

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 Llewellym. *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.<sup>1</sup>

**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. Llewellyn 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