both challenges and opportunities for industry. The changes in the DSM-5 criteria may translate into changes for how we go about developing medical products to treat psychiatric disorders, including ADHD. There will need to be an investment in research and education, and sponsors must examine the possibility of developing new endpoints and outcome assessments for use in clinical trials.

PRM185

THE IMPACT OF DSM-5 ON THE DEVELOPMENT OF DRUGS TO TREAT MAJOR DEPRESSIVE DISORDER

Meyers OI

Truven Health Analytics, Cleveland, OH, USA

OBJECTIVES: In May 2013, the American Psychiatric Association released the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), a classification system for psychiatric conditions. DSM-5 brings significant changes to many diagnostic categories as compared to the previous edition. The objective of this review was to examine the changes in the Major Depressive Disorder (MDD) criteria and discuss the impact these changes may have for industry. **METHODS:** A line-by-line review of the DSM-5 and DSM-IV criteria for MDD was undertaken. Significant changes were highlighted and discussed from the point of view of sponsors of clinical trials for psychopharmacologic agents being developed to treat MDD. **RESULTS:** The primary symptom criterion for MDD remains unchanged, requiring five of nine symptoms, over a two-week period. The changes of note have to do with the differential diagnoses and specifiers. One change that received significant attention in the time leading up to the publication of DSM-5 was the elimination of the bereavement exclusion, which discounted bereavement after the loss a loved one within the first two months as part of the normal grief process. In terms of specifiers, a new addition in MDD is "with anxious distress," referring to episodes of depression characterized by at least two of five symptoms of anxiety. DSM-5 notes that this is associated with "greater likelihood of treatment nonresponse." Therefore, this is a factor sponsors may wish to consider in developing their trial inclusion/exclusion criteria. **CONCLUSIONS:** The significant changes in DSM-5 pose both challenges and opportunities for industry. The changes in the DSM-5 criteria translate into changes for how we go about developing medical products to treat psychiatric disorders, including MDD. There will need to be an investment in research and education, and sponsors must examine the possibility of developing new endpoints and outcome assessments for use in clinical trials.

PRM186

PATIENT NETWORKS AS A DATA SOURCE FOR PATIENT REPORTED OUTCOMES RESEARCH. CARENITY EXPERIENCE

Castejón N[1], Chekroun M[2], Martínez García J[1], Gay CH[2], Rebollo P[1]

[1]*LASER Analytica, Oviedo, Spain,* [2]*carenity.com, Paris, France*

OBJECTIVES: To explore the potential of online patient networks (PN) as a viable source of PRO data for clinical research. Several PNs have emerged in the last few years in different European countries, and as a natural meeting point for chronic patients with an active engaged with their communities, they represent a promising source of patient reported data. In this original, the experience with the French PN "Carenity" is described. **METHODS:** Given the great heterogeneity of the users of "Carenity", and the fact that the test was computer-led by definition, a Computer Adaptive Test (CAT) was considered the best choice. The authors decided to use a culturally adapted version of CAT-Health system, which measures generic health-related quality of life (HRQoL). However, in absence of a calibration for the French population, a selection of the best items was used, using the Spanish calibration as a reference. All patients in the PN were invited to participate in the test. A score was estimated for the test using the Spanish parameters, as a rough approximation of the real score. Age, sex and the main pathology of the subjects were also collected. **RESULTS:** Preliminary results from the first week of data collection show 601 patients answered (Women: 404, Men: 140). The most frequent reported pathologies and their t-scores were multiple sclerosis (N:92,M:37.91,SD:5.85), fibromyalgia (N:81,M:36.65,SD:4.99), ankylosing spondylitis(N:60,M:37.74,SD:5.32) and both types of diabetes (I: N:53,M:50.38,SD:10.94, II: N:41,M:48.12,SD:10.04). Significant differences ($p<0.05$) were found in diabetes patients by sex, and between both types of diabetes and the other 3 most common pathologies. **CONCLUSIONS:** Carenity PN seems to be a fast way to obtain PRO scores directly from patients. Preliminary results show differences in the expected direction.

