Second, the particular aim of this research was to identify optimal fistula mitigation strategies for the ''high risk'' patient, as identified by FRS scores ranging from 7 to 10. We were particularly interested in providing strategic insights to improve outcomes for this most vulnerable patient subset since the clinically-relevant fistula (CR-POPF) rate approaches or exceeds 30% in this scenario. In this context, the exclusion of negligible, low and intermediate risk patients does not represent a bias, but rather is a reflection of this study's focal aim and design. In previous works utilizing all FRS patients from the Pancreas Fistula Study Group dataset, we have performed risk-stratified analyses of internal stents,[2] external stents,[3] octreotide prophylaxis,[4] intraperitoneal drainage,[5] and pancreatogastrostomy reconstruction.[6] Despite these analyses of individual mitigation strategies, we cannot currently determine how these individual strategies interact, nor define an optimal fistula mitigation bundle for the lower FRS risk zones, particularly due to the low encountered incidence (0–4%) of the outcome of interest—CR-POPF. However, a collective summary, FRS zone-by-zone, of the results of these studies has been collated in a tabular form in another publication.[7]

A third question regards the CR-POPF associated mortality in this study population. The overall 90-day mortality rate for these high FRS risk patients was 4.6%, of which nearly two-thirds suffered a CR-POPF (see Table 3 in manuscript).[8] The pancreatic fistula, or its immediate complications, was considered the cause of death for 2.3% of the 522 patients, and thus half of all postoperative mortalities. We suspect these higher numbers reflect the greater fistula rate, and inherent glandular risk, encountered in this particular population.

Finally, it is possible that this so-called technique of ''external venting stents'' as described by Oguro et al[9] may be particularly beneficial for this scenario, although—as you recently note in your extensive review of CR-POPF-related mortality after PD[10]—this strategy was rarely used (0.9% of PDs) and without improvement in mortality ($P = 0.999$). We agree with your enthusiasm to improve the delivery of surgical care by targeting a reduction in CR-POPF-related mortality and are intrigued by the data you provide from Oguro et al. However, we would like to clarify for the general readership that this strategy relies on a 2-stage operative approach. In the first stage, the jejunum is apparently approximated to the cut-edge of the pancreas using just a posterior attachment, without formation of an actual enterotomy. The pancreatic duct is then cannulated with an externalized stent, creating, in effect, a controlled pancreatico-atmospheric fistula (thus, all patients actually have a CR-POPF). The patient is then subjected to a second operation, 3 months or later, to ''take down'' the fistula and convert it to a formal pancreatico-jejunostomy at that point. In our opinion, this approach seems excessive, inefficient, and impractical. It remains to be determined if these results can be confirmed at other institutions or, furthermore, in a randomized controlled trial.

The authors report no conflicts of interest.

Brett L. Ecker, MD
Laura Maggino, MD
Charles M. Vollmer Jr, MD
University of Pennsylvania
Philadelphia, PA
Charles.Vollmer@uphs.upenn.edu

REFERENCES

1. Pedrazzoli S, Sperti C. Prevention of clinically-relevant postoperative pancreatic fistula after pancreticoduodenectomy. *Ann Surg*. 2018.
2. Sachs TE, Pratt WB, Kent TS, et al. The pancreaticojejunal anastomotic stent: friend or foe? *Surgery*. 2013;153:651–662.
3. McMillan MT, Ecker BL, Behrman SW, et al. Externalized stents for pancreatoduodenectomy provide value only in high-risk scenarios. *J Gastrointest Surg*. 2016;20:2052–2062.
4. McMillan MT, Christein JD, Callery MP, et al. Prophylactic octreotide for pancreatoduodenectomy: more harm than good? *HPB (Oxford)*. 2014;16:954–962.
5. McMillan MT, Malleo G, Bassi C, et al. Multicenter, prospective trial of selective drain management for pancreatoduodenectomy using risk stratification. *Ann Surg*. 2017;265:1209–1218.
6. Ecker BL, McMillan MT, Maggino L, et al. Pancreatogastrostomy vs. pancreatojejunostomy: a risk-stratified analysis of 5316 pancreatoduodenectomies. *J Gastrointest Surg*. 2018;22:68–76.
7. Puri P, Vollmer CM. Strategies for prevention and treatment of pancreatic fistula. In: Shen P, Rocha F, eds. *Optimizing Outcomes for Liver and Pancreas Surgery*. 1st ed, New York: Springer; 2017.
8. Ecker BL, McMillan MT, Asbun HJ, et al. Characterization and optimal management of high-risk pancreatic anastomoses during pancreatoduodenectomy. *Ann Surg*. 2018;267:608–616.
9. Oguro S, Yoshimoto J, Imamura H, et al. Three hundred and sixty-eight consecutive pancreaticoduodenectomies with zero mortality. *J Hepatobiliary Pancreat Sci*. 2017;24:226–234.
10. Pedrazzoli S. Pancreatoduodenectomy (PD) and postoperative pancreatic fistula (POPF): a systematic review and analysis of the POPF-related mortality rate in 60,739 patients retrieved from the English literature published between 1990 and 2015. *Medicine (Baltimore)*. 2017;96:e6858.

