The empirical distribution of tau from IQWiG reports for the application in Bayesian meta analysis with random effects

Presenter: Ralf Bender

Ralf Bender, Matthias Maiworm, Christoph Schürmann, Sibylle Sturtz Institute for Quality and Efficiency in Health Care (IQWiG), Cologne

Introduction: Meta-analysis is the method of choice in systematic reviews to summarize the effect estimates of the included studies. Frequently, models with random effects are applied, which require the estimation of the heterogeneity parameter t. However, in the case of very few studies, t cannot be reliably estimated, leading to broad confidence intervals. In such situations, the application of Bayesian methods with weakly informative prior distributions is an option (Bender et al., Res. Syn. Methods 2018). Different choices for prior distributions for t are possible according to several proposals given in the literature (Röver et al., Res. Syn. Methods 2021).

Aims: The goal of the talk is to explore the empirical distribution of τ from IQWiG reports in order to inform future Bayesian random-effects meta-analysis in the case of very few studies.

Methods: We collected all published meta-analyses from IQWiG reports for the period 2005 to June 2020 and re-estimated τ by applying random-effects meta-analyses and the Paule-Mandel method. We applied the effect measures SMD for continuous data, HR for time-to-event data, whereas both OR and RR were considered for binary data. We summarized the empirical distributions of τ for the different effect measures and compared these distributions with the proposals for prior distributions in the literature.

Results: The empirical distributions of τ can be derived from IQWiG reports for the different effect measures.

Multi-step estimators of between-study variances and covariances, and their relationship with the Paule-Mandel estimator

Presenter: Dan Jackson

Introduction: Moment based estimation methods for the between-study variance are well established in univariate meta-analysis but can perform poorly. Multi-step moment based estimators have been proposed as a way to improve performance, whilst retaining computational and conceptual simplicity.

Aims: This talk will, in an “equation-light” way, explain how multi-step estimators are calculated. It will focus on univariate analyses but will also outline work in progress that explores the multivariate setting. The main aims are to disseminate some recently published findings that give greater credibility to multi-step estimators, that establish their relationship with the Paule-Mandel estimator, and to give an indication of the directions of our current unpublished work.

Methods: We illustrate the use of multi-step estimators using several different examples, both univariate and multivariate. We will outline the main findings from preliminary simulation studies in the multivariate setting (and will explain why extensive univariate simulation studies are not a priority).

Results: Our results illustrate the (recently published, and proved mathematically) relationship between the multi-step estimator and the Paule-Mandel estimator in the univariate setting. We will show that this relationship becomes weaker for multivariate meta-analyses but will also tentatively propose how multi-step estimators could be used to define a multivariate Paule-Mandel estimator.

A Robust and Computational-efficient Method for Multiple-outcome Network Meta-analysis

Presenter: Yong Chen

Introduction: In many biomedical settings, there is an increasing number of interventions available for a disease condition. It is critical for clinical decision-making to accurately evaluate and compare the relative efficacy and safety, as well as other patient centered outcomes of these interventions.

Aims: We propose a network meta-analysis model for multiple clinical outcomes.

Results: Inspired by the idea of composite likelihood, the proposed method only requires specification of the marginal distribution of each outcome, and a pseudolikelihood is then constructed under a working independence assumption. We also develop a novel inferential procedure with associated efficient computational algorithm, which is statistically robust (i.e., requires minimal distributional assumptions) and computational stable and fast. We will illustrate our method through multiple case studies including a network meta-analysis of comparing 12 labor induction methods.

Conclusion: The proposed composite likelihood based multivariate network meta-analysis method leads to a computationally efficient algorithm with robust statistical inference, while being able to take multiple outcomes into consideration.

4

On presenting heterogeneity in random-effects dose-response meta-analysis

Presenter: Nicola Orsini, PhD, Associate Professor of Medical Statistics

Department of Global Public Health, Karolinska Institutet

Introduction: Statistical heterogeneity across studies is commonly taken into account in estimation of dose-response meta-analysis by inclusion of random-effects but seldomly presented in either a numerical or graphical form in research articles.

Aim: To illustrate how to derive and graph study-specific dose-response relationships based on a weighted mixed-effects dose-response models estimated on tables of aggregated data.

