

## Comment on: “Measuring the Volume–Outcome Relation for Complex Hospital Surgery”

Harold S. Luft<sup>1</sup>

Published online: 3 June 2017  
© Springer International Publishing Switzerland 2017

The recent paper by Kim et al., published in *Applied Health Economics and Health Policy* [1] questions analytic approaches to the oft-observed association between volume and outcome hospital care (V–O). They suggest estimating fixed and random effects, rather than simple logistic models and that reliance on inappropriately applied models may yield incorrect policy recommendations. While important, such conclusions derive from how one should generally do research, rather than their specific empirical results.

Unless simply mining data to see what surfaces, research begins with underlying hypotheses or plausible models. So, why would one hypothesize a V–O relationship? Bunker’s [2] original observation was that with more experience, surgeons got better—they were “learning by doing.” Early learning typically occurs during training. Beyond learning the technique, however, is knowing what to do in rare situations of unusual anatomy or a “slip of the scalpel.” Cumulative surgical experience would be more relevant here than current volume.

The usual “story” one tells is about surgeons, but most empirical work focuses on hospital volume. Perhaps it is the volume of the anesthesiologist, the operating room team, or the post-operative staff. If so, the focus should be the volume of similar procedures, not each specific procedure [3].

The above implicitly assumes cumulative experience is what matters, but volume per se could matter if skills decayed rapidly, requiring constant honing. If so, above some point additional volume might have little marginal impact [4].

Some hospital practices markedly reduce infection rates and other causes of death [5–7]. Does volume foster such practices? Very low volume hospitals may not even recognize their worse-than-average outcomes because their patient deaths appear infrequent and random. Some practices may require sufficient specialized staff, i.e. overall size, not procedure-specific volumes [3]. Hospitals may also improve practices over time without changing volume [8].

The causality may be reversed, i.e. selective referral. Good outcomes (perhaps a surgeon with outstanding skills) may attract more patients. None of these explanations are mutually exclusive, their importance may vary by procedure, and the policy implications of each are quite different. Careful procedure-specific empirical work is needed for meaningful recommendations. Ideally, one would have data from many hospitals with a range of volumes, preferably at a surgeon and team level, well-defined surgical conditions, outcomes associated with both patient-level factors and (potentially, surgeon and team volume), and plausible tests for selective referral.

Kim et al. [1] have 12 years of data from Florida, New Jersey, and New York—a major advance. They compare fixed and random effects models with simple logistic models ignoring hospital effects, but stop with significance tests, rather than exploring hypotheses. They look for significant coefficients on the volume variable, rather than seeking a critical minimum volume level and argue against selective referral based on the literature.

Being able to detect a “signal” is critical in empirical work. The 279,414 patients in their data are spread over

---

This comment refers to the article available at doi:[10.1007/s40258-016-0241-6](https://doi.org/10.1007/s40258-016-0241-6).

---

✉ Harold S. Luft  
lufth@pamfri.org

<sup>1</sup> Palo Alto Medical Foundation Research Institute, Palo Alto, CA, USA

500+ hospitals and 12 years. The coefficients of variation for volume over time at the hospital level are reasonable, but largely because the means are so low. For three procedures, median hospital-year volume is 2; for two the medians are 5 and 7 patients. Mortality rates are typically in single digits so most years most hospital have no deaths.

Much of the early V–O work focused on procedures that were risky, involved special surgical skills, and offered improved quality or life expectancy, such as coronary artery bypass graft or total hip replacement. Those studies recognized the limitations of discharge abstract data to account for important patient risk factors.

In-patient mortality is especially problematic as an “outcome” for cancer procedures [9, 10]. For some patients with cancer, surgical intervention offers a reasonable chance of cure. For others, surgery offers little incremental benefit and skilled surgeons may advise against it. Other surgeons, perhaps less busy ones, may accede to a patient wanting to “try everything.” Noting the presence of metastases won’t account fully for this, especially if those metastatic patients having procedures are operated on by poorer quality surgeons. Metastases strongly predict a short lifespan, not death due to the surgical procedure.

