberg D[2]
[1]Reading Hospital, West Reading, PA, USA, [2]HealthCore Inc., Wilmington, DE, USA
OBJECTIVES: Propensity Score Matching (PSM) is a common method in many retrospective studies to control for differential treatments. PSM controls for variables where patients are selected for one treatment over another based on aspects of their care that are unknown to the researcher or not a part of the study. This study uses simulated data comparing two cohorts within a population treated for a common psychiatric disorder. Data are analyzed to determine if regression artifacts (RA) are present in the data, uncontrolled by PSM. RA in this context are Type I errors. **METHODS:** Variables commonly used to diagnose patients with Major Depression were simulated: Age, Gender, Ethnicity, Global Assessment of Functioning, Beck Depression and Beck Anxiety scores. Distributions of N=100,000 were simulated for each variable using population values. From these distributions, samples of n=100, n=250 and n=500 were drawn based on typical values that would be seen in a patient with Major Depression. The outcome measure Dependent Variable was the score on the Beck Depression scale, using success of treatment values from 10-15 percent, and correlated with the pretest score using Chomsky's decomposition. PSM was used on a ratio of 1:1. Analysis methods were group and paired t-tests as well as a difference in difference analysis at the end of the study. **RESULTS:** Type I error occurred in each simulation and were correlated with sample size. RA, leading to Type I error were more common at lower sample sizes, in excess of 70%, to a minimum of 54% for n=500. **CONCLUSIONS:** This study demonstrates that RA occur in basic experiments designed to specify treatment effects. Researchers who use PSM methods need to be aware of situations where RA are likely to occur. Standard statistical controls for RA are being tested to see if they correct for RA and Type I error when PSM is used.

PRM218
APPLICATION OF SIMPLE IMPUTATION TECHNIQUES FOR MISSING PAIRWISE CONTRASTS FROM MULTI-ARM TRIALS WHEN USING FREQUENTIST NETWORK META ANALYSIS
Petto H[1], Brnabic A[2], Kadziola Z[1], Belger M[3]
[1]Eli Lilly Regional Operations GmbH, Vienna, Austria, [2]Eli Lilly, Sydney, Australia, [3]Eli Lilly and Company Ltd, Windlesham, UK
OBJECTIVES: When conducting frequentist (fixed effects or random effects) network meta-analysis (NMA), input data is usually required in contrast form. In practice, multiple-arm trials are quite common and results for only the contrast relative to one treatment group are presented. However, some frequentist NMA require all possible pairwise treatment effects and standard errors combinations. While the missing effect sizes can still be directly derived, additional assumptions about covariances are needed to calculate standard errors. **METHODS:** Simple imputation techniques are used for substituting the standard errors of the missing comparisons and this has been applied to both simulated data as well as a real world data example. After imputation data is analyzed using standard frequentist NMA, incorporating multi arm studies by the method described in Rücker (2015). **RESULTS:** We derive simple imputations techniques by (1) assuming independence between contrasts, (2) estimating missing co-variances from the available contrasts in the multi arm trials and (3) from the other two arm studies in the network. Comparable results to networks including all pairwise contrasts can be obtained, especially if only few contrasts are missing in multi arm studies and if variances of the comparisons are not too different. In the first case, even (1) can give acceptable results. If variances differ, but are similar to that from two arm studies then (3) might be preferable over (2). **CONCLUSIONS:** Our results suggests that from a practical point of view, simple imputation techniques might be useful tools for incorporating multi arm trials with incomplete pairwise contrasts into frequentist NMA, although limitations need to be carefully considered. Rücker G: Network meta-analysis, electrical networks and graph theory. Research Synthesis Methods, 2012, 3, 312–324.

PRM219
INDIRECT COMPARISONS IN BENEFIT ASSESSMENT
Kühnast S[1], Schiffner-Rohe J[2], Rahnenführer J[1], Leverkus F[2]
[1]Technical University of Dortmund, Dortmund, Germany, [2]Pfizer Deutschland GmbH, Berlin, Germany
OBJECTIVES: With the Act on the Reform of the Market for Medicinal Products (AMNOG) in Germany, pharmaceutical entrepreneurs must submit a dossier demonstrating additional benefit of a new drug compared to an appropriate comparator. Underlying evidence was planned for registration purposes and therefore often does not meet the appropriate comparator as defined by the Federal Joint Committee (G-BA). For this reason AMNOG allows indirect comparisons (ICs) to assess the extent of additional benefit. This study evaluates the applicability of available IC methods in several situations common to benefit assessment in oncological indications. **METHODS:** An extensive literature search on available statistical methods for performing ICs is performed. Additionally, benefit dossiers containing ICs are analyzed regarding the applied methodology. We use simulation studies to evaluate and compare adjusted (Bucher) and unadjusted methods regarding their properties under different circumstances. **RESULTS:** Adjusted ICs are deemed to be "state of the art". Due to their requirements they are, nevertheless, often not applicable. In most cases reasons are lacking comparability of the trials, e.g. concerning the common comparator, the study population and the study design. Simulations of Hazard Ratios for endpoints overall survival and progression free survival were performed considering various "extents of additional benefit" according to IQWiG criteria. Starting with a setting of identical studies we stepwise modified study population and various attributes in study design. Finally the common comparator was omitted. Discrepancies between ICs and true values are compared graphically and on the basis of statistical measures. **CONCLUSIONS:** ICs imply a set of requirements to be able to derive valid statements. Prerequisites for adjusted ICs are often not met as necessary studies and publications are not available. With respect to the progress of benefit assessment and the subsequent price negotiation it would be helpful having alternatives with acceptable properties in order to estimate the extent of additional benefit.