Methods: Realistic linear and non-linear dose-response mechanisms arising from either experimental or observational studies are simulated in the presence of statistical heterogeneity. Given estimates of fixed and random-effects, marginal and conditional dose-response relationships are computed and visualized.

Results: Consider a dose-response meta-analysis of 30 prospective cohort studies where, for the average study, the mortality age-adjusted mortality rate ratio decreases with higher walking levels with a plateau at 2 hours/week. Using a piece-wise linear random-effects model, the estimated age-adjusted mortality hazard ratio conferred by a generic level x (ranging from 0 to 4 hours per week) of walking is $\hat{HR} = e^{-0.47\,x + 0.54\,I(x>2)(x-2)}$. Given the estimated variance 0.077 of the random effects associated with the slope before 2 hours/week, the meta-analyst can expect the middle 95% of the studies to have an age-adjusted mortality hazard ratio comparing 0 vs 2 hours per week between 0.9 and 8; $e^{-0.47\,(0-2)\,\pm\,1.96\sqrt{0.077(0-2)^2}}$.

Conclusions: Exploring the extent of heterogeneity requires the ability to derive desired contrasts based on estimates of both fixed and random-effects. A challenge in dose-response meta-analysis is the limited amount of empirical data. Code and worked examples are available at http://stats4life.se/software/drmeta/

Restoring credibility: Weighted and iterative least squares

Presenter: Tom Stanley

Introduction: The credibility of conventional meta-analysis (MA) has been a casualty of the 'replication crisis.' MA is routinely biased with high rates of false positives. Although several methods are said to correct publication selection biases, all have limitations, and none outperform the others consistently in applications.

Aim: We introduce a new tool, the weighted and iterated least squares (WILS), that greatly reduces selective publication bias and inflated rates of false positives when present at surprisingly little statistical cost.

Methods: WILS is a simple weighted average that makes no assumption about the nature of the selection process or how publication selection should be modelled. Essentially, WILS is the fixed effect weighted average that fully accommodates heterogeneity. Its algorithm identifies excess statistical significance when present and then removes those studies most responsible for the exaggeration of effect size.

Results: By comparing MA estimators: WILS, random effects, selection models and trim and fill to preregistered multi-lab replication results, we show that WILS greatly reduces (by 80%) the substantial biases (> 0.2 d) and high rates of false positives of these alternative methods in actual applications. Comprehensive, evidence-based simulations establish that WILS has negligible bias, lower MSE, and acceptable type I errors.

Conclusion: WILS is widely applicable to meta-analyses across the disciplines, typically reducing publication selection bias to scientific insignificance. Because it is so effective and easy to use, WILS offers a widespread practical solution that can restore credibility to meta-analysis. A complete paper is available.

Consequences of Simulation Design for Apparent Performance of Methods of Meta-analysis

Presenter: Elena Kulinskaya

Elena Kulinskaya, David Hoaglin and Ilyas Bakbergenuly

Introduction: Contemporary statistical publications rely on simulation to evaluate performance of new methods and compare them with established methods. In the context of random-effects meta-analysis of log-odds-ratios, we investigate how choices in generating data affect such conclusions.

Aims: To evaluate how the choice of the value of the overall log-odds-ratio, the distribution of probabilities in the control arm, and the distribution of study-level sample sizes affects comparative performance of standard meta-analytic methods.

Methods: To examine the impact of the components of simulations, we assess the performance of four methods: the best available inverse-variance-weighted two-stage method, a two-stage method with constant sample-size-based weights, and two generalized linear mixed models. We retain the customary normal distribution of study-level effects.

Results: The results show no important differences between fixed and random sample sizes. In contrast, we found considerable differences among data-generation models in estimation of heterogeneity variance and overall log-odds-ratio.

Conclusion: This sensitivity to design poses challenges for use of simulation in choosing methods of meta-analysis.

7

Can systematic reviews reliably assess harms? Not without a paradigm shift.

Presenter: Tianjing Li

Riaz Qureshi, Thanitsara Rittiphairoj, Evan Mayo-Wilson, Tianjing Li

Introduction: Research has shown that the methods for assessing harms (“adverse events”) in systematic reviews (SRs) is suboptimal.

