intervals HPV-based screening was more effective than cytology alone, with a relative reduction in cervical cancer incidence of 49%-90% compared to 33%-80% with cytology alone (depending on screening intervals). In HPV- compared to cytology screening the incremental gain in effectiveness was higher with extended screening intervals and the increase in harms lower. Based on the BHF, 12 of 17 screening strategies were dominated, including annual cytology, the current recommended standard in Germany. Biennial HPV-screening was similarly effective as annual cytology and reduced unnecessary treatment. Moving from biennial HPV- with cytological triage to annual HPV-screening alone results in an incremental harm-benefit ratio of 15-533 unnecessary treatments per additional prevented cervical cancer case (depending on screening adherence rate). **CONCLUSIONS:** The benefit-harm frontier is a useful tool to demonstrate the trade-off between expected gains and risks of different screening strategies. Based on our analyses, HPV-based cervical cancer screening is more effective than cytology alone, but has a higher risk of overtreatment when used in annual screening. In the German health care context, depending on screening adherence rates biennial or triennial HPV-screening for women ≥30 years is similarly effective as annual cytology with significantly reduced unnecessary treatments.

PRM2
EVALUATING WHETHER INCONSISTENCIES ARE PRESENT IN A MIXED TREATMENT COMPARISON OF TROUGH FORCED EXPIRATORY VOLUME IN 1 SECOND AT 12 WEEKS

Buckley F[1], Baldwin M[2], Keininger D[3], Jansen J[1], Cope S[4]
[1]Mapi, Boston, MA, USA, [2]Novartis Pharmaceuticals UK Limited, Horsham, West Sussex, UK, [3]Novartis Pharma AG, Basel, Switzerland, [4]Mapi, Toronto, ON, Canada

OBJECTIVES: To evaluate whether there are inconsistencies in the network of randomized controlled trials (RCTs) used for a network meta-analysis (NMA) comparing alternative long-acting bronchodilators among patients with moderate to severe chronic obstructive pulmonary disease (COPD) in terms of trough forced expiratory volume in 1 second (FEV1) at 12 weeks. **METHODS:** The change from baseline (CFB) in FEV as observed with placebo, tiotropium 18µg/5µg once daily (OD), salmeterol 50µg twice daily (BID), formoterol 12µg BID, aclidinium 400µg BID, glycopyrronium 50µg OD, indacaterol 75/150/300µg OD, formoterol 12µg+ tiotroipum 18µg BID/OD, indacaterol 150µg+ tiotropium 18µg OD, and indacaterol 110µg+ glycopyrronium 50µg OD in RCTs identified with a systematic literature review were synthesized with a NMA. Where possible, treatment estimates from fixed effect (FE) and random effects (RE) NMA models (assuming consistency between direct and indirect evidence) and independent means (IM) models (pooled direct evidence) were compared to assess whether any inconsistencies in the network were present. **RESULTS:** Thirty-two RCTs identified through a systematic literature review were included in the analysis. Direct evidence was available for the monotherapies versus placebo, the combination therapies versus tiotropium, for indacaterol+ glycopyrronium versus placebo, and for tiotropium versus salmeterol. The largest differences between the estimated treatment effect estimates from the NMA and the IM models were observed for the comparisons between indacaterol 150µg versus tiotropium (FE difference=0.025 [95% Credible Intervals (95%CrI): 0.002, 0.047]; RE difference=0.027 [95%CrI: -0.007, 0.61]), indacaterol+ glycopyrronium versus placebo (FE difference=-0.022 [95%CrI: -0.053, 0.008]; RE difference=-0.018 [95%CrI: -0.059, 0.022]), and indacaterol+ glycopyrronium versus tiotropium (FE difference=0.011 [95%CrI: -0.014, 0.036]; RE difference=0.015 [95%CrI: -0.024, 0.053]). **CONCLUSIONS:** Based on a comparison of the findings of a NMA and IM models, some minor inconsistencies in treatment effects for trough FEV1 at 12 weeks were identified that will be explored through additional sensitivity analyses.

