Editorial

The role of statistical significance in health risk assessment and in the decision-making process

MARCO PANTEGHINI[1], MOHAMMAD ALÌ MANSOURNIA[2]

[1] Department of Obstetrics and Pediatrics, Azienda Unità Sanitaria Locale, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Reggio Emilia, Italy; [2] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran

Dear Editor,

We appreciated the article by Filippini and Vinceti [1]. It highlights one of the most important recent changes in scientific methodology: the role of "statistical significance" "in the establishment of causal relations in science, including toxicology and biomedical sciences […] psychological and economic research" [1]. The authors claim that "a complete demise of this simplistic approach appears fully justified in both public law and health risk assessment in favour of a more challenging but methodologically correct method based on the comprehensive assessment of the strengths and limitations of all the available evidence" and they stress that "statistical significance testing has been the pillar and the tenet of risk assessment and biostatistics for decades" [1].

When studying causal relations between two variables in a population, for instance between a possible cancerogenic agent and cancer, the best method is to study the entire population. This allows us to calculate the relative risk (RR) related to the exposure, comparing the absolute risk of cancer in the population exposed to the possible cancerogenic agent and the absolute risk of cancer in the population not exposed to this agent.

When the population is large, it is not always possible to study the entire population, so we have to calculate the relative risk of cancer by comparing the absolute risk of cancer in a population sample exposed to the possible cancerogenic agent and in a population sample not exposed to this agent. In so doing, we obtain a relative risk (RR), for instance RR = 3, meaning that in the sample studied, the frequency of cancer is three times more frequent in the sample exposed to the agent.

Is this relative risk representative of the population or is this result related to population sampling? In other words, is this result due to the sampling variation around the true value or could it be representative of the true value in the population?

"Statistical significance" is a way to try to answer this question.

It starts from the hypothesis (the so-called "null hypothesis") that in the population there is no relation between the two variables considered (the possible cancerogenic agent and the cancer, *i.e.* that the RR = 1 and that the possible cancerogenic agent is not cancerogenic). If this hypothesis is true, what is the probability (p-value) of obtaining a result, such as the one observed in the population sample (RR:3) or more distant from the above hypothesis?

If this probability is low, traditionally < 0.05 (< 5%), the hypothesis that the real value of relative risk in the population is 1, that is, that the result observed (RR = 3) is due to sampling variability, is not considered a valid explanation, the "null hypothesis" is refused, and the result (RR = 3) is considered "statistically significant". Conversely, if the p-value is > 0.05, the hypothesis that the real value in the population is 1, i.e. that the sampling variability is a valid explanation of the observed result, it is "not refused" and the observed result is considered "not statistically significant".

It has been proposed to replace the term "significance" with "compatibility" and "significance" test with "hypothesis" test, in order to emphasise that hypothesis tests evaluate the compatibility between a hypothesis or a model and observed data [2, 3]. Specifically, it is important to report that: if the test's hypothesis is true, the probability to obtain results, that are equally or less compatible with (or equally or more distant from) the above hypothesis than those observed, is p.

We would like to make two observations about the proposal of the authors and many scientists to abandon the use of "statistical significance" [4].

First, "statistical significance", as defined in relation to null hypothesis statistical testing, has been proposed to prevent false positive results, that is, considering that there is an association between two variables studied in a population sample, when this association is absent in the population and the result observed in the population sample is due to sampling fluctuation around the true value. Therefore, we have to be aware that completely abandoning "statistical significance" may lead to an increase in false positive results [5] and to a decrease in false negative results. Anyway, the best method to prevent false positive and false negative results is to increase the sample size. Increasing the sample size decreases the probability of refusing a result when that result is true and decreases the probability of not refusing a result when it is false.

"How feasible is it to abandon statistical significance?" [6]. It is feasible by simply reporting the results of clinical trials as relative and absolute risk

between the treated group and the untreated group, without doing the hypothesis test. These results, if derived from well-designed trials (e.g. prospective double-blind randomized trial between two homogeneous groups) "count as evidence" [5] of association between two variables in a population. The only way to determine whether a result is due to sample variability is to repeat the trial and to increase the sample size.

Secondly, the American Statistical Association (ASA) released a statement on statistical significance in 2016 which states that statistical significance "is not equivalent to scientific, human, or economic significance" [7], meaning that a scientific result may be important or "significant", even if it is not "statistically significant". We want to stress the importance of this concept, because, in patient-centred medicine, the patient must be put at the centre of the decision-making process. The patient must decide if a scientific result is significant for him/her, considering his/her needs and values. Similarly, in a decision-making process, we have to put the decision-maker (e.g., a public health agency) at the centre of the decision-making process. The use of "statistical significance", as a means to decide if a result is scientifically important or not, does not put the decision-maker at the centre of the decision-making process, because the test discriminates between significant and non-significant results before the evaluation of the decision-maker. This implies that reporting the results of a scientific study, without classifying them as "statistically significant or not statistically significant", may put the decision-maker at the centre of the decision-making process, promoting patient-centred medicine [8].

The signatories of the petition for retiring statistical significance [4] were asked about their intentions: specifically, about how likely they are to use the concept of "statistical significance" in their future publications [5]. Forty-two percent declared they are neutral or likely to use it in future publications, and 58% declared that they expect to never to use it again or they said it would be unlikely they would use it in future publications. The use of hypothesis testing in publications depends on several factors, first of all on the editor's willingness to accept both studies in which it is used and studies in which the hypothesis test is not used. We agree that it is important to "promote more education among researchers and users of scientific evidence" [6] about statistical significance. At the same time, we agree with ASA, that it is important to "open a fresh discussion and draw renewed and vigorous attention to changing the practice of science" [7].

Funding

Conflict of interest statement

The authors declare no conflict of interest.

Authors' contribution

MP conceived and drafted the manuscript. MAM revised the manuscript, especially the statistical section. All authors have read and approved the latest version of the paper for publication.

References

[1] Filippini T, Vinceti SR. The role of statistical significance testing in public law and health risk assessment. J Prev Med Hyg 2022;63:E161-E165. https://doi.org/10.15167/2421-4248/jpmh2022.63.1.2394

[2] Greenland S, Mansournia MA, Joffe M. To curb research misreporting, replace significance and confidence by compatibility: A Preventive Medicine golden jubilee article. Prev Med 2022;164:107127. https://doi.org/10.1016/j.ypmed.2022.107127

[3] Mansournia MA, Nazemipour M, Etminan M. P-value, compatibility, and S-value. Global Epidemiology 2022;4:100085. https://doi.org/10.1016/j.gloepi.2022.100085

[4] Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature 2019;567:305-307. https://doi.org/10.1038/d41586-019-00857-9

[5] Mayo DG, Hand D. Statistical significance and its critics: practicing damaging science, or damaging scientific practice? Synthese 2022;200:220. https://doi.org/10.1007/s11229-022-03692-0

[6] Diaz-Quijano FA, Calixto FM, da Silva JMN. How feasible is it to abandon statistical significance? A reflection based on a short survey. BMC Med Res Methodol 2020;20:140. https://doi.org/10.1186/s12874-020-01030-x

[7] Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. Am Stat 2016;70:129-133. https://doi.org/10.1080/00031305.2016.1154108

[8] Panteghini M, Tonni G, Mansournia MA. To Move Beyond Statistical Significance and Consider Clinical Relevance. Clinical Oncology 2022;34:e315. https://doi.org/10.1016/j.clon.2021.06.008