Aims: To examine methods used to assess harms among a cohort of SRs of gabapentin and whether these reviews have found consistent signals of harms.

Methods: We searched for SRs of gabapentin that assessed at least one harm. Among the reviews that were classified as “reliable”, we extracted and compared the methods on pre-specification of harms, searching, analysis, and reporting of harms. We compared the results for harms between pairs of reviews with a high degree of overlap in their included studies as determined by corrected coverage area (CCA).

Results: From 4320 records, we identified 70 reliable SRs. At least one harm was pre-specified in 43/70 (61%) reviews. Reviews rarely searched for observational or unpublished data. Most reviews (51/70, 73%) assessed harms only narratively or with a mix of narrative and quantitative analyses. Among the 44 reviews with a meta-analysis, the Mantel-Haenszel model was most commonly used (19/44, 43%). Data on harms were often incomplete due to reviewers’ use of selection criteria in reporting harms. Of the 514 unique included reports for harms, half (244/514, 48%) were cited in a single review and most (458/514, 89%) were not cited as evidence for more than one condition. Among 18 pairs of reviews with CCA ≥ 50%, the selection of harms for assessment and the language used to describe harms varied between reviews.

Conclusions: To fundamentally revise how we approach synthesis of harms, immediate actions can include: describing any limitations to the assessment of harms; using standardized language when referring to harms; and presenting rationale for decisions made in selecting harms to assess and report.

8

A non-parametric approach for combining evidence on progression free and overall survival

Presenter: Nicky Welton

Nicky Welton (University of Bristol), Caitlin Daly (University of Bristol), Ross Maconachie (National Institute for Health and Care Excellence), Tony Ades (University of Bristol)

Introduction: Cost-effectiveness analyses of cancer treatments typically require an evidence synthesis of progression free survival (PFS) and overall survival (OS) outcomes. Existing methods either rely on the proportional hazards assumption or make parametric assumptions which may not capture the diverse survival curve shapes across studies.

Aims: Our aim was to develop a non-parametric approach for jointly synthesising evidence from published Kaplan-Meier survival curves of PFS and OS without assuming proportional hazards, to obtain the inputs required for cost-effectiveness models.

Methods: Relative treatment effects are pooled as differences or ratios of restricted mean survival time (RMST), estimated by the area under the survival curves (AUCs). The within trial correlation between the AUCs for PFS and OS is estimated using non-parametric bootstrap sampling. Network meta-analysis models are given for AUCs for PFS and post-progression survival (PPS) to ensure that OS=PFS + PPS. The relative treatment effects are applied to a baseline RMST for PFS and PPS to obtain estimates of RMST and discounted RMST for each treatment. The methods are applied to treatments for Stage IIIA-N2 Non-Small Cell Lung Cancer.

Results: Estimates conformed to the constraint that OS is greater than PFS. Treatments effects on RMST differed for PFS but were comparable for PPS. Conclusions: The model was simple to implement and fitted the data well. The results may be combined with external registry evidence beyond the restricted follow-up time used in the evidence synthesis, to produce the mean time in PFS and PPS states for the time-horizon required in economic models.

9

Can Bayesian Meta-Analysis Be Made Easy Enough for Everyone?

Presenter: David Rindskopf

Introduction: Bayesians believe that their methods are more complete, acknowledging all sources of uncertainty, and that the interpretation of results is more natural. But many people have been reluctant to switch to Bayesian methods. Why? Three possible reasons: (1) philosophy of statistics: parameters are fixed, not random, and even if they were random, prior distributions are too subjective; (2) Computational (programming) burden to specify the model; (3) interpreting the results.

Aims: All problems can be addressed; I will deal a little with the first, but mostly with the second, asking whether default values can make Bayesian meta-analysis simple enough. The answer in most cases is yes; the demonstration is made using a little-used program, hblm, by William DuMouchel.

Methods: In this section, I demonstrate how to do and interpret a simple Bayesian meta-analysis without knowing any theory of Bayesian statistics, and without equations. For non-Bayesians sake, we do not use technical terms; for example, we say distribution, not posterior distribution, since we do not even have to talk about priors. Everyone is intuitively a Bayesian.