Don't Calculate Post-hoc Power Using Observed Estimate of Effect Size

An article recently published in the *Annals of Surgery* states: ''as 80% power is difficult to achieve in surgical studies, we argue that the CONSORT and STROBE guidelines should be modified to include the disclosure of power—even if <80%—with the given sample size and effect size observed in that study''.[1]

This would be a bad idea. The problem is that the (estimated) effect size observed in a study is noisy, especially so in the sorts of studies discussed by the authors. Using estimated effect size can give a terrible estimate of power, and in many cases can lead to drastic overestimates of power (thus, extreme overconfidence of the sort that is rightly deplored in the full article[1]), with the problem becoming even worse for studies that happen to achieve statistical significance.

The problem is well known in the statistical and medical literatures; see, for example, references[2,3]. For some discussion of the systemic consequences of biased power calculations based on noisy estimates of effect size, see,[4] and for an alternative approach to design and power analysis, see.[5]

That said, I agree with much of the content of.[1] I agree that the routine assumption of 80% power is a mistake, and that requirements of 80% power encourage researchers to exaggerate effect sizes in their experimental designs, to cheat in their analyses to attain the statistical significance that they was supposedly so nearly being assured.[6] More generally, demands for near-certainty, along with the availability of statistical analysis tools that can yield statistical significance even in the absence of real effects, have led to replication crisis and general corruption in many areas of science,[7] a problem which I believe is structural and persists even in the presence of honest intentions of many or most participants in the process.[8]

I appreciate the concerns of[1] and I agree with their goals and general recommendations, including their conclusion that ''we need to begin to convey the uncertainty associated with our studies so that patients and providers can be empowered to make appropriate decisions.'' There is a just a problem with their recommendation to calculate power using observed effect sizes.

I thank Aleksi Reito for bringing this article to my attention.

Andrew Gelman, PhD
Department of Statistics, Columbia University, New York, NY.
gelman@stat.columbia.edu

REFERENCES

1. Bababekov YJ, Stapleton SM, Mueller JL, et al. A proposal to mitigate the consequences of type 2 error in surgical science. *Ann Surg*. 2018;267:621–622.
2. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*. 1994;121:200–206.
3. Senn SJ. Power is indeed irrelevant in interpreting completed studies. *BMJ*. 2002;325. Article 1304.
4. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14:1–12.
5. Gelman A, Carlin JB. Beyond power calculations: assessing Type S (sign) and Type M (magnitude) errors. *Perspect Psychol Sci*. 2014;9:641–651.
6. Gelman, A. The ''80% power'' lie. Statistical Modeling, Causal Inference, and Social Science blog, December 4, 2017. http://andrewgelman.com/2017/12/04/80-power-lie/. Accessed December 4, 2017.
7. Ioannidis J. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol*. 2016;73:82–86.
8. Gelman A. Honesty and transparency are not enough. *Chance*. 2017;30:37–39.