PRM3
TESTING THE EUNETHTA INTERNAL VALIDITY OF RANDOMIZED CONTROLLED TRIALS GUIDELINE AND TOOL IN HUNGARY

Huszti Z, Juhasz J
National Institute for Quality- and Organizational Development in Healthcare and Medicines, Budapest, Hungary

OBJECTIVES: The reliability of the results of a randomized trial depends on the extent to which potential sources of bias have been avoided. We tested the EUnetHTA Internal Validity guideline so that to harmonize our risk of bias assessments with the European standard and finally to improve the reliability of relative efficacy and cost effectiveness assessments for decision makers in Hungary. **METHODS:** We translated the risk of bias standardized assessment questions of the EUnetHTA Internal Validity (of randomized controlled trials) into Hungarian. We first chose ten studies for internal validity testing from the ones that were submitted for reimbursement at the beginning of 2012, and their results were used for health economy assessment. **RESULTS:** We found adequate randomization sequence generation in seven studies and we marked it unclear in three trials (e.g.: lack of information, age related sequence generation). The allocation concealment was labeled suitable in six studies (e.g.: IVRS, IWRS) and unclear in four trials. All studies could be classified according the the type of blinding. We found selective reporting in one trial where the non-inferiority results in the per-protocol population were not published. We rated the risk of bias low for eight trials and high for two trials due to unclear sequence generation and publication bias. We also evaluated 77 endpoints and we labelled 22 endpoints with high risk of bias. The most common reasons for high risk ratings were the not appropriately implemented ITT principle and selective reporting. **CONCLUSIONS:** The EUnetHTA guideline gives an opportunity to estimate the risk of bias of randomized controlled trials in a structured and harmonized way without leaving out any important considerations. The results of the internal validity evaluation can lead the focus of interest to those endpoints where the sensitivity analysis is requisite in the health economic models.

PRM4
CLINICAL OUTCOMES ASSESSMENTS IN SCHIZOPHRENIA: A SYSTEMATIC LITERATURE REVIEW

Cooke C[1], Heatley R[2], Galani Berardo C[3], Johnson KI[2], Kasper S[4]
[1]Complete Clarity, Macclesfield, UK, [2]Complete Clarity, Macclesfield, Cheshire, UK, [3]F. Hoffman La Roche, Basel, Switzerland, [4]Medical University of Vienna, Vienna, Austria

OBJECTIVES: There is a growing interest from health technology assessment agencies in determining the clinical outcomes assessments and endpoint strategies that can establish treatment benefits. We describe a systematic literature review of endpoints and outcomes used in schizophrenia trials to determine treatment benefit. **METHODS:** The therapies selected in the search strategy included pharmacological interventions, cognitive-behavioural therapies, family intervention, and music therapy. These were chosen to reflect the range of interventions in current use, and to allow comparison between outcomes reported for different therapies. The search terms were designed to include all outcomes for each therapy area, and were used to search four electronic databases for published English language studies. Randomised controlled trials (RCTs) were retrieved if they included patients with schizophrenia treated with the chosen therapies, and clinical outcomes from a predefined list (e.g. symptom improvement, functionality, quality of life, remission rates, response rates, and recovery). **RESULTS:** Of 2,221 RCTs identified, 271 progressed to data extraction; 225 assessed pharmacological interventions and 46 non-pharmacological interventions. Approximately 76 outcomes were measured across the trials. The most common scale used in pharmacological trials was the Positive and Negative Syndrome Scale (PANSS) total score (76.9%), and the PANSS positive subscale in non-pharmacological trials (50%). However, even within the common outcomes, the specified level of reduction to define a relevant response varied; among trials reporting PANSS total, five different levels of reduction were defined (≥20%, ≥25%, ≥30%, ≥40%, ≥50%). Common outcomes were also measured differently in terms of improvement from baseline and proportion achieving response/remission, with little consensus on clinical meaningfulness. **CONCLUSIONS:** The RCTs included in this review reported a broad range of outcomes, making comparison of different therapies a complex task. The disparity in outcomes between pharmacological and non-pharmacological outcomes scales highlights the challenges in designing trials to demonstrate clinical benefit.