Results: The demonstration shows (hopefully) that it is possible for a person untrained in Bayesian analysis to run and interpret the results of a simple Bayesian meta-analysis. This should make the promulgation of Bayesian methods easier.

Conclusion: Bayesian statistics has a great attraction for a segment of statisticians. Its mathematical and computational burdens have made it difficult to overcome the attractions of the ability to combine prior information with information from data, and the simple interpretation of results (compared with the counterfactuals required for frequentists).

10

Loop-splitting in network meta-analysis: a new approach to evaluating loop inconsistency

Presenter: Becky Turner

Becky Turner[1], Tim Band[2], Tim Morris[1], David Fisher[1], James Carpenter[1], Julian Higgins[3], Ian White[1]

[1]MRC Clinical Trials Unit, UCL, UK

[2]Department of Earth Sciences, UCL, UK

[3]Population Health Sciences, University of Bristol, UK

Introduction: Network meta-analysis relies on an assumption of consistency, meaning that direct andindirect evidence should agree for each treatment comparison. Existing tests do not handle treatments symmetrically and existing tests for inconsistency across the whole network based on a design-by-treatment interaction approach lack power.

Aims: We propose new tests for inconsistency and demonstrate their application to two example networks.

Methods: We apply a local test to a loop of treatments in the network. We define a model with one inconsistency parameter that can be interpreted as loop inconsistency. To provide a global test for inconsistency, we extend the model across multiple independent loops and describe how to identify independent loops within a network.

Results: The models are applied first to a small network meta-analysis comparing four treatments for promoting smoking cessation. Local tests for inconsistency are applied to each loop. Global tests are applied to every combination of independent loops. We demonstrate invariance to choice of loops and find no global evidence of inconsistency (p=0.67). Next, the models are applied to a large network meta-analysis comparing 12 antidepressant drugs, including 31 independent loops, and find no global evidence of inconsistency (p=0.51).

Conclusions: Our proposed models handle treatments symmetrically and are invariant to choice of reference treatment, which makes interpretation easier. The global model is invariant to choice of independent loops. In comparison with the existing approach to testing for global inconsistency in network meta-analysis, our model uses fewer degrees of freedom and is expected to improve power.

11

On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis

Presenter: Christian Röver

Introduction: Random-effects meta-analyses involving only few studies (say, less than 10) are commonly conducted. A Bayesian approach can appropriately account for the uncertainty in estimating the between-trial variance, which is particularly relevant in the case of only few available studies. However, it is important to specify proper (weakly) informative prior distributions for the heterogeneity parameter (tau). While prior specification for the effect parameter (overall mean) is typically chosen as non-informative and is therefore straightforward, consensus on the best choice of prior for the between-trial standard deviation is lacking.

Aims: To provide guidance on constructing and motivating weakly informative heterogeneity priors in specific contexts and for wider application areas

Methods: We reviewed the statistical literature for recommendations on the choice of heterogeneity priors and systematically evaluated motivating arguments for choice of different effect scales.

Results: Ideas and arguments are illustrated using a number of examples and effect scales, including mean differences, standardized mean differences, log-transformed effect sizes, regression slopes, and correlation coefficients. We provide guiding questions for judging whether a particular prior distribution is reasonable. For example, a half-normal prior distribution with scale 0.5 should often be appropriate for log-transformed effect sizes, such as log-ORs.

Conclusion: Transparent motivation and pre-specification of prior distributions for heterogeneity parameters are commonly feasible.

References:

C. Röver, R. Bender, S. Dias, C. H. Schmid, H. Schmidli, S. Sturtz, S. Weber, T. Friede. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. Research Synthesis Methods 2021. https://doi.org/10.1002/jrsm.1475

Drinking from a Firehose: Evidence Synthesis in the Era of COVID-19

Presenter: Haley Holmer

Introduction: The COVID-19 pandemic has stress-tested the systems of evidence-based practice, in particular, rapid review methodology. Some methodologists argue that a rapid review should include all components of a systematic review, done faster. The COVID-19 evidence environment has presented opportunities to advance rapid review methods.

Aims: To evaluate the suitability of rapid review methods in a public health emergency, and to pinpoint weak links in methods for evidence synthesis.

