

EDUTORIAL

The Problem of Multiple Hypothesis Testing and the Assumption of Causality

“Scientia potentia est.” (Sir Francis Bacon, London 1561–1626). We all agree that information is power because data give us the knowledge to make correct decisions. Although we currently have access to a great amount of information, our problem is to select all the useful data, separate it from the invalid and non-significant, and interpret it in the proper way.

Databases are one of the possible sources of information that can cause problems, which on the other hand, when used in a correct way are essential in our research. The problem arises when we use databases that have been designed for other purposes (often healthcare management or clinical practice) including a great amount of information that we try to recycle afterwards. If a pre-defined objective is not established before using a database, we cannot be assured of the quality control of data or of the completeness of the information, that is to say all outcomes we need are included and well defined and the sample size is large enough. In addition, even if this information is complete and of high quality, the way we analyse it could direct our conclusions to the wrong ones. For instance, the use of multiple hypothesis testing may lead to the detection of spurious relationships and lead to incorrect conclusions. Remember that a type I error of 5% ($p < .05$) as a significant result is the probability of asserting a statistical difference when the null hypothesis is really true (no difference). Therefore, when we randomly test a lot of outcomes trying to find a significant “ p ” value, we must take into account that there is a chance of defining any relationship as significant, without it being true, also that this chance would depend on the number of outcomes tested.

To avoid all those kinds of problems, it is important to design the study before the storage of data and even better, with a prospective and public registration of trial protocols, taking into account that some journals require this prior registration to publish any research. This will help us to be clear about some concepts, like the questions we need to answer, the main and secondary outcomes, controlling possible confounders, and pre-defining the sample size. This is the only way to assure a correct and powerful interpretation of information from a database.

Nevertheless, if we want to analyse a database by multiple hypothesis testing, trying to discover some significant result, we can apply some methodological tips, for instance the Bonferroni correction¹ or the false discovery rates controlling procedure of Benjamini and Hochberg.² These statistical methods, and many others, try to help researchers to avoid Type I errors in a more or less conservative way, meaning with more or less false negatives (Type II errors). We must keep in mind that the quality of the evidence from this type of analysis is low and it must be confirmed with specific well designed research.

In addition, we must address the key point of causality, because a significant correlation relationship does not imply causation,

which cannot be stated with a specific statistical test. At this point, we must refer to the Bradford Hill criteria whose compliance gives information about a possible causal relationship:³ strength of association (the stronger the association, the more likely there is a causal relationship); consistency (the relationship is observed repeatedly); specificity (a given factor is related specifically to a unique effect; this criterion is not mandatory because usually different causes can produce different effects); temporality (the cause must precede the effect; this can only be confirmed with a prospective design); biological gradient (increases in the exposure increase the magnitude of the effect); plausibility (the association must be reasonably explained); coherence (the effect must be coherent according to current knowledge); design of the study (an observational cross sectional historical study is not to the same as a randomised controlled trial); analogy (similar outcomes have already produced similar effects). The accomplishment of all nine criteria is almost impossible due to different problems that could be related to the design of the study, the characteristics of the outcomes, current knowledge, or any other criteria related to the specific hypothesis. However, a global evaluation of these criteria should offer us information to suspect or doubt a causality relationship.

In conclusion, we must be cautious when analysing multiple hypotheses squeezed from a database, by trying to differentiate between true and spurious relationships. The use of some techniques can help us make wiser decisions. Multiple hypothesis testing can be a good way to generate scientific questions, however these must be confirmed later with specific studies. The use of prospectively registered databases focused on pre-defined objectives is the best way to begin our research. Finally, conclusions about causality cannot be stated based on a single statistical test and some key criteria should be considered before asserting the cause and effect relationship.

REFERENCES

- 1 Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995;310:170.
- 2 Green GH1, Diggle PJ. On the operational characteristics of the Benjamini and Hochberg false discovery rate procedure. *Stat Appl Genet Mol Biol* 2007;6:27.
- 3 Höfler M. The Bradford Hill considerations on causality: a counterfactual perspective. *Emerg Themes Epidemiol* 2005;2:11.

Sergi Bellmunt-Montoya*

Department of Vascular Surgery/Endovascular Surgery and Angiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Spain

*Hospital Universitari Vall d'Hebron, Passeig de la Vall d'Hebron, 119–129, 08035, Barcelona, Spain.

Email-address: sbellmunt@vhebron.net