

results section, we split up descriptive information from primary studies (sex rates) and analytic approaches (considering sex/gender in the assessment of risk of bias, presenting disaggregated data by sex/gender or subgroup or heterogeneity analyses). We used 'not applicable' to denote a situation where insufficient primary studies or data on estimates did not enable to conduct the intended analyses (e.g. meta-analysis, subgroup analysis). We performed descriptive statistics and regression analyses to assess associations between authors' gender and sex/gender reporting.

**Results** 556 reviews were screened, of which 91 were excluded due to withdrawal (19,8%) or sex-specific disease (80,2%). Our analysis comprised 465 studies, including 2 prognosis, 4 methodology, 5 overview, 20 diagnostic and 421 intervention reviews. Women represented 53,8% (n = 250) and 38,9% (n = 181) of first and last authorship, respectively, while in 25,3% of reviews both authors were women. 85,6% of authors belonged to high-income countries. 7.5% (n = 35) of reviews reported on sex in the abstract, 17.6% (n = 82) in the methods and 61.4% (n = 285) in the results section. Of these 285, 64.7% provided descriptive results and 16,3% had an analytic approach. In the discussion section, 13.5% (n = 63) of reviews addresses sex-related findings. Only 4 studies scored positive in all 4 sections. Studies with female first-last authorship had a non-significant increased probability of reporting on sex (RR [95% CI] 1,24 [0.68-1.92])

**Conclusions** Consideration of sex and gender in Cochrane reviews is scarce. This prevents from generating inclusive and unbiased health research and inhibits effective population-wide translation of results. While women are less likely to be last authors, studies with a female first or last author show an increased (although non-significant) probability of reporting on sex.

## 28 PSYCHOSOCIAL CONSEQUENCES OF INVITATION TO COLORECTAL CANCER SCREENING – A DANISH LONGITUDINAL COHORT STUDY

<sup>1,2</sup>Jessica Malmqvist, <sup>1</sup>Volkert Siersma, <sup>1</sup>Mie Sara Hestbech, <sup>1</sup>Dagný Rós Nicolaisdóttir, <sup>3</sup>Christine Winther Bang, <sup>2,3</sup>John Brodersen. <sup>1</sup>Center for Research and Education in General Practice, Copenhagen, Denmark; <sup>2</sup>Research Unit for General Practice in Region Zealand, Copenhagen, Denmark; <sup>3</sup>The Center for Research and Education in General Practice, Copenhagen, Denmark

10.1136/bmjebm-2019-EBMLive.36

**Objectives** Harms of screening are generally not well reported. Harms of colorectal cancer (CRC) screening may include negative psychosocial consequences from the attention being drawn to disease through the screening invitation, the screening procedure itself and from fear of the screening result. Hence, it is important to measure psychosocial consequences before and after invitation to screening, at screening and after screening. Moreover, measurement of psychosocial consequences in cancer screening settings should be performed using questionnaires with high content validity and adequate measurement properties. Despite this fact, few CRC screening studies investigating psychosocial consequences have performed a baseline measurement before and after invitation to screening in both a screening and a control cohort. Moreover, none of the studies has used a questionnaire with high content validity and adequate measurement properties.

Therefore, the objective of this study was to investigate the psychosocial consequences of invitation to CRC screening in a

Danish CRC screening naïve cohort using a condition-specific questionnaire validated with Item Response Theory Rasch Models.

**Method** A random sample of 1000 screening invitees scheduled for screening in March 2017 was matched with 1000 control persons scheduled for screening in November and December 2017. We sent a questionnaire to both study groups five weeks prior to the invitation of the invitees and another questionnaire to both groups three days after the invitees' screening invitations were sent by post.

The primary outcome was psychosocial status, measured with the condition-specific questionnaire Consequences of Screening for CRC (COS-CRC).

We analysed the mean COS-CRC score at each measurement for each COS-CRC scale compared between the study groups using multivariable regression models, adjusting for potential confounders. We adjusted for differential non-response by weighting the observations that were available at the follow-up measurement by the inverse of the probability of not being missing. We also adjusted for repeated measures and weighting. To allow for multiple testing, a p-value <0.01 was considered significant.

