Letter to the Editor

What all surgeons should know about sub-group analysis

Chadli Dziri[1,2]^, Selman Uranues[3], Min-Hua Zheng[4], Abe Fingerhut[3,4]

[1]Tunis Faculty of Medicine, The University of Tunis El Manar, Tunis, Tunisia; [2]Honoris Medical Simulation Center, Tunis, Tunisia; [3]Section for Surgical Research, Department of Surgery, Medical University of Graz, Graz, Austria; [4]Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Minimally Invasive Surgery Center, Shanghai, China

Correspondence to: Abe Fingerhut, MD, FACS (hon), FRCPS (g), FRCS (Ed) (hon), FASCRS (hon). Section for Surgical Research, Department of Surgery, Medical University of Graz, Neue Stiftingtalstraße 6, 8010 Graz, Austria; Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Minimally Invasive Surgery Center, No. 197, Ruijin 2nd Road, Huangpu District, Shanghai 200023, China. Email: abefingerhut@aol.com.

Received: 12 January 2024; Accepted: 08 March 2024; Published online: 01 April 2024.
doi: 10.21037/ales-24-3
View this article at: https://dx.doi.org/10.21037/ales-24-3

Subgroup analysis is widely practiced and reported in clinical studies (1-3). Assmann *et al*., in their 2000 analysis of such trials published up to 1997, found that 38 of 50 emphasized simple outcome comparisons between treatments, unadjusted for baseline covariates (1). Of note, however, *post-hoc* subgroup analyses (after the study has been completed) are notoriously disposed to methodological faults such as false positive results due to multiple testing, inadequate sample size and therefore low power, and, most frequently, inappropriate statistical interpretation. Hence, the impossibility to come to any conclusions, and in particular, to infer any causality.

Reliance on traditional null hypothesis analysis to obtain subgroup P values is misleading. Even if the difference in the overall result is not statistically significant, almost inevitably, analysis of some subgroups will eventually show a statistically significant difference depending on chance.

However, subgroup analysis can be of use to the scientific community in that results might lead to a new hypothesis. One way to try to use this information in a statistically pertinent and sensible manner is to use interaction analysis; interaction analysis, however, is not widely practiced among scientists, and in particular, surgeons.

What is interaction? In statistics, an interaction describes a situation where two or more (independent) variables interact to affect a third (dependent) variable in a non-additive manner.

The methodology is specific. In order to try to explain this simply to the surgical readership of this journal, we will provide a scenario and step by step analysis of how to proceed and what may be gleaned from such analysis.

Scenario

In a recently published paper (4), The efficacy, safety and tolerability of Hemopatch® applied to the pancreatic stump after distal pancreatectomy (DP) were evaluated in the prevention of clinically relevant postoperative pancreatic fistulas (POPF) (grades B and C according to Bassi *et al.* (5,6). Of 631 eligible patients, 360 were randomized; once the study was closed, 315 patient records were analyzed [155 in the standard closure (stapled or hand-sewn) group; 160 in the Hemopatch® group (standard closure + Hemopatch®)]. While the difference in efficacy of the primary endpoint (B/C POPF) was not statistically significant (P=0.120; 24.5% and 16.3%, in the two groups with or without Hemopatch®, respectively), 17 out of 65 and seven out of 70 patients sustained a B/C POPF in the standard and Hemopatch® groups [26.2% *vs*. 10.0%, respectively (P=0.014)], in the 135 patients where hand-sewn closure of the pancreatic stump was performed.

^ ORCID: 0000-0003-2779-0930.

Statistical analysis

The interaction methodology is well described (2,7). Specifically, in relation to our example, we want to determine whether the two dichotomic independent variables i.e., outcome "Hemopatch® or basic treatment" [Treatment (T)] and "hand-sewn closure or stapled closure" [Baseline factor (B)] interact with the onset of B/C POPF, the dependent variable. This involves logistic regression according to the following formula (2):

Interaction:

$$\text{Product}(T)\times(B)=\left[\text{Hemopatch}^{®}\ vs.\ \text{Basic}(T)\right]\times\left[\text{Handsewn}\ vs.\ \text{stapler}(B)\right] \quad [1]$$

To begin with, we can calculate the Relative Excess Risk due to Interaction or "RERI" (2). The RERI is the difference between the combined effect of the treatment and one of the demographic factors compared with their effects if they were considered separately. RERI ranges from -∞ to +∞ and may indicate super-additive (RERI >0) or sub-additive (RERI <0) interaction effects.

The formulas used are RERI = RR (T+, B+) – RR (T+, B–) – RR (T–, B+) + 1 (stratification) when the absolute risks are available, and RERI=$e^{\beta1+\beta2+\beta3}$ – $e^{\beta1\ (T)}$ – $e^{\beta2\ (B)}$ +1 (interaction modeling) when they are not (2).

In the example provided above, we found an interaction between T and B (P=0.034) with the product (TXB) β=1.288.

In our example, the absolute risks were available, so we were able to calculate the RERI as follows.

RR_{T+B+} is the relative risk when both risk factors (hand sewn and Hemopatch®) are present (=0.47), RR_{T+B-} is the relative risk when one risk factor (T) is present (=1.24), and RR_{T-B+} is the relative risk when the other risk factor (B) is present (=1).