Letter to Editor: A Proposal to Mitigate the Consequences of Type 2 Error in Surgical Science

We read the article by Bababekov et al[1] published under Surgical Perspectives with great interest. There is no doubt that inadequate power to detect significant findings, or type 2 error, is a concern in surgical science, especially when the interpretation of the results is made inadequately. This belongs to fundamental reasons behind the so called ''replication crisis.''[2,3] As the authors write, absence of evidence is not evidence of absence. The authors conclude that CONSORT and STROBE guidelines should ''include the disclosure of power with the given sample size and effect size observed in that study.''

When stipulating power using the study sample size and observed effect size, one is calculating the post-hoc power. The authors make a serious and well-known error in their article when suggesting the use of post-hoc power to mitigate type 2 error. Statistical power (a priori power usually) is the probability of finding a significant effect, if there is a true difference to be found. There is a great difference between post-hoc power and a priori power which estimates the power of the study to detect a predefined effect size of interest using the study sample size. We suggest readers to refer to article by Hoenig and Heisey stating that ''because of the one-to-one relationship between *P* values and observed power, non-significant *P*-values always correspond to low powers.''[4] Moreover O'Keefe writes ''Second, this [post-hoc] power value provides an answer to question that doesn't much matter, namely: 'What chance was there of producing a statistically significant result, assuming that the population effect is exactly equal to the observed sample effect?'''[5]

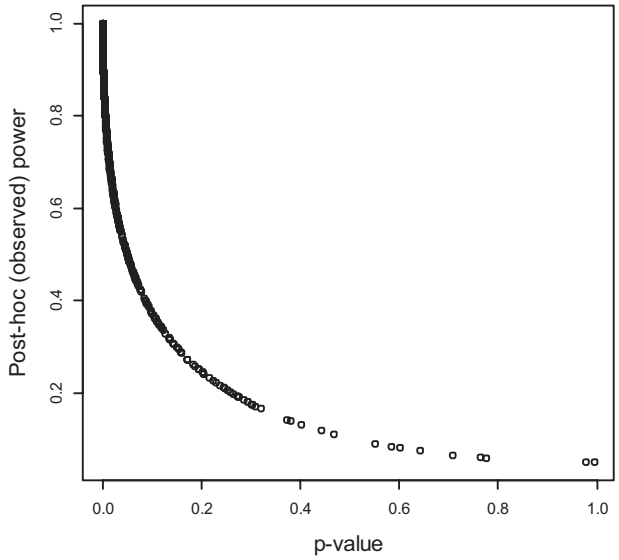

FIGURE 1. Correlation between observed, that is post-hoc power and *P* value for 500 simulated *t* tests with a power of 79.5% to detect an effect size of 0.5. *P* values higher than 0.05 are more sparse than those <0.05.

As the same information is required to calculate post-hoc power and *P* value, the natural outcome is that post-hoc power is inherently associated with the calculated *P* value. Post-hoc power of 50% always provides *P* value of 0.05 (Fig. 1). This direct correlation has been reported in statistical literature a number of times.[4–6] Post-hoc power is nothing more than a report of *P* value in a different way and therefore provides no answer to type 2 error. Nonsignificant studies always have low observed power.

The authors use the search of keys in a house as a common sense example of type 2 error. As the authors state, ''reporting the power of a study is analogous to disclosing the extent of a search effort.'' This is true and in fact is the case in a priori power analysis, that is the probability of finding the keys in prespecified areas of the house. Using study sample size and observed effect size, that is the post hoc power, is only a 1-time retrospective look at the data, just like the *P* value.

Instead of reporting the post-hoc power, the scientist could explain how likely it was to observe a significant effect based on the sample size and assumed effect size of interest, similar to reporting the a priori power calculations, as suggested in CONSORT and STROBE guidelines.

The authors report no conflicts of interest.

Olli Helminen, MD, PhD
Department of Surgery, Central Finland Central Hospital, Jyväskylä, Finland
Cancer and Translational Medicine Research Unit, Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland
olli.helminen@ksshp.fi

Aleksi Reito, MD, PhD
Department of Surgery, Central Finland Central Hospital, Jyväskylä, Finland
Coxa Hospital for Joint Replacement, Tampere, Finland