RESEARCH ON METHODS – Statistical Methods

PRM187

DEFINING THE PROPER METHODOLOGY TO USE IN A DATA-PEEK FOR POWER (DPP)

Eisenberg D, Wasser T

HealthCore, Inc., Wilmington, DE, USA

OBJECTIVES: Late phase research is conducted outside the RCT setting where there is uncertainty as to how many subjects are needed to find differences between groups. Due to the lack of real-world information (non-RCT) in late phase designs, there are no tangible inputs for power calculations. This research defines a bias-free method to examine data while determining sample size. **METHODS:** As an example of the application of a DPP, a study examining the decrease of HGbA1c values in two different insulin delivery methods was examined in patients with several comorbid conditions. Literature examined found little to no data and a DPP was used to determine effect size (ES) and standard deviations (SDs) once 30 patients had been enrolled in each group. The DPP procedure was: 1) Determine the test statistic; 2) Identify the power formula most appropriate to the test statistic; 3) Determine the ES, variation and assumptions needed for the data-peek in the form required by the formula; 4) Construct a matrix of possible sample size values; and 5) Select a sample size that is obtainable and answers the research question. **RESULTS:** Data for group A demonstrated a mean reduction of 2.75% ±0.760, group B mean reduction of 3.01% ±0.636. Exact power analysis showed 113 subjects per group would be needed. A matrix of likely sample size based on these values ranged from 44 to 193 per group. Based on this DPP, a sample of 120 per group was selected as the sample size that would deliver clinically meaningful results. **CONCLUSIONS:** A DPP is useful in late phase research to define appropriate sample size where no data exist. It is important to note that DPP methods do not require significance testing, but the benefit is no need for a correction for multiple comparisons at the time of the final analysis.

PRM188

FAULTY CONNECTIONS: CAN CRITICISMS OF NETWORK META-ANALYSIS IN NICE SUBMISSIONS BE AVOIDED?

Martin A, Rizzo M, Iheanacho I

Evidera, London, UK

OBJECTIVES: To assess 1) how network meta-analyses (NMAs) included within manufacturer submissions to the National Institute for Health and Care Excellence (NICE) have been criticised by its Evidence Review Groups (ERGs); 2) how some of these criticisms might be avoided in future submissions; and 3) the extent to which such avoidance might increase the likelihood of a new intervention being approved. **METHODS:** We reviewed the ERG reports of all NICE technology appraisals published since January 2007 to identify those where the manufacturer's submission included an NMA. Subsequently, all criticisms made by the ERG of such analyses were analysed to seek common themes; and assess how often any one type of criticism was associated with a rejection by NICE. **RESULTS:** A total of 181 NICE technology appraisal reports were evaluated. These covered 243 separate interventions, 83 (34%) of which were drugs for cancer. Overall 37–64% of submissions cited NMAs, of which 43–83% were criticised, with this proportion having increased over time. Avoidable criticisms related to flaws in the systematic review methodology used to identify relevant RCTs for the analysis; inappropriate pooling of data from heterogeneous studies; and use of suboptimal statistical approaches in conducting the NMA. Unavoidable criticisms related to the lack of RCTs available for competitor drugs in the population of interest. However, no association was found between flaws in the NMA and a decision by NICE not to approve the use of the intervention. Instead, such rejection was associated mainly with a lack of evidence of clinical efficacy or cost-effectiveness in the target population. **CONCLUSIONS:** Most criticisms of NMAs could be avoided by a more rigorous and transparent approach to conducting and reporting the underlying systematic review and statistical analysis. However, rejection of submissions remains a considerable risk where the underlying evidence is weak.

