

Methods of Observing Variations in Physicians' Decisions: The Opportunities of Clinical Vignettes

Lara Converse, MA, Kirsten Barrett, PhD, Eugene Rich, MD, and James Reschovsky, PhD

Mathematica Policy Research, Washington, DC, USA.

To support their efforts to promote high quality and efficient care, policymakers need to better understand the key factors associated with variations in physicians' decisions, and in particular, physician deviations from evidence-based care. Clinical vignette survey instruments hold potential for research in this area as an approach that both allows for practical, large-scale study and overcomes the data quality challenges posed by analysis of clinical data. These surveys present respondents with a narrative description of a hypothetical patient case and solicit responses to one or more questions regarding the care of the patient. In this review, we describe various methods for measuring variations in physicians' decisions and highlight a range of design features researchers should consider when developing a clinical vignette survey. We conclude by identifying areas for future research.

KEY WORDS: clinical vignettes; evidence-based care; clinical decisions.

J Gen Intern Med 30(Suppl 3):S586–94

DOI: 10.1007/s11606-015-3365-8

© Society of General Internal Medicine 2015

INTRODUCTION

Physicians' clinical decisions frequently deviate from evidence-based care as reflected, for example, in clinical practice guidelines.^{1,2} These deviations from evidence-based guidelines, resulting in variations in clinical practice, may be appropriate for selected patients, but policymakers must understand the nature and extent of these deviations to be reassured that these clinical decisions are not causing harm or increased costs. The challenge has been to measure the variation in care and account for contributing factors, some of which (for example, case mix or patients' financial status) are not under clinicians' control, and to detect those unwarranted variations that can be associated with inefficient resource use and—sometimes—unnecessary risk to patients.

In this literature review, we assess different methods of measuring variations in physician decisions. We focus in particular on techniques that may support research into what organizational features and payment policies promote evidence-based decisions in individual clinical scenarios that contribute substantially to health care use and costs. Although each method has strengths and weaknesses, we devote most of our attention to clinical vignettes as an approach worthy of further research, given that these tools are presently the most

feasible method to measure variations in individual physician decisions about pertinent diagnostic and treatment options.

COMMON APPROACHES TO MEASURING POINT-OF-CARE CLINICAL DECISIONS

Although technology in the field is evolving, researchers have regularly used several methods to measure variations in physician point-of-care decisions (Appendix A). Two methods—medical record abstraction and claims data analysis—are based on readily available data on the care of actual patients, but using these approaches generally requires sophisticated statistical analysis to control for differences in patient case mix among providers or across settings. Two other options—standardized patients and clinical vignettes—require primary data collection from clinicians, thus increasing the clinician burden of the research. By using the latter approaches, however, one can directly measure physician decisions and control for case mix by soliciting a decision on a single case or a consistent set of cases from all sampled providers (Table 1).^{3–7}

Medical Record Abstraction. Medical record abstraction relies on a trained chart abstractor to review clinical records and produce a data set of physician decisions as physicians themselves record them.³ The availability of chart records, as well as the fairly low burden placed on physicians or medical practices to provide these data (which are generated in the course of routine patient care), are strong advantages of this method. However, medical information that cannot be extracted automatically from an electronic health record must be abstracted manually by trained researchers, the time^{4,7} for and expense^{3,4,5,6} of which may severely limit the sample size that can be included in an analysis. Both handwritten and electronic medical records also suffer from “recording bias,” in that not all relevant medical data or services may be recorded.^{3,5,6,8}

Claims. As a record of physician point-of-care decisions, computerized administrative claims data share many of the advantages of medical record abstraction, being both widely available and requiring no provider time for data collection. These data are also fairly inexpensive to gather, avoiding the costs of hiring medical record abstractors, administering

