the number of people approached or assessed for eligibility, so that their data can contribute to evidence on optimising RCT participation.

22 COMPARING STATISTICAL SIGNIFICANCE BASED ON P-VALUES WITH THE PROBABILITY OF REPLICATING A RESULT LESS EXTREME THAN THE NULL HYPOTHESIS TO MAKE EVIDENCE MORE REPLICABLE

Huw Llewelym. *Aberystwyth University, Aberystwyth, UK*

10.1136/bmjebm-2019-EBMLive.103

Objectives It has been suggested that the term 'statistically significant' in relation to P-values should be 'retired' as it is regarded as a major cause of misinterpretation of scientific data. It is also associated with less frequent than expected replication of studies, especially in the social sciences and medicine, which also may be having a negative impact on evidence-based healthcare. The object of this presentation is to use the familiar reasoning processes of medicine as a guide to reasoning with scientific data in order to avoid current pitfalls. The steps used in interpreting clinical data include checking the reliability of symptoms, signs and test results and if these facts are reliable then using them as evidence to make predictions and using the predictions and their probabilities to make decisions. The object of this presentation is the above first step: to assess the reliability of scientific study results using analogous reasoning.

Method The reliability of a diagnostic test or study result can be assessed by estimating the probability with which it will fall within any 'true' range if the study or observation were repeated until nearly an infinite number of observations were made. For example, this range might be on one side of a test's 'lower end of normal' or a 'null hypothesis' (e.g. a true result of no difference between the effect of a treatment or placebo). The calculation is based on regarding a data set as a subset of a larger data set within which the probability of each possible true value is set to be the same. The resulting calculated 'idealistic' probability of long-term replication is based on the ideal situation when the observations are conducted in an impeccably consistent way. Failing this the probability of 'non-impeccability' is used to estimate a 'realistic' probability of replication.

Results If the data can be modelled with a 'normal' or other symmetrical distribution, then the probability of a true result less extreme than the null hypothesis is shown to be exactly 1-P. If the distribution is not symmetrical then it is only approximately equal to 1-P. If a Bayesian prior probability is specified, then it is shown that this can be incorporated in to the calculated probability of replication. The result of a prior pilot study and the result of subsequent studies (e.g. in a meta-analysis) can also be incorporated to provide a probability of replication within any specified range given all the evidence. The probability of getting the same P value or better after repeating a study is 50%. If the probability of impeccably consistent methodology is low (e.g. due to cherry picking, etc) then the realistic probability of replication will be lower than the idealistic probability.[1]

Conclusions This approach is based on an improved understanding of the relationship between P-values and the probability of replication according to well recognized Bayesian principles. Instead of assuming uniform prior probabilities, it actually creates this condition with a new universal set that contains the study data and within which the prior probabilities of true outcomes are uniform. It thus replaces the confusing definition of the P value (the probability of the study result that was seen and others more extreme that were not seen conditional on a null hypothesis) with a probability of replication less extreme than the null hypothesis. This approach avoids the pitfalls associated with classifying P-values as being 'significant' or 'not significant' and over-estimating the probability of replication.

REFERENCES

1. Llewelyn H. Replacing P-values with frequentist posterior probabilities of replication—When possible parameter values must have uniform marginal prior probabilities. *PLoS ONE*, 2019, **14**(2): e0212302. https://doi.org/10.1371/journal.pone.0212302

23 EVIDENCE-BASED MEDICINE CHALLENGES IN NEW ANTICANCER DRUGS

Tatiane Ribeiro. *University of Sao Paulo, Sao Paulo, Brazil*

10.1136/bmjebm-2019-EBMLive.104

Cancer become epidemic, it caused 9,5millions deaths in 2018 and the disease incidence yearly increase. Anticancer drugs costs have been increasing and its alarming worldwide healthcare systems, so evidence-based medicine in this area impact directly on financial aspects, patient access and best healthcare practice. Lately Worldwide Health Agencies such as FDA, EMA and ANVISA (Brazilian Health Agency) have made flexible requirements for drug approval. Phase 2 and Non-randomized studies can be used as single reference to new drug approval, and mainly in oncology, usually those trials use surrogate outcomes as the primary endpoints, such as progressive-free survival (PFS) and objective response rate (ORR). Those surrogate endpoints have major impact on evidence-based assessment, they are not patient-important outcomes. Patients have to be active on their healthcare decisions, and those outcomes are barriers to shared decision making. Moreover, outcomes based on PFS can be biased due to many causes: First, there are no standard for measurement (commonly use radiological parameters); Second, progression is subjective, and the measurement can be different among researchers; Third, frequency of assessment can influence the results and the last, there is no relationship that the progression control can generate benefit to patient. In oncology, the majority of papers published use those surrogate endpoints and when we focus on new technologies approved by the FDA, it is more significant. How do the physicians deal critically with this study limitation? How physicians' critical point of view would be influenced by pharmaceutical industry marketing (always attending in their office and medical conferences)? On cancer and rare diseases treatment, there is also the patients' association pressure regarding new technologies. Patients with advanced cancer are facing death and they want fast and effective solutions for their health problems. This society pressure for 'unmet medical needs' are in many times used as argument for approval of drugs with phase 2 trials or even non-randomized studies, considered still experimental. New technologies access is important, and we face an EBM Manifesto critique, the clinical trials cost and duration. Clinical trials and real-world data could be associated to generate information about efficiency and safety of new drugs. In my

opinion, the insertion of real-world data in the process of post market is important and could be mandatory to pharmaceutical industry, the public presentation of those data would be essential (as it was stated in AllTrails initiative) and if the rule was not fulfilled, it could result, for example, in a big fine or the drug registry suspension. Currently, when a new drug enters in the Market, patients and sometimes also physicians don't know the risks and uncertainty regarding those new treatment choices, in many cases this drug haven't yet confirmatory data or they are based on non-randomized studies. Patients should be clearly informed, I suggest patient education about that uncertainty of using some new medicines and also training for physicians on shared decision making. Transparency is the key for good healthcare practices based in evidence.