PRM189

METHODOLOGICAL CHALLENGES IN COMPARING TOPICAL THERAPIES IN DERMATOLOGY IN THE ABSENCE OF HEAD TO HEAD STUDIES

Greiner RA[1], Batscheider A[1], Eheberg D[1], Gohlke A[1], Pfiffner C[1], Ehlken B[1], Macheleidt O[2], Hutt HJ[2]

[1]*IMS Health, Munich, Germany,* [2]*LEO Pharma GmbH, Neu-Isenburg, Germany*

OBJECTIVES: German HTA agency requires evidence about the additional benefit of a new pharmaceutical versus an appropriate comparator as basis for price negotiations. This is challenging when head-to-head studies (H2HS) or randomized placebo-controlled trials (RPCTs) are missing and particularly in dermatology, where topical therapies in registration trials are usually compared to their individual vehicle. The aim of this research was to describe different approaches to assess the additional benefit of a new topical therapy under these limitations. **METHODS:** For ingenolmebutate-gel (IMG) and the appropriate comparator diclofenac-hyaluronic-acid (DHA) bibliographic literature search was conducted for RCTs followed by sequential screening on H2HS, comparable endpoints, RPCTs, common bridge comparator, H2HS of vehicles alone, RPCTs of vehicles. The similarity of vehicles was assessed by comparison of efficacy and safety profile. The lack of H2HS demands to conduct the following approaches depending on the comparability of vehicles: 1. An adjusted indirect comparison due to Bucher 1997 (vehicles are placebo-like or adequately similar) 2. Linkage of direct comparisons due to Wells 2009 (possible when H2HS or RPCTs of vehicles are available) 3. Mixed treatment comparison (MTC) (prerequisites as mentioned for Bucher). **RESULTS:** 5 RCTs for IMG versus 3 RCTs for DHA were identified with comparable endpoints. No RPCTs for topical therapies or for their vehicles, no H2HS of vehicles, no bridge comparator and no clear evidence for the adequate similarity of both vehicles could be detected. Therefore, the prerequisites of all available statistical methods are not met and cannot thoroughly be applied. Notwithstanding these limitations, Bucher (RR[95%KI]: 4.14[2.03;8.47]) and MTC both favor IMG significantly while Wells showed non-inferiority (RR[95%KI]: 0.8[2,03;8,47]) in the primary endpoint of IMG versus DHA. **CONCLUSIONS:** A definition of adequate similarity for vehicles by German HTA agencies is needed to enable the use of methodologically sound indirect comparisons or MTCs in reimbursement dossiers.

PRM190

USE OF MULTIVARIATE BAYESIAN EVIDENCE SYNTHESIS TO REDUCE UNCERTAINTY AROUND CLINICAL EFFECTIVENESS AND QUALITY OF LIFE ESTIMATES

Bujkiewicz S[1], Thompson J[1], Sutton A[1], Cooper N[1], Harrison M[2], Symmons DP[2], Abrams KR[1]

[1]*University of Leicester, Leicester, UK,* [2]*University of Manchester, Manchester, UK*

OBJECTIVES: In health technology assessment, decisions about reimbursement of new health technologies are largely based on effectiveness estimates. These estimates are sometimes also used to predict the health-related quality of life outcomes, such as EQ-5D, as part of economic evaluation. However, sometimes these effectiveness estimates are not readily available. When many alternative instruments measuring these outcomes are being used (and are not all reported) or an extended follow-up time of clinical trials is needed to evaluate long-term endpoints (and drug development is at an early stage), data on relevant outcomes may be limited. The aim of this study was to develop methodology that would allow synthesis of all available evidence to assess interventions early and reduce uncertainty around relevant outcomes. **METHODS:** Bayesian multivariate meta-analysis

methods have been developed for synthesis of diverse sources of evidence: multiple outcomes (including surrogate, potentially short-term endpoints) and other external evidence. These methods were applied to an example in rheumatoid arthritis where outcomes such as the Health Assessment Questionnaire (HAQ), the Disease Activity Score (DAS-28) and the American College of Rheumatology (ACR20) are synthesized. External information about correlations between the outcomes was included in the form of informative prior distributions. Estimates of HAQ were then mapped onto EQ-5D. Also in an alternative approach, the multivariate framework was applied to model jointly the utility estimates and the clinical effectiveness outcomes. **RESULTS:** The use of multivariate meta-analysis led to reduced uncertainty around the effectiveness and utility estimates. Combining the HAQ with DAS-28 gave a 19% reduction in the uncertainty around the estimate of HAQ and also 16% around the estimate of EQ-5D. **CONCLUSIONS:** By allowing all relevant data to be incorporated in economic evaluations of new health technologies, this multivariate approach to meta-analysis can lead to reduced uncertainty and hence more efficient decision-making in health care.

