tended to be protective of discontinuations due to adverse events relative to standard dose EFV. Use of arm-specific regression supplemented by endonodal trials led to tighter confidence intervals facilitating decision-making. Specifically, DTG was consequently superior to EFV with respect to CD4 cell counts and raltegravir was distinguishable from EFV when it was not otherwise. **CONCLUSIONS:** Making full use of available evidence is the focal strength of NMA methodology. Therefore, we recommend use of endonodal trials to further supplement evidence bases requiring arm-specific meta-regression.

PRM9

CRITICAL APPRAISAL OF REAL WORLD EVIDENCE – A REVIEW OF RECOMMENDED AND COMMONLY USED TOOLS

Quigley JM[1], Thompson JC[2], Halfpenny NJ[2], Scott DA[1]
[1]*ICON Health Economics and Epidemiology, Oxford, UK,* [2]*ICON Health Economics, Oxford, UK*

OBJECTIVES: In an absence of randomized controlled trials (RCTs) and to verify RCT evidence, health technology assessment (HTA) agencies commonly rely on real-world (RW) studies to provide efficacy evidence for healthcare interventions. RW study designs can introduce considerable bias into a systematic review (SR) and several methodologies exist to evaluate the risk of bias in such studies. We conducted a series of reviews to identify which tools are commonly used and which are recommended by HTA bodies. **METHODS:** A targeted search of SRs including RW studies, conducted in MEDLINE and EMBASE (OVID SP), identified reviews published January 2013-June 2015. Studies identified were reviewed to determine which appraisal tool was used. Secondly, recommendations for the critical appraisal of RW studies by expert review groups (Cochrane, CRD) and HTA bodies (NICE, SMC, NCPE, AWMSG, IQWiG, PBAC, AMCP, AHRQ and CADTH) were reviewed. **RESULTS:** 1885 studies were identified and screened. Commonly used tools included Downs & Black, Chalmers, the Newcastle-Ottawa Scale, and the CriSTal checklist. Neither Cochrane nor CRD recommend a particular risk of bias instrument. The AHRQ developed the MORE checklist following a SR of existing critical appraisal tools. Of the other HTA bodies only CADTH recommend use of a specific critical appraisal tool; SIGN 50 (for cohort or case-control studies). The tools identified examine a variety of criteria including reporting, external validity, bias, confounding, and power. **CONCLUSIONS:** There is no consensus on a preferred instrument that allows for the assessment of all types of RW evidence and critical appraisal of RW evidence is often omitted from HTA submissions. There is thus a need for cross communication between groups to reach a consensus and develop a suitable tool. Until a suitable tool is developed, reviewers should select the most appropriate checklist for the design of the studies identified in a particular SR.

PRM10

DOES ATTRITION IN SUBJECT-BASED STUDIES OF DRUG SAFETY LEAD TO BIAS RELATED TO MORBIDITY?

Pignot M[1], Klamert A[2], Pisa G[1], Potthoff P[3]
[1]*Kantar Health, Munich, Germany,* [2]*Kantar Health Germany, Munich, Germany,* [3]*Kantar Health, Munichn, Germany*

OBJECTIVES: Sample quality in prospective long-term drug safety studies can be impaired by selective lost-to-follow-up. Attrition will especially bias the sample, when patients with relevant risk factors selectively drop out. In this case, effects in endpoints cannot be related to study-relevant independent variables. The present contribution will demonstrate how careful follow-up procedures can prevent disease-related drop-out bias of the sample. **METHODS:** For a long-term prospective safety study of new Oral Contraceptive (OC) 25,213 women aged 20 to 40 years were enrolled from gynecological practices in Germany. The women filled-in a baseline questionnaire and were followed-up over two years with four follow-up questionnaires in total. Whenever safety-relevant signs were reported in the questionnaires, physicians validated the report. After two years 12,823 women were still in the sample and completed the fourth questionnaire. Disease differences between the "Retained" and the "Lost" group, which could indicate sample bias, were analysed using multivariate methods. **RESULTS:** The "retained" and the "lost-to-follow-up" group did not differ in initial disease status or in risk factor at study start: High blood pressure: 2.9% in "Retained Group"; 2.7% in "Lost Group" (phi=.006; n.s.), diabetes : 0.6% vs. 0.6% (phi=.001; n.s.), high cholesterol: 2.6% vs. 2.4% (phi=.007; n.s.),venous thrombosis : .8% vs. .8% (phi=.002; n.s.), smoker-rate : 34.9% vs. 42.3% (phi=.127), BMI>30: r=.004 (n.s.), age r=.048 (n.s.). The results show that drop-out of the initial sample is not related to study relevant morbidity and that sample bias cannot be concluded. **CONCLUSIONS:** Careful follow-up methods guarantee low lost-to-follow-up in longterm prospective studies of drug safety. Since drop-out cannot be attributed to study-relevant confounders, attrition does not lead to sample bias.

