On Residency Selection and the Quantitative Fallacy

J. Bryan Carmody, MD, MPH

It was 1964, and Robert McNamara had a problem. As the US Secretary of Defense, McNamara was responsible for the Vietnam War, and it was going poorly. A renowned intellect with experience in both academics and industry, McNamara had a reputation for using quantitative methods to solve difficult problems, and he set to work applying rigorous numerical analysis to the war effort.

To McNamara, the complexities of conflict could be reduced to simple mathematical equations: so long as the "body count" for hostile Viet Cong soldiers was greater than that of US personnel, victory was inevitable. McNamara directed the deployment of increasing numbers of offensive ground troops to Vietnam, and obsessively reviewed body counts, prisoners taken, weapons seized, and tons of bombs dropped—figures that rose continuously, even as the United States slowly lost the war. By 1967, McNamara grew skeptical that the war was winnable, and he tendered his resignation.

In this issue of the *Journal of Graduate Medical Education*, Sharma and colleagues[1] use quantitative methods to take on a more benign yet nonetheless complex problem: How can we predict which applicants will succeed in residency training programs? Unlike most previous studies, which have used narrow definitions of "residency success," the authors evaluated multiple domains of physician competency, ranging from patient and faculty evaluations to the American Board of Internal Medicine (ABIM) Certifying Examination. Using application data for 167 residents at a single center's internal medicine residency program from 2007 to 2014, the authors found that United States Medical Licensing Examination (USMLE) Step 2 Clinical Knowledge (CK) scores were the best predictor of residency performance across these varied domains. Notably, scores from USMLE Step 2 CK more consistently predicted measures of residency success than scores from USMLE Step 1.

The authors deserve credit for taking on an issue of such practical importance. The number of residency applications program directors must review has been increasing each year. In 2018, the average internal medicine residency program sought to fill only 14.6 positions in the Match,[2] but program directors reported receiving a mean of 2220 applications with which to do so.[3] Program directors rightly search for measures that can provide insight into an individual candidate's future success in their program—and if those measures are numerically precise and allow rapid interpretation (like USMLE scores), all the better. In light of the findings of this study, should program directors rely more heavily on USMLE Step 2 CK?

Closer inspection demonstrates the limitations of such an approach. While the association between Step 2 CK scores and outcome measures was statistically significant for all the outcomes studied, the predictive value varied greatly. The practical utility of the statistical association must be considered in light of the effect size.

As seen in prior studies, multiple-choice tests were best at predicting the results of future multiple-choice tests. The multivariable model for in-training examination (ITE) scores—which included both USMLE Step 1 and Step 2 scores—explained 55% to 57% of variation on the first 2 ITEs, while the concordance statistic for Step 2 CK in predicting passage on the ABIM Certification Examination was 0.82.

However, when USMLE Step 2 CK scores were used to predict faculty or patient evaluations, their predictive capability was much more limited. The R^2 values for the scores provided by faculty and patient ratings range from 0.03 to 0.11, and only 6% of the variation in overall resident class rank was attributable to variation in Step 2 CK scores. Moreover, the effect size for even *large* differences in Step 2 CK scores was modest. Using the regression coefficients from the multivariable models, a 50-point difference in USMLE Step 2 CK score portends just a 0.15 point change on the faculty evaluations and ~0.40 point on patient evaluations (both of which were scored on a 6-point scale).

In putting these findings from Sharma and colleagues into practice, we must be careful to avoid falling into the so-called McNamara fallacy.[4] Named after the Secretary of Defense, the McNamara or

DOI: http://dx.doi.org/10.4300/JGME-D-19-00453.1

quantitative fallacy has been summarized as the logical snare that results from the following progression:

1. Measure whatever can be easily measured.
2. Disregard things that cannot be measured easily.
3. Presume things that cannot be measured easily are not important.
4. Presume that things that are not measured easily do not exist.[5]

Against this backdrop, it is perhaps more understandable why USMLE Step 2 CK scores were the only measure associated with higher residency performance for each of the outcomes the authors studied. Much of the information that was not easily measurable was considered crudely or not at all. For instance, research experience was quantified as the number of publications and posters, and dichotomized to ≤ 5 or > 5. Receipt of medical school awards (any number and type) was considered only as a binary variable, while class rank and clerkship grades were ignored altogether since the various scales used by different schools made statistical inference difficult.

Admittedly, evaluating non-numerical data is less efficient and requires subjective judgment. But does that mean such information should be disregarded as unimportant, or that only quantitative data are capable of making meaningful predictions? After all, despite the statistically significant association, a multiple-choice test of clinical knowledge seems like a poor instrument with which to foretell a future resident physician's ability to listen to patients.

McNamara's steadfast belief in quantitative methods led him to believe that the United States was winning the war in Vietnam—even as commanders began to tell him the opposite. In using numbers to inform residency selection, we must be careful not to overextend what these metrics can really tell us, lest we fall prey to the quantitative fallacy ourselves.

References

1. Sharma A, Schauer DP, Kelleher M, Kinnear B, Sall D, Warm EJ. USMLE Step 2 CK: best predictor of multimodal performance in an internal medicine residency. *J Grad Med Educ*. 2019;11(4):412–419.
2. National Resident Matching Program; Data Release and Research Committee. Results of the 2018 NRMP Program Director Survey. Washington, DC: National Resident Matching Program; 2018. https://mk0nrmpcikgb8jxyd19h.kinstacdn.com/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf. Accessed June 26, 2019.
3. National Resident Matching Program. Results and Data: 2018 Main Residency Match. Washington, DC: National Resident Matching Program; 2018. https://mk0nrmpcikgb8jxyd19h.kinstacdn.com/wp-content/uploads/2018/04/Main-Match-Result-and-Data-2018.pdf. Accessed June 26, 2019.
4. O'Mahony S. Medicine and the McNamara fallacy. *J R Coll Physicians Edinb*. 2017;47(3):281–287. doi:10.4997/JRCPE.2017.315.
5. Basler MH. Utility of the McNamara fallacy. *BMJ*. 2009;339:b3141. https://doi.org/10.1136/bmj.b3141.

J. Bryan Carmody, MD, MPH, is Associate Residency Program Director and Assistant Professor of Pediatrics, Eastern Virginia Medical School.

Corresponding author: J. Bryan Carmody, MD, MPH, Eastern Virginia Medical School, 601 Children's Lane, Norfolk, VA 23507, 757.668.7244, james.carmody@chkd.org