PRM191

NETWORK META-ANALYSIS OF MULTIPLE OUTCOMES: A SIMULATION STUDY AND APPLICATION

Achana F, Dequen P, Gray L, Cooper NJ, Abrams KR, Owen RK
University of Leicester, Leicester, UK

The usefulness of a multivariate approach to compare treatments in the context of pairwise meta-analysis has been widely demonstrated in the literature. However, this approach has not yet been considered for multiple treatment comparisons. We believe that extending such methodology to network meta-analysis (NMA) will increase the primary evidence base allowing us to compare more interventions across multiple outcomes measures. Borrowing strength between outcome measures using multivariate NMA can also potentially increase the precision of relative treatment effect estimates and reduce the impact of outcome reporting bias.**OBJECTIVES:** To extend standard NMA to incorporate multiple outcomes of interest and evaluate the use of multivariate NMA models through simulated and real datasets. **METHODS:** We developed a random effects multivariate NMA model to account for the correlation between multiple outcome measures. The potential benefits of this method were demonstrated in a simulated example comparing univariate and bivariate NMAs for continuous outcome measures. We further explored the application of our multivariate NMA model using a case study comparing anti-obesity pharmacological interventions for waist circumference, weight change and BMI change from baseline. **RESULTS:** The simulation study showed that through use of multivariate NMA the precision in mean relative treatment effects increased compared to a standard univariate NMA. This held true under multiple scenarios testing model parameters including both within- and between-outcome correlations. Similar findings were obtained from the application to the example dataset in obesity. **CONCLUSIONS:** Our method proves particularly useful in reducing uncertainty around relative effectiveness estimates when the outcomes included for analysis are highly correlated. However, the advantages of the multivariate NMA are limited where there is little correlation between outcome measures. Further work will explore the applicability of multivariate NMA methods to different types of outcomes such as binary outcome measures.

PRM192

HANDLING VARIABILITY IN TIME ENDPOINTS IN MULTI-CENTRE TIME AND MOTION (T&M) STUDIES: A CASE STUDY OF ERYTHROPOIESIS-STIMULATING AGENTS FOR ANAEMIA MANAGEMENT IN 13 CENTRES IN ITALY

Kritikou P[1], De Cock E[2], Proskorovsky I[3], Payne KA[3], Tomic R[4]
[1]*Evidera, London, UK,* [2]*Evidera, Barcelona, Spain,* [3]*Evidera, Dorval, QC, Canada,* [4]*Roche S.p.A., Monza, Italy*

OBJECTIVES: In multi-centre Time and Motion (T&M) studies, time endpoints can be highly variable due to differences in centre practices. Our aim was to assess the impact of the type of analysis employed on the results of a T&M study. **METHODS:** Data from 13 centres were analyzed in relation to each of the following: drug preparation, distribution, and injection, using three methods. Base case methodology included a random intercept generalized linear mixed effect model assuming gamma distribution with log link function to account for potential centre clustering effect and non-normality of the outcome measure. The two alternative methods were: standard linear regression (assuming time data are normally distributed) and gamma regression with log link function (assuming time data are positively skewed), both of which do not account for centre clustering effect. Sample means and variability as measured by 95% confidence interval (CI)) were also compared. **RESULTS:** For the base case, mean time was 0.53 min (95% CI: 0.33-0.85) for "preparation", 0.30 min (95% CI: 0.22-0.40) for "distribution", and 0.81 min (95% CI: 0.59-1.11) for "injection". Mean time resulting from the standard linear regression was markedly higher for "preparation": 0.66 (95% CI: 0.59-0.73), and similar for "distribution" and "injection": 0.34 (95% CI: 0.30-0.37) and 0.84 minutes (95% CI: 0.79-0.88), respectively. Using the gamma regression yielded similar results to standard linear regression: 0.65 (95% CI: 0.59-0.71), 0.31 (95% CI: 0.29-0.34), and 0.83 minutes (95% CI: 0.79-0.88), respectively. The base case scenario detected a "centre-clustering" effect, hence producing substantially wider CIs compared to both alternative methods which ignore dependence in the data. **CONCLUSIONS:** Although mean task times remained relatively stable across the various methods, 95% CIs were substantially wider for random intercept model. If "centre-clustering" is detected, random effects regression models must be employed to produce valid confidence intervals around point estimates.

