How Do You Deliver a Good Obstetrician? Outcome-Based Evaluation of Medical Education

David A. Asch, MD, Sean Nicholson, PhD, Sindhu K. Srinivas, MD, MSCE, Jeph Herrin, PhD, and Andrew J. Epstein, PhD, MPP

Abstract

The goal of medical education is the production of a workforce capable of improving the health and health care of patients and populations, but it is hard to use a goal that lofty, that broad, and that distant as a standard against which to judge the success of schools or training programs or particular elements within them. For that reason, the evaluation of medical education often focuses on elements of its structure and process, or on the assessment of competencies that could be considered intermediate outcomes. These measures are more practical because they are easier to collect, and they are valuable when they reflect activities in important positions along the pathway to clinical outcomes. But they are all substitutes for measuring whether educational efforts produce doctors who take good care of patients.

The authors argue that the evaluation of medical education can become more closely tethered to the clinical outcomes medical education aims to achieve. They focus on a specific clinical outcome—maternal complications of obstetrical delivery—and show how examining various observable elements of physicians’ training and experience helps reveal which of those elements lead to better outcomes. Does it matter where obstetricians trained? Does it matter how much experience they have? Does it matter how good they were to start? Each of these questions reflects a component of the production of a good obstetrician and, most important, defines a good obstetrician as one whose patients in the end do well.


If someone were to ask which obstetrical residency program in the United States is the best, or even which programs are better than others, what information would we use to answer the question? We can judge the subjective quality of a program’s faculty or the halo of the hospitals that surrounds it, but those measures might not go very deep. More important, those measures may not be relevant for the very patients who receive the care that will be provided by the graduates of these programs. Although it is convenient to assess medical education using plausible and easily measurable process standards (do the students or residents get enough cases, enough lectures, enough sleep?) or by educational milestones, judging residency programs by actual patient outcomes is not only more patient-centered, it better supports innovation. When training is judged by process, the process becomes fixed. But when training is judged by outcomes, multiple pathways toward those outcomes may emerge, some of which might yield better outcomes. Our view is that although intermediate approaches to judging and assuring the quality of medical training have their place—and perhaps should retain a central role in assessment for practical purposes—we should be forever on the lookout for more clinically proximate outcomes against which to judge the ways in which we produce doctors.

We illustrate this thinking by reviewing some work we have done in the field of obstetrics, where we used an important clinical outcome, in this case maternal complications of delivery, to assess the training of obstetricians. On the basis of maternal outcomes, we examine whether it matters where obstetricians train, whether experience matters, and whether their initial skill matters. Each of these questions relates to how our medical education system produces a competent physician workforce, potentially providing answers to the question of how to make obstetricians better in ways that are directly relevant to women.

Does It Matter Where the Obstetrician Trained?

Most physicians seem to think that training program matters. Medical students aim to get accepted into prestigious residencies. Established physicians brag about where they trained. Prospective employers may give hiring advantages to graduates of certain residencies. Each of these observations suggests that people believe that it matters where doctors train, but these might be empty signals. For these intuitions to be meaningful, we would need to be able to understand something about doctors’ ability to improve the health of the patients they later care for based on where they trained.

In this context, our results in obstetrics are promising. We analyzed all hospital-based deliveries in New York and Florida in the 16 years from 1992 to 2007. This sample was defined by those deliveries performed by licensed obstetricians who had completed a U.S. obstetrics–gynecology residency and who had performed at least 10 deliveries after residency. To evaluate residency programs, we further required that each physician be a graduate of a residency program that contributed at least 10 physicians to the sample. In the end,
we reviewed 4,906,169 hospital-based deliveries performed by 4,124 physicians who were graduates of 107 residency programs (of 249 in existence) from 22 states and the District of Columbia.

We found substantial and stable differences in maternal complication rates across programs. The top quintile of programs produced graduates with an average maternal complication rate of 10.3%, but the bottom quintile of programs produced graduates whose maternal complication rates were one-third higher, 13.6%. These differences were sustained over time, and programs fell into the same pattern across vaginal, cesarean, and total delivery outcomes. Programs had similar complication rates for hemorrhage, infection, laceration, and operative complications. These results were preserved after adjusting for observable differences in maternal health, but those adjustments were minor because women who deliver babies are generally very healthy.

So, some residency programs consistently turn out graduates who take better care of patients. But they might turn out better graduates only because they admit better graduating medical students in the first place. A woman planning to have a baby should not care whether her doctor developed talent before residency or after residency, but medical educators ought to care, because medical educators want to know what elements contribute to training effects and not just selection effects. To tease apart these two factors, we adjusted our previous analyses on the basis of each physician's licensing examination scores. These scores are largely determined by individuals' knowledge before residency and so might adjust for differences in who goes to which residency program. But adding these scores into our model hardly changed the results at all: The ranking of individual programs was unchanged for composite or individual measures of quality, and the difference between the best and worst programs was reduced by less than 0.1% in absolute terms, suggesting that licensing examination scores contribute almost nothing to observable clinical quality in this case.

These results confirm the belief that it matters where doctors train. Women might be well advised to pick obstetricians on the basis of whether they trained in particular residency programs, and medical educators ought to try to find out what it is about those residency programs at the top that allows them to produce physicians who are consistently better at achieving important clinical outcomes.

**Does Experience Matter?**

Most people also seem to think that experience matters. Many patients instinctively seek out experienced doctors. Empirically, a considerable literature across different clinical domains associates higher volumes with better outcomes, but although this might imply that practice makes perfect—the theory behind the intuitive quest for experience—the same pattern might result if women preferentially seek out obstetricians with good outcomes. Instead of volume driving outcomes, outcomes might drive volume, and this causal direction might be particularly active in obstetrics, where women have time to shop around and, through their social connections, might have the word-of-mouth information on which to base their comparisons. Therefore, we focused not on the number of babies delivered but on a physician's years of practice—because although it is possible that good quality can lead to delivering more babies, it should not lead to being older.