**Results** Results and conclusions will be presented at the conference.

**Conclusions** Results and conclusions will be presented at the conference.

## 29 INCIDENCE OF SUBSTUDIES PUBLICATION AFTER ORIGINALLY NEGATIVE CLINICAL TRIALS AND RATE OF POSITIVATION: A CALL FOR SCIENTIFIC INTEGRITY

<sup>1,2</sup>Luis Correia, <sup>1</sup>Diego Rabelo, <sup>1</sup>Alessandra Caldas, <sup>1</sup>Bruno Góes, <sup>1</sup>Janine Magalhães, <sup>3</sup>Daniel Medina, <sup>1</sup>Denise Matias. <sup>1</sup>Bahiana Medical and Public Health School, Salvador, Brazil; <sup>2</sup>Hospital São Rafael, D'Or Institute for Research and Education, Salvador, Brazil; <sup>3</sup>School of Medicine, Federal University of Bahia, Santo Antonio de Jesus, Brazil

10.1136/bmjebm-2019-EBMLive.37

**Objectives** After a negative study, substudies may emerge providing a positive spin on the hypothesis, based on subgroup or secondary outcome analyses. This approach suffers from multiplicity problem, imprecision and higher risk of type I error.

To describe the incidence of substudies testing the same hypothesis over a period of 5 years after publication of a primarily negative randomized clinical trial; to access the most frequent subanalysis method, how explicit the exploratory nature of the analysis were and the rate of successful positivation of results.

**Method** We searched all negative randomized trials published in the *New England Journal of Medicine* (NEJM) during the year 2014. Then, we made a highly sensible PUBMED search over the following 5 years in order to detect subsequent articles testing the same hypothesis in the same dataset throughout secondary analysis. Methods utilized, lack of recognition as secondary analysis was described and rate of positivation were described.

**Results** During the year 2014, 46 negative randomized clinical trials were published in the NEJM. Over five years, 14 of those articles had subsequent publication of subanalysis in PUBMED indexed journals, an incidence of 30% (95% CI = 19% - 45%). The most frequent method was secondary endpoints analysis (67% of substudies) and the remaining made use of subgroup analysis. Half of those analysis were not

defined *a priori* (*post-hoc* analysis). Thirty-two percent of studies did not leave explicit in the conclusion it was an exploratory analysis and the success rate of positivation was 91% (95% CI = 80% - 97%).

**Conclusions** Subsequent publication of subanalysis from originally negative trials is frequent, commonly not defined *a priori*, commonly not explicit about the exploratory nature and highly successful in positivating results. This suggests lack of ecosystem scientific integrity.

30

### APPLYING A THRESHOLD CONTROLLED TRIAL IN ORDER TO ASSESS THE EXTERNAL VALIDITY OF RANDOMIZED CONTROLLED TRIALS AND TO OPTIMIZE THE SELECTION OF PATIENTS FOR TREATMENT TO MAKE EVIDENCE MORE RELEVANT

Huw Llewelyn. *Aberystwyth University, Aberystwyth, UK*

10.1136/bmjebm-2019-EBMLive.38

**Objectives** Randomized controlled trials (RCTs) are the 'gold standard' method used to assess the efficacy of treatments. However, patients may be reluctant to be subjected to another RCT once efficacy has been firmly established, e.g. in order to examine the effect on efficacy of using a different selection criterion for treatment. A different approach to RCTs would help in such a situation. Regression discontinuity is one way of doing this in order to assess the external validity of a RCT but it cannot be used also to assess the effect of changing treatment selection criteria. A method will be described here that allows the result of a RCT to be predicted based on the assumption of constant odds ratio after patients above and below a threshold are offered different interventions.

**Method** Subjects are allocated to a control limb if the results of the test used to select them for the trial are on one side of some threshold and allocating them to a treatment limb if the results are on the other side of the threshold. The results are interpreted by assuming that the distribution of pre-treatment test results in those with a subsequent outcome is the same for those in the treatment and placebo limbs. This is the assumption made when using relative risk and odds ratios to apply the result of an RCT to patients with different baseline risks. This approach is illustrated with a data set from a RCT where the diagnostic test was the albumin excretion rate, the treatment was an angiotensin receptor blocker and the outcome was biochemical nephropathy. The result of the full RCT is compared with the result of a 'threshold controlled trial'.

