Multiple Comparisons in Variation of Care Research

Vinay Prasad, MD, MPH[1]; Andrew Oseran, MD, MBA[2]

[1] Oregon Health & Science University
[2] Massachusetts General Hospital

ABSTRACT

Research in hospital variation is important and currently very popular. However, due to the methods employed in such studies—namely, the retrospective mining of large datasets and the use of several alternative variation groupings—some results may be spurious. In this commentary, we perform an empirical analysis of the 50 most highly cited and the 50 most recent papers focusing on variation in medical care. Across these studies, we identify at least 13 unique groupings and could find no single instance where a medical practice was found not to vary. We go on to discuss one example of variation—statin use—in more detail to elucidate the tensions that these studies often create. Together, these results suggest that multiple hypothesis testing is a concern for variation research. Finally, we outline strategies to mitigate this concern.

RÉSUMÉ

La recherche sur la variation hospitalière est importante et actuellement très populaire. Toutefois, en raison des méthodes employées dans de telles études—notamment, l'extraction rétrospective de grands ensembles de données et l'utilisation de plusieurs groupements de variation alternatifs—certains résultats peuvent être fautifs. Dans ce commentaire, nous effectuons une analyse empirique des 50 articles les plus citées et des 50 articles les plus récents se concentrant sur la variation dans les soins médicaux. Dans ces études, nous identifions au moins 13 groupements uniques, et ne pouvions trouver aucun cas où une pratique médicale ne variait pas. Nous discutons ensuite d'un exemple de variation—dans l'utilisation de statines—en plus de détails afin d'élucider les tensions que ces études suscitent souvent. Collectivement, ces résultats suggèrent que la mise à l'essai de multiples hypothèses est une préoccupation lors de la recherche sur la variation. Finalement, nous décrivons des stratégies pour atténuer cette préoccupation.

INTRODUCTION

Variation in medical care that is not explained by patient preferences or characteristics and does not result in improved health outcomes is inappropriate. This metric often serves as the subject of public policy deliberation and intervention. The body of evidence now documenting such variation is immense. A PubMed search for "variation in care" yields over 23,000 entries (July 8, 2013), and the same terms generate over 2 million results on Google Scholar (July 8, 2013). The seminal illustration of variation research, the Dartmouth Health Atlas, describes variation across 306 hospital referral regions (HRRs) in the United States [1].

Others have analyzed variation across groupings other than HRRs. For instance, Zhang et al. examined variation across 3,436 hospital service areas (HSAs) in the United States. They found that the highest spending HSAs were only loosely correlated with high spending HRRs [1]. These results suggest that targeting variation by HRR is neither a sensitive nor specific strategy to identify the source of unwarranted practice [1]. Other research efforts have examined variation based on race, physician specialty, individual physicians, individual hospitals, patient age, other patient characteristics, countries, states, cities, neighbourhoods, and more. Importantly, these studies are often performed in large data sets, such as Medicare administrative databases, and are thus likely overpowered to identify small but significant differences between groups.

Unexplained variation research is very popular and often provocative, likely thousands of researchers embark on projects studying variation. Because these studies are nearly always done retrospectively, and because such analyses can be conducted easily with modern computing, for all of the variation studies that have been published, many more analyses—possibly thousands—remain unpublished [2]. In short, research papers that find unexplained variation in medicine may represent only the tip of the iceberg, with many more studies hidden beneath the surface. When many observational analyses are conducted with diverse definitions, datasets, and hypotheses, the possibility exists that a large portion of the literature is spurious—reflecting what researchers and editors believe is plausible, rather than the truth about patterns of inappropriate care [2,3].

To illustrate this point, we set out to characterize the number of ways in which variation is analyzed in a representative sample of variation of care research. Specifically, using Thomson Reuters Web of Science, we identified the 50 most highly cited and the 50 most recent studies of variation in a medical practice in the biomedical literature. Our search was conducted on January 31, 2013, using "Variation" in the title field, and restricting results to