Methods: As a consultant for the World Health Organization (WHO), I used existing rapid review methods to answer questions for WHO decision-makers. Using case studies of COVID-19 rapid reviews, I assessed the risk of errors that could affect the accuracy or value of the review, and identified unnecessary steps in developing an informative and trustworthy review.

Results: Since February 2020, WHO developed >90 rapid review products; approximately half were conducted within 24-72 hours. In this environment, even on tight timelines, risk of missing important studies was low, and risk of interpretation errors was moderate. Omitting synthesis was acceptable for some questions. Not all steps were necessary to produce actionable output; specifically, topic refinement and systematic searches could be achieved in <24 hours. In addition, quality assessment was not always necessary, specifically for non-causal questions.

Conclusions: Systematically identifying and selecting data to inform decision-making without synthesis can be successful. In the COVID-19 setting, to use time efficiently and minimize potential harm, reviewers should adhere to practices for systematic search and study selection, and maintain flexibility on other steps while assessing potential risk.

13

Meta-analysis of dichotomous and polytomous diagnostic tests without a gold standard

Presenter: Enzo Cerullo (GMT)

Enzo Cerullo [1,2,] Hayley E. Jones [3], Terry J. Quinn [4,] Nicola J. Cooper [1,2,] and Alex J. Sutton [1,2]

[1]Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK

[2]Complex Reviews Support Unit, University of Leicester & University of Glasgow, Glasgow, UK

[3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[4]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

Introduction: Standard methods for the meta-analysis of diagnostic tests without a gold standard are limited to dichotomous data. Multivariate probit models are used to analyze correlated binary data, and can be extended to multivariate ordered probit models to model polytomous (i.e. non-binary) data. Within the context of an imperfect gold standard, they have previously been used for the analysis of dichotomous and polytomous diagnostic tests in a single study, and for the meta-analysis of dichotomous tests.

Aims: To extend previously proposed multivariate probit models for the meta-analysis of dichotomous and polytomous tests without assuming a perfect gold standard.

Methods: Dichotomous tests use multivariate probit regression likelihoods and polytomous tests use ordinal multivariate probit regression likelihoods. The model can accommodate a hierarchical partial pooling model on the conditional within-study correlations, which allow us to obtain summary estimates of joint test accuracy. We fitted the models using Stan, which uses a state-of-the-art Hamiltonian Monte Carlo algorithm. We applied the models to a dataset in which studies evaluated the accuracy of tests, and combinations of tests, for deep vein thrombosis.

Results: Whilst modelling the conditional dependence between tests, we found that assuming that the 'reference test' (ultrasound) was perfect would result in substantial underestimation in the specificity of the Wells score (6%), D-Dimer (9%) and up to 10% underestimation in joint test accuracy estimates.

Conclusions: We developed a hierarchical, latent class multivariate probit model for the meta-analysis of polytomous and dichotomous diagnostic tests without a gold standard.

*Corresponding Author

14

Assessing the Risk Of Bias due to Missing Evidence in Network meta-analysis: the ROB-MEN tool

Presenter: Virginia Chiocchia

Virginia Chiocchia[1], Adriani Nikolakopoulou[2,] Theodoros Papakonstantinou[1], Andrea Cipriani[3], Toshi A Furukawa[4], Julian PT Higgins[5], Matthew J Page6, Matthias Egger[1,5], Georgia Salanti[1]

[1] Institute of Social and Preventive Medicine, University of Bern, Switzerland

[2] Institute of Medical Biometry and Statistics, University of Freiburg, Germany

[3] Department of Psychiatry, University of Oxford, UK

[4] Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Japan

[5] Population Health Sciences, Bristol Medical School, University of Bristol, United Kingdom

[6] School of Public Health and Preventive Medicine, Monash University, Australia

Introduction & Aim: The risk of bias due to missing evidence threatens the validity of systematic reviews and meta-analysis and ultimately can affect clinical decision-making. A rigorous methodology to evaluate the impact of this bias on the meta-analysis results of a network of interventions is still lacking. We developed a framework and tool to assess the Risk Of Bias due to Missing Evidence in Network meta-analysis (ROB-MEN) by expanding the tools previously developed for pairwise meta-analysis (ROB-ME, http://www.riskofbias.info).

