The use of propensity score matching does not protect against regression artifacts (regression towards the mean)

PRM217

Carlos C1, Wasser T2, Eisenberg D1

1Reading Hospital, West Reading, PA, USA, 2HealthCore 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. 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 pre-simulated to be a Beta variable using prior 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 measured with the met 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 controlled 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 in the medical literature and that PSM methodology 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.

APPLICATION OF SIMPLE IMPUTATION TECHNIQUES FOR MISSING PAIRWISE CONTRASTS FROM MULTI-ARM TRIALS WHEN USING FREQUENTIST NETWORK META-ANALYSIS

PRM218

Peto H1, Ibrahim A2, Kadziola Z2, Belger M

1Eli Lilly Regional Operations GmbH, Vienna, Austria, 2Eli Lilly, Sydney, Australia, 3Eli Lilly and Company Ltd, Windlesham, UK

OBJECTIVES: When conducting frequentist (fixed effects or random effects) network meta-analysis (NMA), data is usually required in contrast form. In practice, multiple arms are 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. When the missing effect sizes can still be directly derived, additional assumptions about co-variances are needed to calculate standard errors. METHODS: Simple imputation techniques are used for substituting the standard errors of the missing comparisons and this has 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 imputation techniques by (1) assuming independence between contrasts, (2) estimating missing co-variances from the available contrasts in the multi arm trials, and from these estimating other co-variances. RESULTS: Two included randomized clinical trials, 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 practice, however, even if the 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 used in tools for performing multi arm trials with increases in missing effect sizes, in which frequentist NMA, although limited in NMA, might be carefully considered. Rücker G: Network meta-analysis, electrical networks and graph theory. Research Synthesis Methods, 2012, 3, 312–324.

INDIRECT COMPARISONS IN BENEFIT ASSESSMENT

PRM219

Kühnast S1, Schiffner-Rohe J2, Rahnenführer J1, Leverkus F2

1Technical University of Dortmund, Dortmund, Germany, 2Pfizer 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-RA). For this reason AMNOG allows indirect comparisons (ICs) to assess the extent of additional benefit. This study evaluates the applicability of available IC methods for 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 doctrines containing ICs are analyzed regarding the applied methodology. We use simulation studies to evaluate and compare the proportion of additional benefit (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 the results are lacking comparability of the benefits for 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 90% (n=24), 70% (n=11), otherwise simulation settings modified the externalities 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 impede the set of requirements to be able to derive valid statistical results. Prerequisites for adjusted ICs are often not merely as necessary studies and publications are not available. With respect to the progress of benefit assessment and the frequent price negotiation it would be helpful having alternatives with acceptable properties in order to estimate the extent of additional benefit.

THE USE OF INTERQUARTILE DEVIATION IN ESTABLISHING DELPHI PANEL CONSENSUS: A PRIORITIZATION OF INTRAVENTRUS IMMUNOGLOBULIN UTILIZATION

PRM220

Orange J1, Renner B2, Rane P3, Eaddy M

1Teax Children’s Hospital, Houston, TX, USA, 2Xonita, Palm Harbor, FL, USA, 3University 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 (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 ≥0.5 was used to determine consensus for disease states within each domain. Disease states reaching consensus in both domains were ranked according to a 940 algorithmic score to establish priority. RESULTS: The panel reached consensus on the severity of all disease 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 agamaglobulinemia, common variable immunodeficiency, primary immunodeficiency with absent B-cells) were considered to be life-threatening (DS1TA2) and required therapy (DS1TA3). The priority distribution of disorders based on the algorithm is as follows: DS1TA1=0, DS1TA2=1, DS1TA3=1, DS1TA4=1 DS2TA3=4, DS2TA4=4, DS2TA5=4, DS1TA1=1, DS1TA2=7, DS1TA3=11, DS1TA4=5, DS1TA5=2, DS2TA3=11, DS2TA4=0, DS2TA5=0. CONCLUSIONS: The application of interquartile deviation in establishing consensus across two 4-point Likert scales resulted in prioritizing 98% of disease states where IVIG can be used. Additional consensus-building rounds will be needed to prioritize the remaining disease states.

NETWORK META-ANALYSIS FOR HEALTH TECHNOLOGY SUBMISSIONS WORLDWIDE: A REPORT CHECKLIST FOR NETWORK META-ANALYSIS BEST PRACTICES GLOBALLY

PRM221

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 report a method for HTA bodies to produce a checklist for reporting NMA 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 was based on 16 of 24 guidelines covering methodological, statistical methodology, analyses performed, presentation of results, and technical issues. CONCLUSIONS: This reporting checklist provides practical support to health technology assessment (HTA) bodies in 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 NMA within HTA.

RESEARCH ON METHODS – Study Design

PRM222

TRANSPARENCY AND REPRODUCIBILITY OF SUPPLEMENTARY SEARCH METHODS IN NICE SINGLE TECHNOLOGY APPRAISAL MANUFACTURER SUBMISSIONS

PRM222

Witkowski MA, Aldhouse N

Abacus International, Ricerter, 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: updated searches performed (n=57), an updated search performed (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 meetings (n=6), and internal experts (n=2). Evidence review groups critically assessed 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 described between ERGs. CONCLUSIONS: Principles of transparency and reproducibility were not fol-