The authors argue that simple logistic regression makes it impossible to explore underlying relationships and that future research should focus more deeply. Although the data they have are less than ideal, they nonetheless use it less than optimally. Instead of focusing on whether volume is significant in six separate procedure-specific regressions, they could attempt to identify hospital-based patterns in outcomes. To allow for changing medical knowledge, they could estimate a logistic regression for each year based just on patient-factors, then compute a Z score for each hospital-year reflecting observed deaths and the estimated probability for each patient [11, 12]. This could be done for each procedure, all taken together, or grouping procedures performed by similarly trained surgeons. With 12 years of data, they could identify hospitals with consistently better-than-expected results—the positive deviants, and explore what characterizes them [13]. Also of interest would be hospitals that had worse-than-expected results and suddenly improved over a year or two—what changed? If outcomes improved without a change in volume, it suggests some organizational shift (or perhaps removal of a particular surgeon). Volumes increasing after outcomes improved would support a selective referral hypothesis.

Much of what has been published on the V–O relationship seems oriented towards proving or disproving its existence; it is not, however, like the General Theory of Relativity. Observing such a relationship in simple surgical procedures suggests problematic data; failing to see it in

certain other procedures would be a surprise. The goal, however, should be understanding what accounts for the relationship when it is observed to then learn how to improve outcomes. In the meantime, a simple rule of thumb for patients might be to avoid very low volume settings if higher volume hospitals are nearby. An even better approach would eschew the general results of regressions and simply ask for the outcome rates (preferably risk-adjusted) at the relevant hospitals. Public disclosures of such data, moreover, are likely to force improvements in care.

#### Compliance with ethical standards

**Funding** No external funding.

**Conflict of interest** No conflicts.

#### References

1. Kim W, Wolff S, Ho V. Measuring the volume–outcome relation for complex hospital surgery. *Appl Health Econ Health Policy*. 2016;14:453–64.
2. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality? *N Engl J Med*. 1979;301:1364–9.
3. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part I: the relation between hospital volume and outcomes for selected diagnostic categories. *Med Care*. 1984;22:98–114.
4. Al-Sahaf M, Lim E. The association between surgical volume, survival, and quality of care. *J Thorac Dis*. 2015;7(S2):S152–5.
5. Flood AB, Ewy W, Scott WR, et al. The relationship between intensity and duration of medical services and outcomes for hospitalized patients. *Med Care*. 1979;17:1088–102.
6. Birkmeyer NJO, Goodney PP, Stukel TA, et al. Do cancer centers designated by the National Cancer Institute have better surgical outcomes? *Cancer*. 2005;103:435–41.
7. Bilimoria KY. Advancing the quality of pancreatic cancer care. *AHRQ National Quality Measures Clearinghouse*. 2012. <https://www.qualitymeasures.ahrq.gov/expert/expert-commentary/36833/advancing-the-quality-of-pancreatic-cancer-care>. Accessed 5 March 2017.
8. Luft HS, Romano PS. Chance, continuity, and change in hospital mortality rates. Coronary artery bypass graft patients in California hospitals, 1983 to 1989. *JAMA*. 1993;270(3):331–7.
9. Birkmeyer JD, Sun Y, Wong SL, et al. Hospital volume and late survival after cancer surgery. *Ann Surg*. 2007;245:777–83.
10. Schneider EB, Hyder O, Wolfgang CL, et al. Patient readmission and mortality after surgery for hepato-pancreato-biliary malignancies. *J Am Coll Surg*. 2012;215:607–15.
11. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part II: the relation between volume and outcomes and other hospital characteristics. *Med Care*. 1984;22:115–25.
12. Luft HS, Brown BW Jr. Calculating the probability of rare events: why settle for an approximation? *Health Serv Res*. 1993;28(4):419–39.
13. Wong SL, Revels SL, Yin H, et al. Variation in hospital mortality rates with inpatient cancer surgery. *Ann Surg*. 2015;261:632–6.