24 QUALITY APPRAISAL OF SYSTEMATIC REVIEWS OF HIV TREATMENT ADHERENCE AND GUIDANCE TO REDUCE RESEARCH WASTE

[1,2]Joanita Lake, [1]Kendra Lawrence, [1]Elena Martinez Alonso, [1]Valerie Gonzales, [1,2]Joanne LaFleur. *[1]University of Utah, Salt Lake City, USA; [2]Salt Lake City Veterans Affairs Health Care Center, Salt Lake City, USA*

10.1136/bmjebm-2019-EBMLive.105

Objectives Guideline developers and healthcare decision makers rely on high-quality evidence to make sound evidence-based decisions. The quality appraisal step is critical to ensuring a balanced representation of the evidence. The overall results of systematic reviews (SRs) should not be accepted as evidence-based if this step was performed inadequately. Impact factor is incorrectly being used as an indicator of the quality of papers. The purpose of this study is to systematically evaluate the quality of SRs that report the association between HIV adherence levels and specific outcomes, to determine the impact factor or reputation of the publication journal, and to provide guidance for reducing research waste.
Method A literature search was conducted in April 2018 in Ovid MEDLINE, EMBASE, CINAHL, PubMed Central, the Cochrane Library, Science Citation Index, Web of Science, ScIELO Citation Index, and Ovid Emcare. Records were screened in Covidence by at least 2 reviewers using pre-specified eligibility criteria and definitions. Methodological quality of the reviews was assessed independently by 2 reviewers using the AMSTAR 2 tool; additional information about the literature searches and conflicts of interest was extracted. The quality assessment was qualitatively compared to the impact factor of the journals in which the papers were published.
Results Our literature search identified 1141 unique records. Ultimately, 9 articles met our inclusion criteria. The overall confidence in the results of 78% of the SRs was critically low (1 critical flaw with or without non-critical weaknesses). Frequent problems identified were lack of protocols, incomplete literature searches, study selection and/or data extraction not done in duplicate, lack of formal quality appraisal tools, inadequate consideration of the effect of risk of bias of individual studies on results, and missing key information on populations, interventions, comparisons, outcomes, study designs included or rationale for studies included, funding sources, and conflicts. Impact factor or the reputation of a journal is not an indication of the overall quality of these SRs.
Conclusions This research emphasizes the importance of using quality appraisal tools and reporting guidance. The majority of SRs do not meet quality standards despite the availability of tools and guidance. The number of published SRs is increasing. This does not necessarily translate to more precise answers based on high-quality evaluations of the evidence for the ultimate goal of improving healthcare decision-making and patient care. Low-quality evidence syntheses are a huge burden on everyone involved and may cause harm. All parties involved in healthcare decisions should require critical appraisal of evidence regardless of the reputation or impact factor of an author, organization, or journal, and be prepared to perform such an evaluation prior to using, applying, or distributing SRs. The problem can be corrected if we work together to find ways in which this can be done and continue to develop innovative methods and tools to streamline the SR process without compromising quality.

25 IS EPISTEMONIKOS THE ANSWER TO KEEPING UP WITH SYSTEMATIC REVIEWS?

[1,2]Joanita Lake, [1]Valerie Gonzales, [3]Mike Clarke. *[1]University of Utah Department of Pharmacotherapy, Salt Lake City, USA; [2]Salt Lake City Veterans Affairs Health Care Center, Salt Lake City, USA; [3]Northern Ireland Methodology Hub, School of Medicine, Dentistry and Biomedical Sciences, Queen's University, Belfast, UK*

10.1136/bmjebm-2019-EBMLive.106

Objectives Guideline developers, healthcare decision makers, and researchers need to identify reliable systematic reviews (SRs) to inform evidence-based medicine (EBM) and underpin guidelines. At the start of this decade, Bastian et al. highlighted the challenge of keeping up with new publications, when 11 reviews were being published daily and this is coupled with difficulties in finding SRs through time-intensive, traditional literature searches. We sought to estimate the current publication rate of SRs and to examine Epistemonikos as a method for identifying SRs by considering transparency of contributors and SR identification methods, researcher awareness and confidence, and its value as a means of finding SRs.
Methods We reviewed the Epistemonikos website and searched Pubmed, Embase, and the Cochrane Library for 'Epistemonikos' to examine awareness. We compared basic search strategies in Epistemonikos with the comprehensive search strategy from an overview of SRs and screened records solely identified in Epistemonikos to determine their eligibility for the overview. We estimated the number of SRs published annually between 1990 and 2018 through various searches, including Epistemonikos.
Results We noted no major concerns for potential conflicts of interest in the compilation of Epistemonikos, but a fuller process description for identifying SRs would be helpful. The word Epistemonikos appeared in 226 abstracts in Pubmed or Embase, and in the full text of 24 of 7960 (0.3%) full Cochrane reviews. Our basic search in Epistemonikos (including treatment, adherence, and outcome terms) identified 67% of the records retrieved by the full search for the overview. A broader search without outcome terms identified 78%, and a very broad search using only treatment terms identified 89%. One key SR (published in 2011) was not indexed in Epistemonikos at the time of our search (March 2019); but was present by April 2019. None of the other records identified solely by the Epistemonikos search were eligible for the overview. The annual number of SRs suggests three distinct periods: a slow rise to the year 2000, a gradual increase in 2000-