PRM220
THE USE OF INTERQUARTILE DEVIATION IN ESTABLISHING DELPHI PANEL CONSENSUS: A PRIORITIZATION OF INTRAVENOUS IMMUNOGLOBULIN UTILIZATION
Orange J[1], Lennert B[2], Rane P[3], Eaddy M[2]
[1]Texas Children's Hospital, Houston, TX, USA, [2]Xcenda, Palm Harbor, FL, USA, [3]University of Houston, Houston, TX, USA
OBJECTIVES: To use consensus-building methodologies to prioritize disease states for intravenous immunoglobulin (IVIG) utilization while considering disease severity and alternative therapeutic options. **METHODS:** A 7-member expert panel independently ranked 50 disease states across 2 domains: (1) Disease severity (DS) (1=immediately life-threatening, 2=life-threatening, 3=life-modifying, 4=other) and (2) the perceived efficacy of therapeutic alternatives (TA) (1=none, 2=low, 3=medium, 4=high). An interquartile deviation of $\leq$0.5 was used to determine consensus for disease states within each domain. Disease states reaching consensus across both domains were ranked according to a 4x4 algorithmic scale to establish priority. **RESULTS:** The panel reached consensus on the severity of all diseases states; however, 11 of the 50 disease states did not reach consensus on the availability of alternative therapeutic options. No disease state was designated as being immediately life-threatening without an available alternative therapeutic option (DS1TA1), while 3 disease states (X-linked agammaglobulinemia, common variable immunodeficiency, primary immunodeficiency with absent B-cells) were designated as life-threatening with no therapeutic alternatives (DS2TA1). The priority distribution of disorders based on the algorithm is as follows: DS1TA1=0, DS1TA2=1, DS1TA3=1, DS1TA4=1 DS2TA1=3, DS2TA2=4, DS2TA3=3, DS2TA4=1 DS3TA1=0, DS3TA2=7, DS3TA3=14, DS3TA4=0 DS4TA1=0, DS4TA2=0, DS4TA3=3, DS4TA4=1 **CONCLUSIONS:** The application of interquartile deviation in establishing consensus across two 4-point Likert scales resulted in prioritizing 80% of disease states where IVIG can be used. Additional consensus-building rounds will be needed to prioritize the remaining disease states.

PRM221
NETWORK META-ANALYSIS FOR HEALTH TECHNOLOGY SUBMISSIONS WORLDWIDE: A REPORT CHECKLIST FOR NETWORK META ANALYSIS BEST PRACTICES GLOBALLY
Batson S, Greenall G
Abacus International, Oxfordshire, UK
OBJECTIVES: Network meta-analysis (NMA) represents an important and developing method for Health Technology Assessment (HTA). The aim of this study was to review submission guidelines issued by HTA bodies worldwide and produce a checklist for reporting NMA within HTA submissions globally. **METHODS:** The web-based repository of country-specific pharmacoeconomic guidelines maintained by ISPOR was reviewed in January 2015. Guidelines from a number of countries providing sufficient guidance for the use of NMA in HTA submissions were identified and independently reviewed. **RESULTS:** Following review of the available guidance from a number of countries, a single common checklist was developed. The checklist included recommendations relating to five main themes: data; statistical methodology; analyses performed; presentation of results; and technical issues. **CONCLUSIONS:** This reporting checklist provides practical support to health technology manufacturers enabling them to assess the suitability of NMA reports in meeting the requirements of global HTA bodies. In addition, this checklist can be seen as a valid quality tool to critically appraise the reporting of NMAs within HTA.

RESEARCH ON METHODS – Study Design

PRM222
TRANSPARENCY AND REPRODUCIBILITY OF SUPPLEMENTARY SEARCH METHODS IN NICE SINGLE TECHNOLOGY APPRAISAL MANUFACTURER SUBMISSIONS
Witkowski MA, Aldhouse N
Abacus International, Bicester, UK
OBJECTIVES: Systematic reviews (SRs) form an important part of National Institute for Health and Care Excellence (NICE) single technology appraisal (STA) manufacturer submissions. To minimise publication bias when conducting SRs, supplementary searches should be conducted, and should follow the same principles of transparency and reproducibility as database searches. This study aimed to evaluate supplementary search methods used in NICE STA manufacturer submissions. **METHODS:** NICE STAs published between 2011 and 2015 were reviewed. Supplementary search details from manufacturer submissions and related critique from corresponding evidence review group (ERG) reports were extracted. Searches were deemed reproducible if the minimum amount of information required to reproduce searches was reported. **RESULTS:** Of 126 STAs identified, 80 were excluded: appraisal reviews/updates (n=20); appraisal terminated (n=12); no full submission available (n=9); appendices (containing search methods) not published online (n=39). Of 46 included manufacturer submissions, 28 reported conference searches, of which 24 provided enough information for searches to be reproduced. Twenty-one reported clinical trials registry searches, but only seven provided enough information to reproduce these. Thirty-six reported conducting other manual searches, including: manufacturer internal databases (n=24); reference lists (n=20); regulatory body websites (n=11); other websites (n=5); internal experts (n=2). Evidence review groups critiqued omission of supplementary searches in 8 of 18 submissions which lacked searches of conference proceedings, and in 8 of 25 submissions which did not report searching clinical trial registries. The evaluation methods differed between ERGs. **CONCLUSIONS:** Principles of transparency and reproducibility were not fol-