Methods & Results: ROB-MEN first evaluates the risk of bias due to missing evidence for each pairwise comparison separately. We consider possible bias due to the presence of studies with unavailable results (known unknowns) and the potential for unpublished studies (unknown unknowns) before reaching an overall judgement about the risk of bias due to missing evidence in each pairwise comparison. The bias and contributions from direct comparisons to the network meta-analysis (NMA) estimates are combined with the likelihood of small-study effects, as evaluated by network meta-regression, and that of bias from unobserved comparisons. Then, we evaluate the risk of bias due to missing evidence in each NMA estimate, which is our tool's final output. Using an R Shiny app, we illustrate the application of ROB-MEN to a NMA of 18 antidepressants.

Conclusions: The ROB-MEN tool is the first tool for evaluating the risk of bias due to missing evidence in NMA and it is applicable to networks of all sizes and shapes, including very complex ones as demonstrated by our example.

15

Novel and Existing Time-to-Event Extrapolation Techniques for Network Meta-Analyses: Recommendations for Health Technology Assessments

Presenter: Bart Heeg

Bart Heeg, PhD[1], Andrea Garcia, MSc[1], Sophie van Beekhuizen, MSc [1], Andre Verhoek, MSc [1], Ilse van Oostrum, MSc [1], Satrajit Roychoudhury, PhD[2], Joseph C. Cappelleri, PhD[2], Maarten J. Postma, PhD [3], Mario J.N.M. Ouwens, PhD [4]

[1]Ingress-Health, Rotterdam, The Netherlands

[2]Pfizer Inc, Groton, Connecticut, USA

[3]Unit of Pharmaco-Epidemiology & Pharmaco-Economics, Department of Pharmacy, University of Groningen, Groningen, The Netherlands

[4]AstraZeneca R&D, Mölndal, Sweden

Introduction: The NICE Decision Support Unit has published guidance on standard survival extrapolation methods, with the most recent Technical Support Document 21 focusing on flexible survival methods. However, further research is needed on relative survival considering general population mortality in flexible survival network meta-analyses (NMA).

Aim: To introduce and compare flexible relative survival methods in the NMA context.

Methods: Using a Bayesian framework, we illustrate the flexible NMA methods with a network of two trials in front-line advanced melanoma, where one compares nivolumab vs. dacarbazine and a second trial compares ipilimumab plus dacarbazine vs. dacarbazine. The following flexible methods are used for the comparison: parametric, mixture- (MCM) and non-mixture-cure (nMCM), piecewise, splines, and fractional polynomial (FP) models. The flexible survival methods are compared with NMAs based on HR.

Results: As cure was clinically plausible in advanced melanoma, we selected the nMCM method for the base case. Of the tested nMCMs, the log-logistic had the best fit in combination with clinical plausible extrapolations, predicting cure rates of 0.16 [0.10;0.23], 0.26 [0.08;0.50], 0.35 [0.20;0.46] and mean survival of 4.71 [3.44;6.13], 6.76 [3.62;12.13], 9.76 [7.38;11.74] for dacarbazine, ipilimumab+dacarbazine, nivolumab respectively. The mean incremental survival when comparing ipilimumab+dacarbazine and nivolumab was 2.86 [-2.79; 6.81]. Both the HR and the parametric NMA approaches predicted much shorter mean life years, while the FP results were in line with the log-logistic nMCM results

Conclusions: We introduced relative survival NMA methods using MCM, nMCM, piecewise, parametric mixture models. We further introduce an improved FP NMA which reduces the number of tested FPs.

NMAstudio: a fully interactive web-application for producing and visualizing network meta-analyses

Presenter: Silvia Metelli

Silvia Metelli and Anna Chaimani

Introduction: Several software tools have been developed in the last few years for network meta-analysis (NMA). However, presentation and interpretation of findings from large networks of interventions remain cumbersome and challenging.

Aim: To develop a novel online tool, called 'NMAstudio', for facilitating the production and visualization of key NMA outputs in a fully interactive environment.