Published online June 24, 2015

Table 1. Traditional Methods of Measuring Variations in Physician Decisions

Method	Description	Advantages	Disadvantages
Medical Record Abstraction	<p><i>Manual</i> Chart abstractor reviews medical records to produce a data set of physician decisions³</p> <p><i>Electronic</i> Data from some EHR fields can be automatically extracted for analysis</p>	<p>Record of actual patient interactions</p> <p>Low burden for physicians</p> <p>EHR data may be inexpensive to extract and manipulate</p>	<p>Manual abstraction is costly³⁻⁶ and time-intensive^{4,7}</p> <p>Sample sizes are likely to be small</p> <p>Medical records often lack detail and may omit decisions that occurred^{3-6,8} and/or include decisions that never occurred⁸</p> <p>Only some EHR data fields can be automatically extracted (e.g., checkboxes, drop-down menus)⁵</p> <p>May be challenging to sufficiently control for differences in providers' patient case mixtures^{3,4}</p> <p>Physicians who permit access to medical records may not be representative</p>
Administrative Claims	Analysis of claims submitted by physicians for payment	<p>No burden for physicians</p> <p>Data is readily available (for at least some patients)</p> <p>Inexpensive data collection</p> <p>Sample sizes likely to be large</p> <p>Results more likely to be generalizable</p>	<p>Data often lacks detail⁴ and contains inaccuracies^{9,10}</p> <p>Some clinical decisions may not be captured (e.g., referrals)⁵</p> <p>Data for some patients may be unavailable or not easily attainable (e.g., uninsured, Medicare Advantage)</p> <p>May be challenging to sufficiently control for differences in providers' patient case mixtures,⁴ and bias in coding of claims may distort results (e.g., upcoding)¹¹</p> <p>Often not possible to attribute a given service to the provider responsible for placing the order (e.g., an imaging study is likely to be attributed to a radiologist, rather than the primary care physician who submitted the order)</p>
Standardized Patients	Trained actors role-play a specific patient case and record the physician's clinical decisions (physician is blinded)	<p>Records actual clinician decisions</p> <p>Able to capture clinical decisions that are not available through claims nor documented in the medical record</p> <p>By presenting a standardized patient case, controls for differences in physicians' case mixture⁶</p>	<p>Hiring/training standardized patients is costly^{3-6,12}</p> <p>Logistically challenging to organize visits of standardized patients⁶</p> <p>Sample sizes are likely to be small</p> <p>High burden for physicians^{3,4}</p> <p>Treating standardized patients reduces physician time for actual patients³</p> <p>Not all decisions may be realistically simulated (e.g., pediatrics, signs of trauma)⁷</p> <p>Physicians who agree to participate in blinded standardized patient study may not be representative</p>

surveys, and using standardized patients (see below). Moreover, these advantages tend to increase the sample size of claims-based analyses, permitting more generalizable results. Yet, for many services (for example, advanced imaging), claims will only reliably identify the provider who is paid to perform the service, while the provider who decides to order the service (and the parameters of that decision) is of greater interest to policymakers. Claims also normally do not contain all the clinical data,⁴ such as patients' symptoms or detailed elements of their medical history, that can shape physicians'

point-of-care decisions but do not affect reimbursement, and many clinical decisions (such as referrals) are not reflected in claims at all.⁵

Standardized Patients. Standardized patients, used in what is often considered the "gold standard" approach to measuring physician decisions, are trained actors who observe physician performance. Actors are asked to portray a particular patient history or set of characteristics (for example, a propensity to

demand tests) during a clinical visit and to document the services they receive during the encounter. Like medical record abstraction, the use of standardized patients is presently valuable on a small scale, but likely unrealistic in large-scale studies of variations in point-of-care decisions across diverse communities and practice settings.^{5,6} The major limitations are the high cost of training and compensating standardized patients,^{3-6,12} and the logistical challenges of organizing and coordinating their visits.⁶ Accordingly, studies using standardized patients will necessarily involve small samples. Importantly, too, providing care for standardized patients takes physician time away from caring for real patients, burdening physicians and their practices to a much greater extent than would other methodological approaches.^{3,4}

INTRODUCTION TO CLINICAL VIGNETTES

Given the challenges posed by the use of medical record abstraction, claims data analysis, and standardized patients to affordably and reliably measure variations in clinical decisions across settings and specialties, the most feasible method may be a fourth option: physician surveys using clinical vignettes—that is, simulated patient cases. A vignette case generally specifies a hypothetical patient's age, gender, medical complaint, and health history (see Appendix A). Based on the details provided in the

case, the respondent is asked to answer one or more questions regarding diagnosis or treatment of the patient (Table 2).⁴

Vignettes may be administered on paper, by telephone, or in person, or they may be computer administered, sometimes incorporating an audio or video recording¹³ of the patient's responses. They have been used in a wide range of settings, including medical licensing and board certifications,⁴ the training of medical students,^{4,7} and continuing medical education courses.¹⁴ Researchers have used this tool to explore variations in physician decisions, both to characterize the extent of variation that exists¹⁵ and to assess factors that might contribute to it.¹⁶ Vignettes have also figured importantly in studies of the influences of patient race¹⁷ and gender¹⁸ on physicians' evaluation and treatment decisions.

