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-