PRM5
MULTI-DIMENSIONAL CAPTURE OF PATIENT-RELEVANT ENDPOINTS IN REGULATORY TRIALS AND HEALTH TECHNOLOGY ASSESSMENTS IN ONCOLOGY TWO YEARS AFTER INTRODUCTION OF THE GERMAN AMNOG HEALTH CARE REFORM

Dintsios CM[1], Knoerzer D[2], Duenne AA[2], Schwartz FW[3], Ruof J[2]
[1]German Association of Research-based Pharmaceutical Companies (ufa), Berlin, Germany, [2]Roche Pharma, Grenzach-Wyhlen, Germany, [3]Hannover Medical School, Hannover, Germany

OBJECTIVES: With the introduction of AMNOG in January 2011, an early benefit assessment (EBA) was required for new medicines in Germany. EBAs are based on the additional therapeutic benefit of a drug on patient-relevant endpoints (PREs). We compared the acceptance of PREs for oncology in regulatory trials, and in EBAs conducted by German health technology assessment (HTA) bodies. **METHODS:** EBAs on oncology drugs and the respective regulatory trials were reviewed. The Federal Joint Committee (G-BA) website was used to obtain manufacturers' value dossiers, Institute for Quality and Efficiency in Health Care (IQWiG) assessments, and G-BA resolutions. Acceptance of endpoints in the dimensions of mortality, morbidity and quality of life (QoL) by HTA bodies, IQWiG and G-BA, were compared to those accepted for regulatory trials. Data on endpoints used in regulatory trials were obtained from the manufacturers' value dossiers. **RESULTS:** Overall survival (OS) and measures of disease morbidity, such as progression-free survival (PFS), were generally accepted in regulatory trials. OS was accepted by IQWiG and G-BA as a mortality endpoint for evaluating additional benefit. Widely accepted morbidity endpoints such as PFS were not deemed patient-relevant by IQWiG and G-BA. In general, QoL questionnaires used in regulatory trials were accepted by the HTA bodies, although minor variability between questionnaires led to some exclusions from the HTA evaluations and the obtained QoL data revealed a number of missing values. **CONCLUSIONS:** HTA and regulatory bodies largely agree on the acceptance of mortality and QoL endpoints typically evaluated in oncology. Considerable variability was observed in the acceptance of PREs in morbidity. Evaluating additional benefit based only on mortality and QoL endpoints underestimate the potential value of new drugs. Multiple endpoints, which capture all three dimensions, should be evaluated in regulatory trials and accepted by IQWiG and G-BA to confirm patient-relevant additional benefit.

PRM6
THRESHOLD SELECTION IN BIOMARKERS USING COX REGRESSION. AN APPLICATION TO NON-SMALL-CELL LUNG CANCER

Luaces P[1], Sánchez L[1], Ballesteros J[2], García B[1], Rodríguez PC[1], Viada C[1], Popa X[1], Crombet T[1], Lage A[1]
[1]Center of Molecular Immunology, Havana, Cuba, [2]University of the Basque Country UPV/EHU, Leioa, Spain

OBJECTIVES: To select thresholds for predictive biomarkers using Cox regression. **METHODS:** We used data from a Cuban trial designed to assess the efficacy of immunotherapy against the epidermal growth factor (EGF) to test our approach. The trial included 122 patients diagnosed with non-small-cell lung cancer (NSCLC) who had basal EGF concentration available. The EGF concentration was analysed as a predictor of immunotherapy success over the range of all possible values of the biomarker (w [a,b]). For each w_i, patients with $w > w_{i0}$ were selected and a Cox model adjusted to assess survival. We then identified the w_{0s} with significant treatment results to find (a) the lowest biomarker threshold where the effect of treatment was significant and also to find (b) the biomarker threshold that reflected the highest difference between treatments. **RESULTS:** For NSCLC we observed that EGF concentration thresholds range from 870 pg/ml to 2000 pg/ml were significant. At the lowest threshold (870 pg/ml) the immunotherapy group showed a 6-month difference for the median survival (p = 0.022) whereas at the threshold that showed the maximum difference between treatments (EGF = 1750 pg/ml) the immunotherapy group presented a 10-month difference for the median survival (p = 0.004). **CONCLUSIONS:** The evaluation of p-values of the effect of treatment for each w_0 [a,b] allows the selection of the thresholds where the treatment result is significant. Whereas the