Clinical vignettes are likely less expensive to use than both standardized patients and manual medical record abstraction,^{3,4,7} perhaps even after the costs of instrument development and administration are taken into account. Just as importantly, they are free from the challenges posed by incomplete patient medical records or claims data.⁴ Logistically, clinical vignettes are more practical⁶ and less burdensome to physicians than using standardized patients,³ and data collection is easier and faster than with medical record abstraction.^{4,7} Importantly, given these advantages, sample size in a vignette study is likely to be substantially larger than is feasible with standardized patients or manual medical record abstraction.^{4,7} (As described in Table 1,

Table 2. Example of a Clinical Vignette Closed-Ended Question Format

Question format	Example
Yes/no	1. Would you order a CT scan? <input type="checkbox"/> Yes <input type="checkbox"/> No
Multiple choice	2. What treatment would you order for this patient? <input type="checkbox"/> Order imaging study <input type="checkbox"/> Schedule a followup visit in three months <input type="checkbox"/> Advise the use of ibuprofen <input type="checkbox"/> Refer to physical therapy
Likert scale	3a. How often would you refer similar patients to a cardiologist? <input type="checkbox"/> Always <input type="checkbox"/> Usually <input type="checkbox"/> Sometimes <input type="checkbox"/> Never 3b. How often would you refer similar patients to a cardiologist? <input type="checkbox"/> 76–100 percent of patients <input type="checkbox"/> 51–75 percent of patients <input type="checkbox"/> 26–50 percent of patients <input type="checkbox"/> 0–25 percent of patients
Fill in the blank (numeric)	4. For what percentage of similar patients would you prescribe warfarin? [Provide a number from 0 to 100 percent]

the low cost of automated EHR abstraction is currently offset by its limitations in many clinical scenarios.)

Clinical vignettes do have a number of limitations, however. Since they inquire about the treatment of hypothetical patients outside of real-world contexts (for example, without the effects of practice-level influences or time constraints on physicians),⁴ physicians' responses may not reflect what occurs in actual practice.^{3,4} Peabody et al. (2000), for instance, raise the notion of "social desirability bias," which may cause physicians to respond to vignettes based on their knowledge of how they *should* practice rather than how they *actually* practice. For instance, a physician who is far behind in seeing patients might not perform examinations he or she recognizes would be recommended (and would likely report in a vignette).^{3,4}

Like any measurement instrument, the clinical vignette approach may also suffer from high costs of instrument development and validation, as well as non-response bias—concerns avoided by claims data analysis, and perhaps, by medical record abstraction. It is important to note that instrument development costs will increase if vignettes must be regularly updated as clinical guidelines change or if different vignettes are required for different specialized roles or practice settings. Lastly, unlike medical record abstraction or claims, clinical vignettes impose burden on physicians (albeit fairly minimal) by requiring them to submit a survey response.⁷

CONSIDERATIONS FOR DESIGNING AND ADMINISTERING CLINICAL VIGNETTES

Research by Peabody et al. (2000, 2004) provides important guidance on designing and administering clinical vignettes

that accurately measure actual physician behavior. Important attributes of vignettes used in their studies include: (1) allowing open-ended responses, (2) presenting realistic time constraints, (3) offering patient cases with varied and realistic levels of clinical complexity, (4) providing real-time information in response to physicians' answers, and (5) using a design that detects both necessary and unnecessary care.^{3,7}

Designing clinical vignettes that yield valid results requires a thorough understanding of the study purpose, insight into the study population, and an appreciation of the need to balance cost and rigor. In this section, we describe different options for vignette design and conclude with tables of decision points (Table 3) and other considerations (Table 4) that should be weighed in designing a relevant, cost-effective vignette likely to generate responses that accurately reflect physician practice.

Selecting Decisions to Study. Not all clinical decisions are appropriate for measurement with vignettes. Decisions are best suited for the vignette approach when they occupy a middle ground with respect to their evidence base; that is, there should be clear evidence indicating an appropriate choice for patients with certain characteristics, but the "right" answer should not be so obvious as to fail to solicit a variation in responses or raise concerns regarding social desirability bias. For instance, vignettes regarding the decision to counsel on smoking cessation are more likely to face social desirability bias (and insufficient variation in responses) than decisions where the best practice is less universally known. The current state of practice among practitioners being surveyed should also occupy a middle ground; assessing decisions that are already known to be