We looked at all hospital-based deliveries in Florida and New York from 1992 to 2010 and focused on 6,704,311 of them (79%) that were performed by 5,175 licensed obstetricians who completed residency after 1969 and delivered babies in more than one year. We used data on these 57,736 physician-years of practice to examine whether maternal complication rates changed with years of experience. We repeated the analysis on the subset of 3,044 physicians who remained in practice from 1992 to 2010 because it is possible that physicians might drop out of obstetrics if they found they weren't any good at it, making more experienced obstetricians look better by comparison. The results were the same either way and revealed some powerful findings.

We found that experience matters. As physicians gained years of experience, their maternal complication rates fell. This was true for complications after vaginal deliveries, after cesarean deliveries, and for all of their deliveries combined. We also found that these quality improvements continued for three decades after residency. After controlling for secular trends and patient factors, complication rates fell two percentage points in the first decade of practice (on a baseline rate of about 15%), by another one percentage point in the second decade, and by another 0.5 percentage points in the third decade. Although we might have guessed that experience matters, we might not have guessed that an obstetrician with 20 years of experience had an advantage over one with 15 years of experience. And although women might be well advised to pick obstetricians on the basis of their years of practice, medical educators are left with the question of what it is that obstetricians learn in their second or third decade that they did not learn in their first. Because we clearly do not want to extend obstetrical residency training to three decades, the question for educators is whether there are ways (like coaching or simulation) to emulate the quality returns that seem to come from experience so that patients can benefit from them sooner.

**Does Initial Skill Matter?**

Finally, if quality later in a physician's career is determined by residency and experience, we might ask, Does initial performance predict future success? The question is important because although we have shown that experience matters, it might also be the case that a physician's relative quality is substantially determined by where he or she starts.

We looked at the 1,864 obstetricians who began practice in New York and Florida between 1992 and 2010 and followed their maternal complication rates through 15,675 physician-years and 2,005,043 deliveries. Comparing those who start out in the best- and worst-performing quartiles, there is considerable spread between complication rates in their first year; and, to no surprise, these two groups approach the overall average over time. But they do so gradually, and they do not quite get there: Over 15 years, the physicians who start out with relatively poor outcomes never quite catch up to everybody else, meaning that the impact of initial skill persists. In fact, we found that variation in overall quality is determined far more by initial skill than it is by
number of deliveries performed. That finding reflects a potentially important contribution to the literature about what makes a good doctor.

Thus, whereas initial performance is not a guarantee of future returns in a mutual fund, initial performance might be particularly useful when selecting an obstetrician. Although it is not so easy for a woman choosing an obstetrician to learn that obstetrician’s skill on first entering practice, these results suggest that hospitals might be well advised to hire star graduates and hold onto them.

Toward Outcome-Based Evaluation of Medical Education

This approach has many limitations, and obstetrical care may be a domain where it works better than others. In particular, although most of health care increasingly involves teamwork, and most physicians take care of patients with many different conditions, delivering babies is what many obstetricians do predominantly, and delivery often involves only a single physician and a limited number of other clinicians like nurses and anesthesiologists. These conditions support more of a one-to-one association between physician and outcome. In addition, most physicians who deliver babies deliver a lot of them, and so generate samples large enough to support statistical resolution around those outcomes. These conditions rarely obtain in other fields. And of course, although we did not have data on neonatal outcomes, arguably these would be the most critical metric for most mothers.

Nevertheless, the approach of evaluating the training and production of physicians by the quality of the product is particularly appealing when that quality is measured by patient outcomes. Patient outcomes, already central to the evaluation of hospitals and individual physicians, are not yet commonly used to evaluate medical education. This is an important goal because evaluating educational programs on the basis of the care provided by their graduates provides a more compelling standard against which to judge educational quality than standardized tests of knowledge or rankings based on structural factors like one sees in publications like U.S. News & World Report. Clinical outcomes are among the most patient-centered, and they most connect our investment in medical education with the goals of that investment. Although it is easier to structure training programs and their evaluations around their processes or intermediate outcomes—work hours, number of cases, competencies, or milestones—the same reasoning supports a view that we invest in ways that make it easier to collect the data we would rather have. In the end, we should advance efforts to make outcomes data more available and more easily tethered to training processes, and we should subordinate the measures of educational quality we currently use to those that reflect what patients value.

Funding/Support: None reported.

Other disclosures: None reported.

References


Ethical approval: Reported as not applicable.

Previous presentations: This commentary is based on a presentation at the Research in Medical Education session at the 2012 Association of American Medical Colleges annual meeting in San Francisco.

Dr. Asch is a physician, Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, professor, Perelman School of Medicine and Wharton School, University of Pennsylvania, and executive director, Penn Medicine Center for Health Care Innovation, Philadelphia, Pennsylvania.

Dr. Nicholson is professor, Department of Policy Analysis and Management, Cornell University, Ithaca, New York, and research associate, National Bureau of Economic Research, Cambridge, Massachusetts.

Dr. Srinivas is assistant professor of obstetrics and gynecology, Division of Maternal Fetal Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Dr. Herrin is assistant professor of cardiology, Yale School of Medicine, Yale University, New Haven, Connecticut, and senior statistician, Health Research & Educational Trust, Chicago, Illinois.

Dr. Epstein is health science specialist, Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, and research associate professor, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.