extracted relevant data about the reported clinical trials. For this assessment, a coding book was operationalized and specifically tailored to the context and role of IBs. The coding book included items on the reporting of a) a search strategy to find relevant evidence b) a method of evaluating the risk of bias of included studies and c) a data collection and syntheses procedure. The coding book included a second set of items to evaluate the reporting quality of all clinical trials listed in the IB.

Results A total number of 321 clinical trials were identified in the 30 IBs. Despite the relatively similar aims of evidence reporting in IBs and systematic reviews, no IB reported a search strategy that could have demonstrated how comprehensive the reported clinical evidence is. Furthermore, no IB presented a risk of bias assessment for the reported studies. All of the included IBs reported the results of relevant clinical trials exclusively in a narrative form and did not use any method to synthesize the evidence to create a concise overview of the results. The majority of the 321 trials reported the sample size (96%, $n = 308$), blinding (83%, $n = 266$), and randomization (66%, $n = 212$). Specific details, however, about sample size calculation, randomization methods or blinding procedures were reported for less than 1% of the 321 studies reported in the IBs. Baseline characteristics and participant flow were reported for 10% ($n = 31$) of all 321 trials.

Conclusions IBs serve the important function to compile the evidence that shall justify the approval and conduct of a planned clinical trial. Our results in a random subsample of 30 IBs that reported on 321 clinical trials show that IBs do not adhere to any established principles of evidence synthesis known from systematic reviews. In sum, our findings cast doubts about the robustness and transparency of decision-making based on the evidence provided by IBs. In particular, it raises the question whether they enable the above-mentioned reviewers to conduct a meaningful risk-benefit assessment that is ethically and legally required to justify clinical trials. In the presentation, we will outline recommendations on how to improve evidence reporting in IBs.

43 GENERATING GUIDELINE RECOMMENDATIONS – EXPLORING PATIENT VALUES?

¹Eve O'Toole, ¹Louise Murphy, ²Leah McClimans. ¹National Cancer Control Programme, Health Service Executive, Dublin, Ireland; ²University College, Cork, Ireland

10.1136/bmjebm-2019-EBMLive.51

Objectives The National Cancer Control Programme (NCCP) began developing national evidence based clinical guidelines in 2012. In 2017 we decided to generate our guideline recommendations using the GRADE process. There is clear guidance on appraising the quality of the evidence and performing bias assessments. However, the challenge begins with the integration of the best evidence, benefit and harm, patient values and resources when generating recommendations. Patients are full members of the guideline development groups; however, the question remains how do we truly capture patient values? By inviting a philosopher/medical ethicist to be a member of the GDG, integrating patient values has been explored in our update of the national clinical guideline diagnosis staging and treatment of patients with prostate cancer.

Methods A guideline development group was established with representatives from diagnostics, surgery, oncology, nursing,

research, patient representatives and a philosopher/medical ethicist. The guideline questions were generated by the GDG. Literature searches, appraisal of the evidence and data extraction were performed by the NCCP research team. Generation of recommendations by the GDG follows a formal standardised protocol using an evidence to decision framework developed in NCCP. The following items are considered; the evidence, its quality its generalisability/directness, benefit and harm, patient values, resources and cost. The evidence statements are written in real time and recommendations graded strong or weak. During the first meeting to generate the guideline recommendations the philosopher challenged the group to consider the meaning behind words such as. generalisability, equity and justice. We also explored patient important values such as pain, certainty, anxiety, quality of life, trust, choice and autonomy.

Results Our evidence to decision framework has been restructured to contain a matrix to explore patient values for each option under discussion. New headings have been added with prompts to stimulate discussion with emphasis on patient important outcomes, equity, acceptability and a richer exploration on the concept of generalisation. Each evidence statement now has clear headings for quality of the evidence, benefit and harm, patient values and resources. Framing the discussion in this manner resulted in a richer conversation. For the first clinical question under review there was high level evidence, however the clinical members in the group were concerned that recommending a course of action could lead to greater inequity within our health system. This has resulted in escalation of the GDG recommendations within the health service with service redevelopment proposed.

Conclusions Involvement of the philosopher/medical ethicist led to the development of clear questions, prompts and a matrix, in the patient values domain. This resulted in the GDG having greater confidence in exploring patient values when generating guideline recommendations. This would challenge guideline developers to expand the membership of their GDG to include this expertise. Of most importance this led to a richer discussion and resulted in more consideration being given to a recommendation even in the presence of high-level evidence. As a result, there is earlier engagement with the services to enable implementation of the recommendations with fidelity to ensure we do not create further inequity within the health services.

44 OPENING THE BLACK BOX OF 'USUAL CARE' AND FINDING A BLACK HOLE: A NUMERICAL SYSTEMATIC REVIEW ON 'USUAL CARE' CONTROL GROUPS IN STROKE REHABILITATION RCTS

¹Chiara Arienti, ²Chiara Arienti, ²Riccardo Buraschi, ²Joel Pollet, ²Massimiliano Gobbo. ¹University of Brescia, Senago, Italy; ²IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy

10.1136/bmjebm-2019-EBMLive.52

Objectives Ethically, in every experimental trial new methods should be tested against the 'best' or, if not defined, the 'usual' care according to current Guidelines. To understand the latter, we focused on stroke rehabilitation as a case in point because stroke is a leading cause of disability and rehabilitation is a complex intervention whose need is increasing due to growth of chronicity and disability (World Health Organization). Our aim was to appraise stroke rehabilitation

interventions provided as 'usual care' and the terminology used, comparing high to low quality randomised-controlled trials (RCTs).