Keywords: Variation; Healthcare Quality; Clinical Outcomes; Comparative Effectiveness
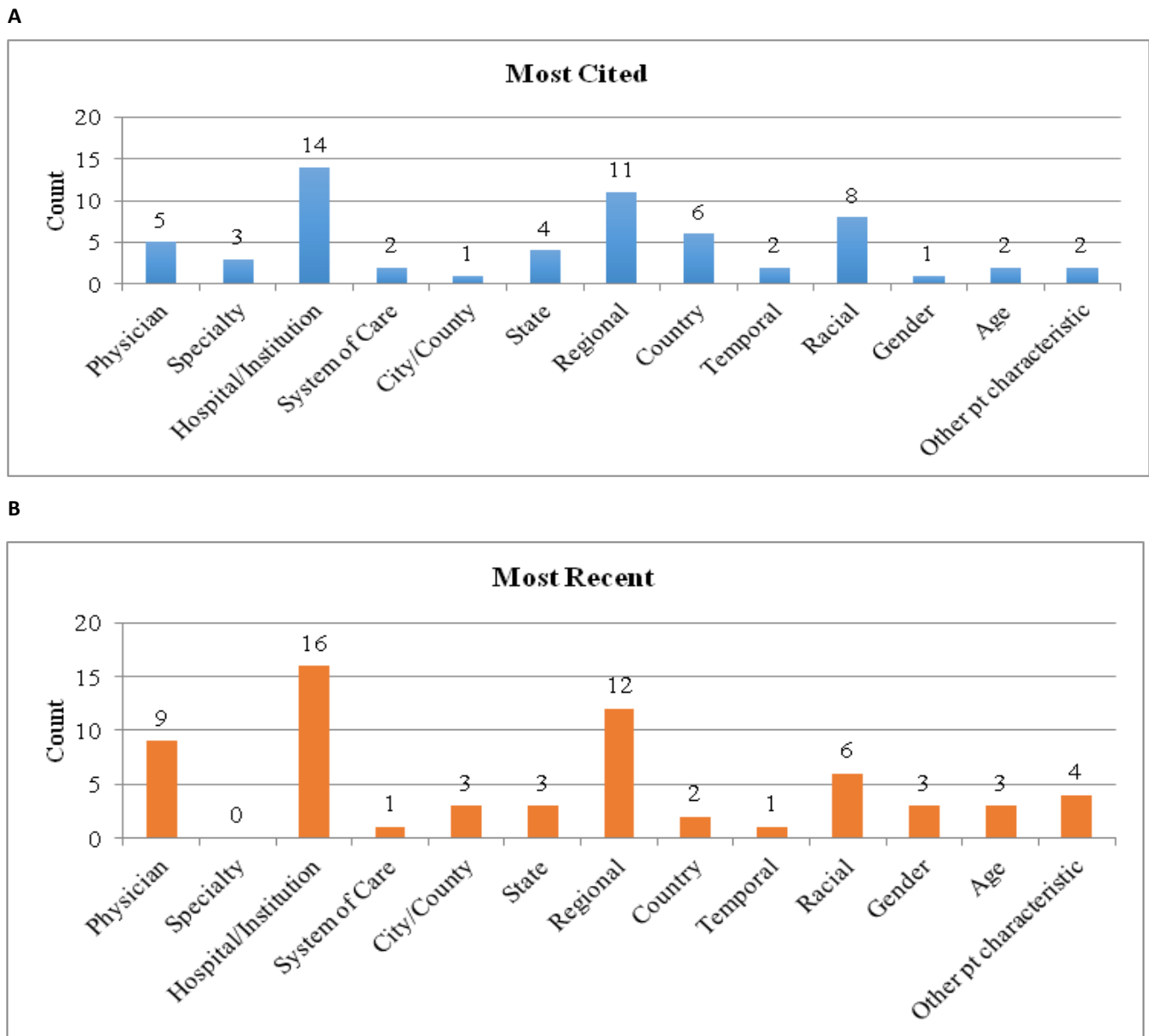
A

B

Figure 1: Count of categorical groupings employed in (A) the 50 most highly cited variation of care papers, and in (B) the 50 most recent variation of care papers

the category of "General Internal Medicine."

Our search generated 2,836 results. One reviewer (A.O.) then selected the 50 most highly cited and the 50 most recent papers that assessed variation in a medical practice. Each included article was read in full, and the metric(s) by which variation was studied was recorded in a Microsoft Excel spreadsheet. As such, each article could perform analyses and subsequently be coded in multiple categories.

The 50 most highly cited articles examined variation 61 times, and the 50 most recent articles examined variation 63 times. The most common comparisons were made between hospitals (14 times in the most cited grouping, and 16 times in the most recent grouping) and between regions (11 and 12, respectively). **Figure 1** shows common ways in which variation was analyzed among the 50 most highly cited and 50 most recent papers of variation in medical care. In short, in a relatively small sample of 100 variation papers, at least 13 alternative groupings were examined for variation.