PRM193

BAYESIAN NETWORK META-ANALYSIS TO ASSESS RELATIVE EFFICACY AND SAFETY OF CANAGLIFLOZIN IN PATIENTS WITH TYPE 2 DIABETES MELLITUS (T2DM) INADEQUATELY CONTROLLED WITH METFORMIN

Pacou M[1], Taieb V[1], Abrams KR[2], Diels J[3], van Sanden S[4], Garg M[5], Schroeder M[6], Kaur V[7], Nielsen ATN[5], Nuhoho S[6], Neslusan C[8], Hemels M[5]
[1]*Amaris, London, UK,* [2]*The Institute of Cancer Research, Sutton, UK,* [3]*Janssen Pharmaceutica, Beerse, Belgium,* [4]*EMEA HEMAR Analytics, Janssen EMEA, Beerse, Belgium,* [5]*Janssen Cilag, Birkerød, Denmark,* [6]*Janssen UK, High Wycombe, UK,* [7]*Consultant, Mumbai, India,* [8]*Janssen Global Services LLC, Raritan, NJ, USA*

OBJECTIVES: To assess the relative efficacy and safety of canagliflozin (CANA), a sodium-glucose co-transporter (SGLT) inhibitor, as add-on to metformin, compared to sulphonylureas (SU), pioglitazone, DPP-4s, GLP-1s and dapagliflozin. **METHODS:** Bayesian network meta-analysis was conducted based on a systematic literature review described separately. Outcomes of interest included HbA1c, weight and hypoglycaemia. Networks were based on treatment- and dose-specific nodes where possible. Non-informative priors were used; selection of fixed versus random-effect model was based on DIC. Studies causing inconsistency (identified through the comparison of direct and indirect evidence in the network) were identified with a clinical expert and excluded from the base case. **RESULTS:** 25/17/7 studies reported results at 26/52/104 weeks (w) respectively. HbA1c-reduction (2) at 26w/52w was best for exenatide 2mg and liraglutide 1.8mg. CANA 300mg had a higher reduction versus DPP-4s (h=-0.11 to -0.39) and dapagliflozin 10mg (==-0.12 to -0.38) across all time points; while CANA 100mg conferred at least as large reductions (==0.01 to -0.30 and 0.00 to -0.26 respectively). The analysis at 104w was conducted based on the pooling of SUs. CANA 300mg and 100mg ranked first/second before liraglutide 1.2mg/1.8mg (s=-0.11/-0.13 and -0.02/-0.04 respectively). Both CANA doses had higher weight-reductions than SU, DPP-4s and pioglitazone, and provided reductions comparable to GLP-1s and dapagliflozin. Odds ratios for hypoglycaemia versus SU ranged from 0.03 to 0.11 for DPP-4 and SGLT. **CONCLUSIONS:** NMA of add-on therapies to metformin suggests that CANA 300mg is associated with increased HbA1c-reduction versus DPP-4s and dapagliflozin while CANA 100mg provides at least similar effects. Additionally, results suggest increasing relative efficacy of CANA over time versus liraglutide and CANA reached at least as large HbA1c reductions as liraglutide at 104w. Weight reduction was comparable to GLP-1s and substantially higher than all other classes. All classes showed significantly less risk of hypoglycaemia compared to SU.