Method Systematic review of RCTs (2006–16). Inclusion criteria were as follows: stroke survivor patients, intervention, rehabilitation, control: rehabilitative 'usual care', outcome: lower limb function. We used the Cochrane 'risk of bias' tool, rating studies as low or high quality. We identified the terminology used to describe the Control Group Rehabilitation Program (CGP), performed a knowledge synthesis process and conducted a frequency analysis to sort the heterogeneity through the itemised identification of the CGP contents. Two quality groups of studies (high-low risk of bias) have been compared.

Results We included 86 publications (23% low-risk of bias). Nine per cent of articles did not describe the CGP. In the remaining 78, we identified 64 interventions: 53 were proposed once, 8 were proposed twice and 3 were proposed in 3 papers. Two interventions (gait and balance) were proposed in 52% and 51% of papers, respectively. Results did not differ in the two RCT quality groups.

Conclusions This case study on stroke rehabilitation has shown that the term 'usual care' in CGP is not adequately used: a good terminology to define interventions is missing and respecting CONSORT checklists does not avoid these problems. These results are clear, and involve authors, reviewers and both field-specific and generalist journals. Nevertheless, they should be verified by future studies in other fields, with this or other methodologies. Reporting guidelines should probably give better expert guidance on this issue.

45 FREQUENCY OF SPIN REPORTING IN RANDOMIZED CONTROLLED TRIAL PUBLISHED BY INDIAN AUTHORS

¹Naveen Kumar, ²Hariohm Katheresapandian, ¹Saravan Kumar. ¹Saveetha college of physiotherapy, Chennai, India; ²Spring physiotherapy center, Chennai, India

10.1136/bmjebm-2019-EBMLive.53

Objectives Reports that distorts the interpretation of results is referred as spin and it has been shown to be common in randomized controlled trials. Spin can be intentional and unintentional, and studies reported its presence in several specialties. It can mislead the health care practitioner in decision making and can affect wellbeing of patient. The objectives of this study were to estimate the frequency of spin in both abstract and main text in physiotherapy clinical trials published from India.

Method Pubmed was searched for randomized controlled trial published by Indian physiotherapists between January and November 2018. Articles included were parallel group randomized controlled trial with identified primary outcome as non-significant and published by Indian authors. Abstracts and full text were retrieved. Two authors independently scored each article for presence of spin in abstract and main text result, discussion and conclusion by using pre-tested strategies defined by Boutron and colleagues. Third author was contacted in case of conflicts. The frequency of spin was estimated in all studies

Results 129 relevant articles were identified in Pubmed search. In which 32 articles met the inclusion criteria. Spin was identified in results and conclusion section of the abstract of 27 (84%) and 30(93.8%) studies, respectively. In main text

results, discussion and conclusion section spin was identified in 22(68%), 26(81.3%) and 29(90.6%) studies, respectively. Spin in title was present in one study.

Conclusions Strategies of spin is frequently included in reporting results in physiotherapy related trials published from India. Readers should be aware of possibilities of reporting bias and should avoid relying on abstract as a reliable report. Peer reviewers and editors need awareness of this issue to avoid the treatment recommendation that are not supported by data.

46 CHILDREN'S FITNESS AND HEALTH: AN EPIC SCANDAL OF POOR METHODOLOGY, INAPPROPRIATE STATISTICS, QUESTIONABLE EDITORIAL PRACTICES AND A GENERATION OF MISINFORMATION

Jo Welsman, Neil Armstrong. University of Exeter, Exeter, UK

10.1136/bmjebm-2019-EBMLive.54

Objectives Over 30 years ago we demonstrated the poor criterion validity of a popular fitness test, the 20 m shuttle run or 'bleep' test (20mSRT). We discounted the test and assumed that others with demonstrable validity and reliability would replace its use in research. Around then, our attention was drawn to an eloquent but obscure paper by JM Tanner (1949) which detailed the fallacy of simple division by body mass to accommodate body size differences in physiological function. Tanner described how incorrect analyses led to patients having 'no more formidable disease than statistical artefact'. Aware of the significance of this paper for our own field, over the next 15 years we published numerous data and tutorial papers demonstrating appropriate methods to measure and interpret cardiorespiratory fitness (CRF) during growth. Not only is the 20mSRT not a valid estimate of measured CRF, it predicts values expressed in simple ratio with body mass.

Method Despite our efforts, the past 10 years have seen a global explosion in published research studies of children's CRF anchored in these flawed methodologies. Data from millions of children worldwide have been collated into international norms, used to examine present and predict future cardiovascular and metabolic health, and to identify *individual* children who warrant intervention to reduce their risk of future cardiovascular disease – the raising of 'clinical red flags'. Data from these studies' present patterns of temporal changes in CRF which directly conflict with rigorously collected and appropriately analysed laboratory data. The 20mSRT test is being supported by international movements as a way of monitoring physical activity levels although objective data reveal the two to be unrelated. Moreover, clinical populations of children with serious life-limiting conditions are being put through maximal laboratory exercise testing with conclusions about their health status being made upon an inappropriate statistical analysis.

Results We believe the continued use of these flawed methodologies in vast numbers of children world-wide to be ethically and morally indefensible. By way of response we have, within the past 12 months: submitted 7 original data papers based upon extensive cross-sectional and longitudinal data founded on over 2000 rigorously determined individual assessments – all of which provide details of and recommendations for statistically justified analytical methods; we have submitted 7 editorial/commentary pieces to paediatric medical, sports