Table 3. Decision Points for Constructing a Clinical Vignette

Consideration	Options	Advantages	Disadvantages
Question type	Open-ended Closed-ended	Avoids social desirability bias or bias from cueing Easier and less expensive to score	Time intensive and expensive to score ⁷ Biased responses possible ^{4,7,9}
Question format	Dichotomous Multiple choice Likert scale Fill in the blank with numeric response	Easy to administer and interpret Easy to administer and interpret Familiar scale Avoids bias from providing response options ¹⁹	Biased responses possible ¹⁹ Biased responses possible Respondents may not interpret the scale in the same way ²⁰ More likely than other question formats to have errors occur in data entry of survey responses ¹⁹
Mode of administration	Hard copy Telephone/in-person interview Computer or tablet	Easier than in-person interviews if respondents are geographically dispersed For vignettes on sensitive topics, less likely than interviews to elicit biased responses Likely least expensive of all modes of administration Telephone interviews are easier than in-person interviews if respondents are geographically dispersed May be designed to allow adaptation of vignettes to physicians' responses Easier than in-person interviews if respondents are geographically dispersed May be designed to allow adaptation of vignettes to physicians' responses For vignettes on sensitive topics, less likely than interviews to elicit biased responses	Greater respondent burden than interview or computer administration if vignette is open-ended In-person interviews are challenging and costly if respondents are geographically dispersed Possibility of biased responses, particularly for vignettes related to sensitive topics Likely most expensive of all modes of administration Not all respondents may be comfortable using computers Likely highly expensive

Table 4. Other Design Considerations for Developing a Vignette Survey

Design factors	Considerations
Selecting decisions to study	Vignette should be based on a clear evidence base Correct answer should not be obvious to respondents Variation in practice should be present and policy-relevant
Realism	Vignettes should present patient complexity similar to that encountered in actual practice ^{4,7} Complexity must be appropriate for avoiding satisficing ²¹
Establishing validity	Have experts review vignettes to ensure they are written and designed appropriately Examine criterion validity by comparing results from the vignette and the use of a similar standardized patient
Pre-testing	Check that instructions and vignette language are clear ⁴ Verify heterogeneity in responses Assess for vignette equivalence ²²
Administration	Emphasize that the vignette is not meant to test the competency of an individual physician, but rather to understand the decisions physicians tend to make when presented with particular cases in practice ⁴ For vignettes administered by computer or in person, consider imposing time constraints ⁷

universally present or absent is likely to be of limited value. Finally, as the Choosing Wisely initiative²³ has recognized, decisions will ideally be such that variability has consequences of interest to policymakers, whether because of variation in cost or variation in risk to patients.

Open-Ended Versus Closed-Ended Questions. Clinical vignettes may use either open-ended or closed-ended questions. Open-ended questions rely on free response; respondents provide written (or typed) answers to how they would care for the patient, without any prompts or limitations to guide them.⁹ Closed-ended questions can be variously structured, requiring respondents to make a selection from a checklist, mark “yes” or “no” for a series of items, select an option from a multiple choice list, rank items, or make a selection within a range on a Likert scale (Table 2).

The open-ended or closed-ended structure of questions used by a clinical vignette can affect data quality. Vignettes that offer open-ended responses allow the physician to report what he or she would do in a given situation, without guidance or cueing. Although closed-ended vignettes are used in many applications, including the U.S. Medical Licensing Examination (USMLE),²⁴ the presentation of response options in closed-ended vignettes may cue a physician to respond in a certain way, especially if he or she views one or more options as “correct” or believes the researcher is seeking a specific answer (social desirability). Choices based on what is thought to be the correct response may be inconsistent with actual behavior or decisions in practice and can result in an overestimate of performance.⁹

Comparing open-ended and closed-ended vignettes directly, Pham et al. (2009) observed that closed-ended vignettes yielded a higher rating of quality of care than responses to identical vignettes presented in an open-ended format.⁹ Closed-ended responses may reflect both a “cueing” effect and testing ability, confounding assessment of clinical decisions. Overall, although closed-ended vignettes may generally accord with actual practice behavior, open-ended vignettes

may result in stronger criterion validity and be better able to distinguish among the decisions of physicians.^{9,25}

Question Format. When constructing a closed-ended vignette, careful selection of question format is essential. One option, a dichotomous “yes/no” response (example 1 in Table 2), is easy to administer and interpret, but may result in bias if some respondents are undecided.¹⁹ Responses to a multiple choice question (example 2), although also simple, may likewise be biased, unless the responses provided represent all options a physician would consider for the given scenario in practice.

