

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.¹

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

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