PRM11

ESTIMATION OF THE PROGRESSION OF COLON CANCER BY JAPANESE LARGE-SCALE INSURANCE BENEFITS DATA ANALYSIS

Iwasaki K[1], Arata H[1], Soeda J[2], Yanai T[2], Hiroi S[3]
[1]*Milliman, Tokyo, Japan,* [2]*Takeda Pharmaceutical Company Limited, Tokyo, Japan,* [3]*Takeda Pharmaceutical Company Ltd., Tokyo, Japan*

OBJECTIVES: Accurate determination of the progression degree in colon cancer is of paramount importance for the decision making in treatment policy. However, it had been difficult to extract the exacerbation status from the real-world data. The objective of the study was to develop the model to determinate the progression degree using the insurance benefits data in Japan. **METHODS:** We conducted analyses using claims data provided by Medical Data Vision Co., Ltd. We extracted target patients by the criteria those who meets all of the following conditions; at least one colon cancer diagnosis (ICD-10 code C18~20), tractable from the first diagnosis to death, and have at least 365 days of observation. We set the progression degree as a scale from 0% to 100%. The degree of 100% indicates the patient death. For the first diagnosis, the scale was adjusted based on the patient's condition. We have developed a linear regression model by using the medication frequency of ATC codes as independent variables and the logit of progression degree as a dependent variable. **RESULTS:** 1,436 target patients were extracted from the database. When the actual progression degree is over 80%, the estimated progression degree rises with the actual degree, however, at the lower progression degrees, the estimated degree was excessively overestimated. **CONCLUSIONS:** We have developed a model to estimate the progression degree in colon cancer. The model estimates the progression degree well only for the last phase patients. The model should further be improved to minimize the bias at the lower degree.

PRM12

BEYOND THE MIDDLE: EVALUATING SURROGACY OF CLINICAL TRIAL ENDPOINTS ACROSS TRIAL DURATIONS

Wasserman M[1], Amin A[2]
[1]*Double Helix Consulting, New York, NY, USA,* [2]*Double Helix Consulting, London, UK*

OBJECTIVES: Overall survival (OS) remains the gold standard measure of clinical efficacy for oncology clinical trials due to its objective nature and consistency between diseases and treatments. However, given recent advances in treatments, and prolonged survival, OS benefits are becoming more challenging to establish, requiring more extensive follow-up. A number of methods to test this rationale have been developed but these often lack adequate data, relying solely on mean or median survival. Recognising these limitations, we developed an alternative methodology whereby surrogacy is established over time, to ensure that a surrogate is not only valid at the mean, but also throughout treatment duration. **METHODS:** A number of different survival points were derived from selected oncology trials by digitizing available survival curves. PlotDigitizer 2.6.4 software was used to establish time points for 10%, 25%, 50%, 75%, and 90% OS and progression-free survival (PFS). Correlation and regression analysis were evaluated at these percentiles based on survival times. Patient populations between the clinical trials were comparable to one another. Statistical analysis was conducted in STATA 12. **RESULTS:** Correlation analysis found the strongest association between PFS and OS between 75% and 25% survival (0.865 to 0.953; $p<0.01$), with a weak association at 90% survival (0.61; $p=0.096$). Regression analysis also found that PFS had the largest influence on OS between 75% and 25% survival ($R2>0.75$). **CONCLUSIONS:** Given the varying nature of how patients progress across and within types of therapies, it is essential to ensure the surrogacy of the endpoint across the full trial duration. For example, patients may progress early on in a disease and surrogacy may not be consistent across different time points. Additionally, in evaluating older studies, using this approach of scanning survival data will provide a richer picture of the disease area that may no longer be available from authors or research institutions.