**Results** When curves are constructed to show the probabilities of an outcome (nephropathy) on placebo and treatment for each diagnostic test result by using all the data from the RCT and from only the part of the data that would have been available from a threshold trial, the results were very similar, the small differences being readily explicable due to minor stochastic variation. In particular, the distribution of pre-treatment AER in those with and without subsequent nephropathy was the same in the treatment and placebo limbs. The proportion estimated to respond to the RCT in the threshold study was the same as in the full RCT. Also, the AER threshold when treatment began to be effective was similar in the threshold study and in the RCT. The results are described in a preprint.<sup>1</sup>

**Conclusions** Provided that suitable controls are in place, (e.g. double blinding) it appears that a threshold study can predict the result of a RCT. This is invaluable as it promises to allow studies to be performed to compare interventions when randomisation is not possible. It would allow the external validity of a RCT result to be assessed during day to day care. It would also allow new diagnostic tests to be assessed in order to see how well they can select patients for treatment in order to avoid over-treatment.

### REFERENCE

1. Llewelyn H. *Estimating the result of randomised controlled trials without randomisation in order to assess the ability of diagnostic tests to predict a treatment outcome*. 2018. <http://arxiv.org/abs/1808.09169>

31

### (IN)CONSISTENCY OF RECOMMENDATIONS FOR EVALUATION AND MANAGEMENT OF HYPERTENSION

<sup>1,2</sup>Martin Mayer, <sup>1,3</sup>Brian Alper, <sup>4,5</sup>Amy Price, <sup>6</sup>Esther van Zuuren, <sup>7</sup>Zbys Fedorowicz, <sup>8</sup>Allen Shaughnessy, <sup>1</sup>Peter Oettgen, <sup>9</sup>Glyn Elwyn, <sup>10</sup>Amir Qaseem, <sup>11</sup>Ilkka Kunnamo, <sup>4</sup>Urvi Gupta, <sup>12</sup>Deborah Carter, <sup>13</sup>Michael Mittelman, <sup>14</sup>Carla Berg-Nelson. <sup>1</sup>EBSCO Health, Ipswich, MA, USA; <sup>2</sup>Cone Health, Greensboro, NC, USA; <sup>3</sup>University of Missouri-Columbia School of Medicine, Columbia, MO, USA; <sup>4</sup>Stanford Medicine X, Stanford, CA, USA; <sup>5</sup>University of Oxford, Oxford, UK; <sup>6</sup>Leiden University Medical Centre, Leiden, Netherlands; <sup>7</sup>Veritas Health Sciences Consultancy, London, UK; <sup>8</sup>Tufts University School of Medicine, Medford, MA, USA; <sup>9</sup>The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH, USA; <sup>10</sup>American College of Physicians, Philadelphia, PA, USA; <sup>11</sup>Duodecim, Helsinki, Finland; <sup>12</sup>Murdy Consultant Group, Newark, OH, USA; <sup>13</sup>American Living Organ Donor Fund, Philadelphia, PA, USA; <sup>14</sup>Society for Participatory Medicine, Nutting Lake, MA, USA

10.1136/bmjebm-2019-EBMLive.39

**Objectives** To systematically assess the consistency of recommendations regarding hypertension management across clinical practice guidelines (CPGs) and electronic point-of-care (POC) resources

**Method** We identified hypertension management recommendations from eight CPGs and two POC resources in April 2018. We described discrete and unambiguous specifications of the population, intervention, and comparison states to define a series of reference recommendations. Three raters reached consensus on coding the direction and strength of each related recommendation made by each CPG and POC resource.

For each reference recommendation, we analyzed the rate of consistency for direction and strength. We did this for the eight CPGs and for the group of ten recommendation sources. We also conducted sensitivity analyses testing the robustness of our findings to the exclusion of recommendation statements of 'insufficient evidence' and to the exclusion of single recommendation sources. We also assessed the CPG and POC resources for evidence of public and patient involvement, patient-facing information, and shared decision-making tools, and we involved patient and public representatives in this assessment.

**Results** Considering all 10 recommendation sources, 12 of 71 recommendations (16.9%) were consistent in direction and strength, 21 (29.6%) consistent in direction but inconsistent in strength, and 38 (53.5%) inconsistent in direction. Considering only the CPGs, 25 recommendations (35.2%) were consistent in direction and strength, 13 (18.3%) consistent in direction but inconsistent in strength, and 33 (46.5%) inconsistent in direction. Excluding 'insufficient evidence' ratings did not explain the inconsistency, and a leave-one-out sensitivity