PRM194

ESTIMATING CHRONIC DISEASE PREVALENCE FROM CLAIMS DATA: REDUCING BIAS BY ACCOUNTING FOR DISEASED INDIVIDUALS WHO DO NOT GENERATE CLAIMS

Signorovitch J[1], De G[2], Drabo E[3], Guerin A[4]
[1]*Analysis Group, Inc., Boston, MA, USA,* [2]*Analysis Group, Inc., New York, NY, USA,* [3]*University of Southern California, Los Angeles, CA, USA,* [4]*Analysis Group, Ltee., Montreal, QC, Canada*

OBJECTIVES: Claims data are often used to estimate the prevalence of chronic diseases, typically by dividing the number of patients with disease-related claims (e.g., ≥1 or ≥2 claims) by the number of studied individuals. Such estimates will have a downward bias because not all diseased patients will generate disease-related claims within their enrollment period. This downward bias can be substantial for underserved diseases that lack effective treatments. We explored whether an empirical Bayes estimator for the number of diseased individuals who do not generate claims could improve the accuracy of claims-based prevalence estimates. **METHODS:** As an example, we studied the prevalence of a rare dermatological condition without any FDA-approved therapies. After accounting for enrollment time, individuals in a large nation-wide claims database were identified as having 0, 1, 2, 3, etc., disease-related claims. These counts were modeled using a mixture of Poisson distributions, with an unknown mixing distribution. Empirical Bayes approaches, which are frequently used to estimate numbers of unobserved species in ecological experiments, were used to estimate the number of diseased individuals without claims, and to provide adjusted prevalence estimates. **RESULTS:** Out of over 4 million individuals with at least one year of continuous enrollment, n=2,026 had disease-related claims, comprised of n=1,422 with one claim, n=317 with two claims, n=134 with 3 claims, etc. The traditional method for estimating prevalence identified 4.9 cases per 10,000 persons. After applying the empirical Bayes approach, the estimated prevalence increased to 7.9 cases per 10,000 persons, and became closer to published prevalence estimates based on non-claims data sources. **CONCLUSIONS:** In this example application, prevalence estimates based on claims data were increased by over 60% by using empirical Bayes approaches to account for large numbers of diseased individuals who did not generate claims. The increased prevalence estimates were more consistent with the published literature.

PRM195

APPLICATION OF COPULAS IN ECONOMIC EVALUATION

Crespo C[1], Monleon A[1], Díaz W[2], Rodriguez C[1], Brosa M[3], Rios M[1]
[1]*University of Barcelona, Barcelona, Spain,* [2]*Universidad de Antioquia, Medellin, Colombia,* [3]*Oblikue Consulting, Barcelona, Spain*

OBJECTIVES: To analyse the applicability of copulas distribution in economic evaluation. **METHODS:** We have analyzed data from an observational prospective study of patients with allergic rhinitis in Spain (n=498). Main data were direct cost (€2012) and Health Related Quality of Life (SF-12). We have calculated the goodness of fit for copulas (Gumbel copula, Clayton Copula, Frank Copula, Normal Copula, Plackett Copula and T copula) based on the empirical process comparing the empirical copula with a parametric estimate of the copula derived under the null hypothesis. We have used inversion of Kendall's tau method to fit copulas. A multivariate independence sample was generated to compare with copulas results. This process was replicated for a 100 times to obtain p-values by bootstrap method. **RESULTS:** Marginal distribution of direct cost was a 3-parameter Gamma distribution (shape=1.856, scale=0.00324, location=10.97). Marginal distribution of Health Related Quality of Life was associated to a 1- gamma (shape 2.9253 and scale 0.16104). P-value range were 0.093 to 0.144 for independent distribution, 0.004 to 0.031 for Gumbel copula, 0.246 to 0.522 for Clayton Copula, 0.545 to 0.814 for Frank Copula, 0.463 to 0.716 for Normal Copula, 0.373 to 0.628 for T Copula and 0.549 to 0.847 for Plackett Copula. Frank Copula and Plackett Copula had the best goodness of fit. Kendall's Tau for Fran Copula showed a correlation of -0.4212. **CONCLUSIONS:** Copulas distribution allows us to adjust better the non-lineal relation between cost and effectiveness. Furthermore, this kind of approach could improve probabilistic sensitivity analyses.