Methods: NMAstudio is a Python web-application that provides a direct visual connection between a customizable network diagram and all NMA outputs. The user interacts with the network diagram by clicking one or more nodes-treatments or edges-comparisons. Based on their selection, different outputs and information are displayed: (a) boxplots of effect modifiers assisting the evaluation of transitivity; (b) pairwise or NMA forest plots and bi-dimensional plots when two outcomes are given; (c) league tables that combine numerical results with risk of bias or confidence ratings from the CINeMA framework; (d) ranking plots; (e) incoherence tests and comparison-adjusted funnel plots for evaluating small-study effects; (f) evolution of the network over time. Pop-up windows providing extra information are enabled. Analyses are performed in R using 'netmeta' and results are transformed to interactive and downloadable visualizations through coding in Python.

Results: We illustrate the tool using a network of 20 drugs for chronic plaque psoriasis. We demonstrate how NMAstudio simplifies the visualization of large and complicated networks as well as facilitates checking of assumptions and interpretation of findings.

Conclusions: Our web-application provides a truly interactive, flexible, and user-friendly tool to display, enhance and communicate the NMA findings.

17

Effectiveness of Deprescribing Interventions: Methodological Challenges when Conducting an Overview of Systematic Literature Reviews

Presenter: Shiyun Chua

Shiyun Chua MD[1], Shelly Gray PharmD MS[1], Adam Todd PhD FRPharmS SFHEA[2], Stephen Hughes BPharm MPhil PhD[2], Andrew Husband PhD FRPharmS PFHEA[2], Zizi Elsisi BPharm MS[1], Yuhan Huang MHS[1], Susan Smith MD MSc BCh[3], Emily Reeve BPharm PhD[4], Tae-jin Park PharmD[1], Jeffrey Harris MD MPH[1], Beth Devine, PhD, PharmD, MBA[1] and the Expert Advisory Panel from the US Deprescribing Network

[1] University of Washington, Seattle WA
[2] Newcastle University, Newcastle-upon-Tyne, UK
[3] Royal College of Surgeons, Dublin, Ireland
[4] University of South Australia, Sydney, Australia

Introduction: Overviews of systematic literature reviews (SLRs) are an increasingly popular method of evidence synthesis. Methodological challenges exist. Until recently, sparse guidance was available. We conducted an overview of SLRs of the effectiveness of deprescribing in older adults.

Aims: We describe four unique challenges encountered and solutions employed.

Methods: We convened a multinational, interdisciplinary team with prior experience conducting SLRs and overviews, and registered our overview with PROSPERO7. As challenges arose, we crafted solutions.

Results: Our search of 11 databases (2005-2020) returned 2,335 unique citations.

(1) In our inclusion criteria, we observed the lack of a standard definition of deprescribing as the field evolved over time, and adapted a definition used *a priori*. We found a lack of structure for examining both interventions and outcomes, and adopted an existing taxonomy for each. These tasks coalesced in the development of a conceptual model.

(2) The broad scope of many SLRs meant only a subset of primary studies met our inclusion criteria. Thus, we synthesized data iteratively at three levels.

(3) Heterogeneous outcomes in SLRs led us to use data visualization techniques to condense large amounts of data into an accessible and reader-friendly format. Inconsistent summary measures across the 13 meta-analyses led us to present meta-analyses results in a table for comparison.

(4) AMSTAR-211 had low discriminatory ability, with 30 of 34 (88%) studies rating critically low. Results were contextualized by describing patterns across domains.

Conclusions: In conducting our overview, we developed novel ways of addressing methodological challenges and gaps in current guidance.

18

Implications of Analyzing Time-to-Event Outcomes as Binary in Meta-analysis

Theodosia Salika[1], Rebecca Turner[1], David Fisher[1], Jayne Tierney[1], Ian R. White[1]

[1]MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, United Kingdom

Presenter: Theodosia Salika

Background: Systematic reviews and meta-analysis of time-to-event outcomes can be analyzed on the hazard ratio (HR) scale but are very often dichotomized and analyzed as binary using effect measures such as odds ratios (OR).

Aims: We investigated the impact of using these different scales by re-analyzing meta-analyses from the Cochrane Database of Systematic Reviews (CDSR) and using individual participant data (IPD).