RERI of POPF in handsewn stump closure [0.47–1–1.24+1=–0.77 (P=0.0204)]. This means that the relative risk of having POPF in patients who had Hemopatch® added to a handsewn stump closure was noted to be 0.77 less than if there were no interaction between Treatment (T) and Closure technique (B).

Another metric that has some interest is the attributable proportion (AP) i.e., the proportion of patients with both exposures (hand-sewn closure and Hemopatch®) who did not sustain the endpoint attributable to the interaction (in this example POPF) according to the formula AP = RERI/RR_{T+B+} (7).

Comments

Brankovic *et al.* have made three postulates relative to interaction analysis (2):

(I) The statistical interaction between the treatment and a baseline factor can be interpreted as a change in the effect measure or as a causal interaction.

(II) The interaction can only be interpreted as causal if both the treatment and the baseline factor directly affect the outcome (8,9).

(III) In randomized controlled clinical trials (RCTs), researchers can assert that the treatment directly affects the outcome, even in subgroups of the baseline factor, due to the randomization of the treatment (10). However, it is not possible to assert that the baseline factor itself is responsible for subgroup effects if the confounding factors of the baseline factor on outcome have not been controlled.

Interaction tests have been criticized for their lack of sensitivity and specificity to detect or exclude truly differential efficacy in the study population. A dichotomous decision on the presence or absence of an interaction on the basis of a statistical test cannot be considered a reliable basis for decision-making (11).

According to Sun *et al.* (12), subgroup effects should not be considered as all black or all white decision-making policies. These authors believe that the "truth" lies somewhere between postulates that are most certainly false and those that are most certainly true.

In conclusion, we would like to make a plea to see more interaction analysis being performed in surgical journals. The mathematics are simple. Interaction analysis provides more credibility than the usual "subgroup" analysis we see in many publications, and paves the pathway for more sound methodological techniques and further randomized studies, in (sub) groups of interest, originally included in larger randomized studies. In agreement with Alrawabdeh *et al.* (13), we believe (I) the increasing use of subgroup analyses, not only in oncology clinical trials but also in other methodologically sound studies, and (II) improved definitions of subgroups that can benefit from a specific treatment, underscore the necessity to use statistically robust approaches to subgroup analysis and to communicate findings with precaution so that authors, readers, journal editors, and reviewers can exercise the warranted but often misunderstood attention required to conduct and interpret results of such analyses.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was a standard submission to the journal. The article has undergone external peer review.

Peer Review File: Available at https://ales.amegroups.com/article/view/10.21037/ales-24-3/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://ales.amegroups.com/article/view/10.21037/ales-24-3/coif). C.D. is an editorial board member of the *American Journal of Surgery*. S.U. reports consulting fees from Baxter Company and he serves as an editorial board member of *Annals of Laparoscopic and Endoscopic Surgery* from September 2022 to August 2024. M.H.Z. serves as the Editor-in-Chief of *Annals of Laparoscopic and Endoscopic Surgery* from April 2016 to April 2026. A.F. serves as the Editor-in-Chief of *Annals of Laparoscopic and Endoscopic Surgery* from April 2016 to April 2026 and reports consulting fees from Baxter Company. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355:1064-9.
2. Brankovic M, Kardys I, Steyerberg EW, et al. Understanding of interaction (subgroup) analysis in clinical trials. Eur J Clin Invest 2019;49:e13145.
3. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189-94.
4. Uranues S, Fingerhut A, Belyaev O, et al. Clinical Impact of Stump Closure Reinforced With Hemopatch on the Prevention of Clinically Relevant Pancreatic Fistula After Distal Pancreatectomy: A Multicenter Randomized Trial. Ann Surg Open 2021;2:e033.
5. Bassi C, Dervenis C, Butturini G, et al. Postoperative pancreatic fistula: an international study group (ISGPF) definition. Surgery 2005;138:8-13.
6. Bassi C, Marchegiani G, Dervenis C, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. Surgery 2017;161:584-91.
7. Knol MJ, van der Tweel I, Grobbee DE, et al. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol 2007;36:1111-8.
8. VanderWeele TJ. On the distinction between interaction and effect modification. Epidemiology 2009;20:863-71.
9. Vander Weele TJ. Confounding and effect modification: distribution and measure. Epidemiol Methods 2012;1:55-82.
10. Lachin JM, Bautista OM. Stratified-adjusted versus unstratified assessment of sample size and power for analyses of proportions. Cancer Treat Res 1995;75:203-23.
11. Hemmings R, Koch A. Commentary on: Subgroup analysis and interpretation for phase 3 confirmatory trials: White Paper of the EFSPI/PSI working group on subgroup analysis by Dane, Spencer, Rosenkranz, Lipkovich, and Parke. Pharm Stat 2019;18:140-4.
12. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. JAMA 2014;311:405-11.
13. Alrawabdeh J, Alzu'bi M, Alzyoud M, et al. Characteristics of post hoc subgroup analyses of oncology clinical trials: a systematic review. JNCI Cancer Spectr 2023;7:pkad100.

doi: 10.21037/ales-24-3
Cite this article as: Dziri C, Uranues S, Zheng MH, Fingerhut A. What all surgeons should know about sub-group analysis. Ann Laparosc Endosc Surg 2024;9:20.