PRM13

NETWORK META-ANALYSIS OF MULTIPLE OUTCOMES INCORPORATING DOSE-RELATED CONSTRAINTS: APPLICATION TO OVERACTIVE BLADDER SYNDROME

Owen RK, Tincello DG, Bujkiewicz S, Abrams K
University of Leicester, Leicester, UK

BACKGROUND: Overactive bladder(OAB) is characterized by symptoms of urgency, incontinence, frequency and nocturia. With the syndromic nature of the condition, clinical trials often solely report the most effective outcome i.e. the symptom with the largest improvement. As a result different interventions are evaluated for different outcomes, which can have severe implications for network meta-analyses, and consequently, decision-making. **OBJECTIVES:** To evaluate the use of multivariate network meta-analysis(MVNMA) to identify the most effective intervention for treating OAB syndrome. **METHODS:** Using Bayesian Markov Chain Monte Carlo methods, we developed MVNMA accounting for the correlation between multiple outcomes to predict treatment effects for missing data. We extended this model to incorporate the exchangeability between treatment effects of the same intervention with different methods of administration (e.g. immediate release, extended release, intravesical etc.) and incorporated dose-response constraints on increasing doses. The outcomes of interest were mean change from baseline in incontinence, and urgency episodes. **RESULTS:** Independently, the datasets included 109 and 56 trials, evaluating 93 and 51 interventions, for incontinence and urgency episodes respectively. Sacral nerve stimulation appeared to be the most effective intervention for reducing incontinence with an estimated mean reduction of -8.9(95%CrI:-10.9,-7) episodes per 24hours relative to placebo. For urgency, sacral nerve stimulation was disconnected from the network and thus could not be not evaluated. Borrowing information between outcomes, the dataset for multivariate analyses included 117 trials evaluating all 95 treatments for OAB. Sacral nerve stimulation appeared to be the most effective intervention for both incontinence and urgency episodes with an estimated mean reduction of -8.3(95%CrI:-10.1,-6.9) and -9.1(95%CrI:-10.5,-7.3) episodes, respectively. **CONCLUSIONS:** Sacral nerve stimulation appeared to be the most effective intervention for treating OAB symptoms. MVNMA allowed us to evaluate all interventions across all outcomes, and in this case also increased precision in treatment effect estimates. Further work includes adjustment for baseline severity.

PRM14

THE VALUE OF PROGRESSION-FREE SURVIVAL (PFS) AS AN ENDPOINT IN ONCOLOGY TRIALS

Quinn C[1], Palmer S[2], Bruns J[3], Grant C[1], Sykes D[4], Kaura S[5]
[1]*PRMA Consulting, Fleet Hampshire, UK,* [2]*University of York, Heslington, York, UK,* [3]*German Cancer Society, Berlin, Germany,* [4]*PRMA Consulting, Hampshire, UK,* [5]*Celgene Corporation, Summit, NJ, USA*

OBJECTIVES: The clinical endpoints selected for oncology trials have to meet the needs of diverse stakeholders: patients, clinicians, regulators, and HTA agencies, each with a different perspective. PFS is becoming a more widely accepted measure of treatment efficacy, but there is tension between regulators and payors regarding its acceptability. This study investigated PFS as a valid and credible endpoint from the perspectives of relevant decision-makers. **METHODS:** Published and gray literature (2005–2015) were searched for regulatory and HTA guidance on PFS as an endpoint. We identified examples of decisions by regulators and HTA agencies in which PFS