Not long ago, a fellow doctor told me that his local health care insurers, in an effort to improve care and rein in costs, had been evaluating physicians and paying them according to their “quality ranking.” With “pay for performance” reimbursement, doctors who had, for example, managed more timely follow-up and achieved better test results with their diabetic or hypertensive patients would rank more highly and earn more in financial bonuses than physicians whose patients failed to meet the insurers’ guidelines.

Despite his own solid “ranking,” my colleague found himself growing more and more disenchanted with these types of reimbursement programs. “It sounds like a great idea,” he said, “but it assumes that what I do or say always has a direct effect on my patients’ health.” He described a long-term patient he had seen earlier in the week; she had poorly controlled diabetes yet failed to come into his office regularly in spite of his numerous requests.

“She just can’t afford to take that much time off from work,” he said, adding with a sigh, “Does that make me a worse doctor?”
For a little over a decade, amid calls for improved quality and greater transparency, insurance companies across the country have been “incentivizing quality” — shifting away from traditional fee-for-service reimbursement and turning to pay-for-performance programs. Based on the experiences of health maintenance organizations in the mid-1990s with assessing quality, public and private third-party payers started applying many of the same evaluation criteria to individual physicians and linking reimbursement bonuses to similar, if not identical, clinical goals. The hope was that these well-established goals, such as regular mammograms and Pap smears and screening for colon cancer and high cholesterol, would not only improve care and cut costs over the long run but also provide payers and the public a standardized way to compare doctors and the care they deliver.

Health care experts applauded these early initiatives and the new focus on patient outcomes. But over time, many of the same experts began tempering their earlier enthusiasm. In opinion pieces published in medical journals, they have voiced concerns about pay-for-performance ranging from the onerous administrative burden of collecting such large amounts of patient data to the potential worsening of clinician morale and widening disparities in health care access.

But there has been no research to date that has directly studied the one concern driving all of these criticisms — that the premise of the evaluation criteria is flawed. Patient outcomes may not be as inextricably linked to doctors as many pay-for-performance programs presume.

Now a study published this month in The Journal of the American Medical Association has confirmed these experts' suspicions. Researchers from the Massachusetts General Hospital in Boston and Harvard Medical School have found that whom doctors care for can have as much of an influence on pay-for-performance rankings as what those doctors do.

Drawing on the experiences of more than 125,000 patients, the researchers first ranked the doctors based on criteria commonly used in pay-for-performance reimbursement plans. While they found that doctors who took care of older or sicker patients tended to rank higher, presumably because of more frequent patient follow-up, the researchers also discovered that primary care practitioners who cared for underinsured, minority and non-English-speaking patients tended to have lower quality rankings than their counterparts.
Using statistical modeling, the researchers then attempted to rerank all the doctors after adjusting for differences in patient characteristics. When they factored patient race, ethnicity, primary language and insurance status into their physician evaluations, many of the original rankings changed, with doctors who worked in community centers — and therefore with more minority and non-English-speaking patients — being more likely to improve in ranking, often by more than 10 percentile points.

“Pay-for-performance can work,” said Dr. Clemens S. Hong, lead author and a general internist at the Massachusetts General Hospital, “but we need more sophisticated measures to make sure we are actually measuring physician quality.”

Addressing the discrepancies highlighted in the study will be challenging. Some health care experts, for example, have suggested paying for improvement rather than performance. Under this model, if over the course of a year doctors are able to increase the numbers of patients who obtain screening tests according to a standardized schedule, their reimbursement would increase.

But such a payment model comes with its own set of issues. “If your patient mammogram rate is 100 percent, there’s no way to improve,” Dr. Hong said. “And do we really want to pay doctors who have only half of their patients getting their mammograms done on time?”

Nevertheless, Dr. Hong and his co-investigators believe that their study only confirms the importance of research efforts focused on developing more sophisticated pay-for-performance measures. “It’s human nature to go to where the incentives are,” Dr. Hong said. Without specific assessment criteria, attributes that are currently measured indirectly or bypassed altogether, like the ability to communicate, build rapport and work with more vulnerable and challenging patients, risk being devalued or ignored.

“When you shine a light on certain measures, you take away from others that are also important,” Dr. Hong said. “Fee-for-service has already driven physicians away from primary care. If we don’t address patient differences, we may do the same thing with pay-for-performance.”