Methods: We extracted CDSR meta-analysis data recorded either as binary (A) or as binary together with observed minus expected ("O-E") and variance ("V") statistics (B); (A) were originally analysed as binary and (B) with HR. We explored change in (A) when re-analysed with HR and complementary log-log (clog-log) link. We compared (B) to re-analysis with HR and clog-log link, or with OR and logit link. Using IPD meta-analyses of time-to-event outcomes, we compared analysis using OR to using HR and clog-log link, the log-rank approach or a Cox model.

Results: Within the CDSR, approximately 19% of meta-analyses provided significant results under one scale and non-significant results under the other. Results from the log-rank approach and Cox model were identical; situations where clog-log link outperformed logit link and vice versa were apparent. Differences between scales arise mainly from (1) high event probability, (2) between-study heterogeneity, (3) greater within-study variation in OR versus HR analyses, (4) percentage censoring, and (5) follow-up time. Conclusions: Dichotomising time-to-event outcomes may be adequate for low event probabilities, low heterogeneity, and shorter follow-up time; these findings guide the use of appropriate methodology for conducting such meta-analyses.

19

PRIME-IPD: A systematic method for preparing Individual participant data for meta-analysis

Presenter: Omar Dewidar

Omar Dewidar, Alison Riddle, Elizabeth Ghogomu, Alomgir Hossain, Paul Arora, Zulfiqar A Bhutta, Robert E Black, Simon Cousens, Michelle F Gaffey, Christine Mathew, Jessica Trawin, Peter Tugwell, Vivian Welch, George A Wells

Introduction: Individual participant data meta-analysis (IPD-MA) is increasingly being conducted in systematic reviews as it enables the investigation of additional hypotheses and consistent analysis among studies. However, managing and preparing IPD may be resource-intensive and time-consuming due to differences in naming conventions, data structures and file formats.

Aim: To describe an approach to prepare IPD for analysis that we applied to an IPD-NMA for mass deworming for children and illustrate as an exemplar.

Methods: We developed this process *a priori* with our research team involving statisticians and clinical experts while reviewing available guidance from Cochrane handbook and working groups and the Get Real IPD Working Group.

Results: Our 5-step approach to prepare IPD for MA entails *Processing, Replication, Imputation, Merging, and Evaluation (PRIME)*. The *Processing* step includes steps to prepare the data using common variable names, identifying inconsistencies in the datasets and replicating the study's reported sample size. The *Replication* step involves recalculating the reported descriptive statistics and study findings as an additional check on data quality. *Imputation* entails dealing with missing data by imputation or other methods as appropriate. *Merging* involves combining the datasets, and in the *Evaluation* step, new variables can be calculated, if needed, for standardized comparison of effects.

Conclusions: PRIME-IPD helped us avoid costly and resource-intensive mistakes and ensured a more efficient process when conducting IPD-MA. Adaptations to the framework may be required depending on nature of data and the purpose of the IPD-MA. More guidance will be developed detailing how each step should be performed.

20

Bias correction and sensitivity analysis for p-hacking in meta-analyses

Presenter: Maya Mathur

Introduction: P-hacking (i.e., manipulating results within studies to attain significance) is distinct from publication bias (i.e., selecting certain studies for publication). Although the distinction is conceptually subtle, I will argue it is statistically important, such that existing methods for publication bias cannot adequately correct for p-hacking.

Aims: To describe and seek feedback on in-progress statistical methods to model and adjust for p-hacking in meta-analyses.

Methods: I will consider a mechanism of p-hacking in which, for a given study, investigators conduct a series of hypothesis tests until they attain the first significant positive result (which they then report) or until they have conducted some maximum number of tests, at which point they give up and do not report the result. I will consider cases involving a finite maximum number of tests as well as cases involving a potentially infinite number of tests (p-hacking as long as needed to obtain a significant positive result). In addition to these p-hacked studies, there may be unhacked studies in which investigators simply report the first result they obtain.

Results: In this framework and under a variety of assumptions, the test statistics from the observed nonsignificant and negative studies follow a truncated distribution that can be used to obtain bias-corrected results.

Conclusions: Developing statistical methods for p-hacking, rather than publication bias alone, remains a nascent area. The preliminary work I will discuss makes progress on this front.