Many of the challenges created by variation research are illustrated by a study of statins among dialysis patients. As of 2012, two large, multicenter randomized control trials (RCTs) showed no benefit of statin therapy among patients who were on hemodialysis, despite their high cardiovascular risk [4,5]. A third RCT, which combined a statin with ezetimibe, and included both dialysis and chronic kidney disease patients, found a reduction in myocardial infarction risk, no improvement in coronary death, and a trend towards increased death from all causes with the use of lipid therapy [6]. Thus, the totality of evidence suggests that statins have no role among hemodialysis patients.

A subsequent variation paper assessed variation in the use of statins among dialysis patients [7]. The results of this paper were as follows: use varied by patient sex, age, race, smoking status, functional status, whether diabetes was the cause of end stage renal disease, substance abuse status, number of comorbidities, and geography (state to state). The authors conclude that the variation they observed "may well reflect a lack of consensus regarding optimal management" [7]. Some of the variation they observed was not surprising; for instance, patients with more comorbidities were more likely to be on a statin. However, others were counterintuitive; for instance, men were less likely than women to use statins. With so many analyses conducted (and the possibility of other unreported analyses), it is difficult to know which of these findings are true. Is there a systematic bias among physicians to give statins to diabetic patients on dialysis? Probably, yes. Is there a bias to withhold statins from men on dialysis? We find this hypothesis implausible. The latter may simply be an artefact of multiple hypothesis testing.

The key question stemming from this line of reasoning becomes: Is knowledge of these variations useful? We already know that no trial has shown which (if any) patients on dialysis benefit from statins, and two trials have shown no benefit. Arguably, no patient on hemodialysis should be on a statin. Regarding variation research, Krumholtz argues that "the goal is not to eliminate variation but to guarantee that its presence throughout health care systems derives from the needs and preferences of patients" [8]. However, when a medication carries a real risk of side effects and no chance of benefit, there is no compelling need for a patient to take it and it is hard to imagine any preference that overrides these facts.

If the authors wish to show that statin use continues in this population despite RCT data suggesting that it is not beneficial, they only need to show its rate of use, not the countless groups among which it varies. If the authors instead doubt the validity of the RCT that yielded negative results on the use of statins in this population, they need to offer alternative data showing the benefit of statins. However, if the authors wish to show that statins continue to be prescribed to some dialysis patients but not to

others, then variation is the right test. With that said, this test should be based on a priori hypotheses, and corrected for multiplicity.

We have outlined some of the problems with variation studies. They are conducted in cases where there is clear evidence, as well as at times when there is no consensus. Datasets used are large and ubiquitous, and variation can be queried across many alternative groupings. As such, many of the results of variation papers may not reflect systemic biases in the use of a treatment, but may rather be an artefact of multiple hypothesis testing.

To improve upon validation studies, we propose that they be conducted only in cases where the evidence base for a practice is genuinely uncertain. When conducted, all proposed groupings should be pre-specified and noted in the paper. Ideally, the protocol for the study should be registered, as others have proposed [2]. The number of dimensions across which variation is queried should be limited to minimize spurious results. Finally, falsification testing should be added to variation papers. Specifically, if variation is used to make claims that reimbursement schemes drive discordant use, then interventions that are not subject to those maligned incentives should be shown not to vary. In fact, in our examination of this subject, we could find no investigation of variation that satisfied Krumholz's mark: where all variation is due solely to legitimate patient characteristics and preferences [8]. If indeed no such practice can be shown to meet this standard, then the standard itself should be reconsidered.

REFERENCES

1. Zhang Y, Baik SH, Fendrick AM, Baicker K. Comparing local and regional variation in health care spending. N Engl J Med. 2012;367(18):1724-31.
2. Ioannidis JA. The importance of potential studies that have not existed and registration of observational data sets. JAMA. 2012;308(6):575-6.
3. Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Intern Med. 2011;171(18):1675-6.
4. Fellström BC, Jardine AG, Schmieder RE, et al. Rosuvastatin and cardiovascular events in patients undergoing hemodialysis. N Engl J Med. 2009;360(14):1395-407.
5. Wanner C, Krane V, März W, et al. Atorvastatin in patients with type 2 diabetes mellitus undergoing hemodialysis. N Engl J Med. 2005;353(3):238-48.
6. Baigent C, Landray MJ, Reith C, et al. The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomised placebo-controlled trial. Lancet. 2011;377(9784):2181-92.
7. Wetmore JB, Mahnken JD, Mukhopadhyay P, et al. Geographic variation in HMG-CoA reductase inhibitor use in dialysis patients. J Gen Intern Med. 2012;27(11):1475-83.
8. Krumholz HM. Variations in health care, patient preferences, and high-quality decision making. JAMA. 2013;310(2):151-2.