Likert scale questions solicit answers to the question of “how likely” or “how often” a physician would make a given decision (examples 3a, 3b). A Likert scale format may be appealing because of its familiarity; however, two physicians may interpret the same term on the scale differently.²⁰ Providing categories with ascending or descending numerical values (example 3b) avoids this weakness, but researchers should ensure the numerical range is distributed equally across categories. For example, if numeric categories are narrower toward the endpoint of a scale, respondents may make conclusions about the average frequency and adjust their responses accordingly.²⁰

A fill-in-the-blank question that solicits a numeric value along a range is, in a sense, a compromise between open-ended and closed-ended formats (example 4). By allowing respondents to provide a free-form numerical answer, the question avoids any bias imposed by providing closed-ended options and is also much easier to “score” than an open-ended question,¹⁹ especially if a numeric response is mandated (as permitted by computerized vignettes). In a mail survey, however, numeric fill-in-the-blank items may be more prone to errors in data entry than numeric Likert scale questions that are truly closed-ended.¹⁹

Mode of Administration. Clinical vignettes can be administered in hard copy (paper and pencil), by telephone

or in-person interviewing, and/or by computer or tablet. Factors to take into account when selecting a mode include location of respondents, respondent access to and comfort with computers, vignette design, social desirability bias (typically greater with telephone or in-person interviews), and budget constraints (see Table 3).

Realism. To simulate most closely the experience of providing patient care, vignettes should present patient cases that are similar in complexity to those seen in actual clinical practice.^{4,7} Incorporating audio or video to present the patient case may be one means of achieving this result.¹³ Another is to use a computerized vignette that imposes a sequential order on the physician's response—that is, a physician may not change his or her planned physical examinations after selecting a treatment plan, for instance.^{3,7}

Vignettes should also avoid prompting “satisficing,” or the act of providing a nonoptimized response—a common occurrence among survey respondents. Satisficing can result when a task is too difficult for the respondent or the respondent's motivation to participate is low. Offering interesting vignettes of appropriate complexity can help reduce satisficing.²¹ Optimizing vignettes to minimize satisficing across a randomized sample of physicians who treat patients of varying complexity will invariably compromise the ability to target the vignette content to the responding physician, however.

Establishing Validity. Ideally, the criterion validity of vignettes would be reported in each research article. This is both time and cost prohibitive, however, as standardized patients are the gold standard comparison group. At a minimum, content validity should be strengthened by having the vignettes reviewed by clinical experts to ensure they accurately depict the situations under examination and that the question types, formats, and response options are appropriate.

Pre-Testing Vignettes. Vignettes should be pre-tested with physicians who have the same or similar characteristics as those in the target population.⁴ Pre-testing (followed by cognitive interviews) enables the researcher to ensure the instructions are easily understood and the vignette is easily interpretable. The questions and response options should be reviewed for clarity and correspondence to the vignette.⁴ Also, the vignette should yield results that have some degree of heterogeneity so differences can be detected. During pre-testing, burden on the respondent should be assessed by recording the time needed to complete the task.

Cognitive testing should assess for vignette equivalence—that is, check that all respondents identically interpret a vignette and do not make additional assumptions about the case being presented.²² For instance, the description of the symptoms of a vignette patient with heart failure should convey the same level of severity to all respondents, and the name

provided for the hypothetical patient should not lead respondents to impose their own assumptions about the patient's insurance status. Researchers should also take care to ensure their vignettes do not include excessive amounts of extraneous information that might “trick” respondents into providing responses that differ from their actual practice.⁴

Administration. The instructions that present a vignette to respondents may affect the accuracy of their responses. Researchers should be clear that the purpose of a vignette survey is not to test “textbook” answers or even to assess the performance of individual physicians, but rather, to obtain an understanding, in the aggregate, of the physician decisions that occur in practice. Offering anonymity to respondents may help elicit truthful responses, but it may also make the resulting data less useful (due to the inability to link to other data sources) and follow up with nonrespondents more difficult.⁴ Promising confidentiality and to limit analyses to only those with reasonably large sample sizes may be an effective compromise.

Validation studies have imposed time constraints on vignette responses to partially replicate the demands of a real clinical setting. The evidence suggests this constraint is important; clinicians who use vignettes as an opportunity to demonstrate their proficiency may act differently when subject to the time pressures of seeing actual patients.²⁶ Although time constraints cannot realistically be enforced in a mail survey, the time to complete the vignette can be monitored when surveys are done online. Respondents given assurances of anonymity also may feel less pressure to perform and are thus less likely to spend an inordinate amount of time completing the vignette (alternatively, respondents might spend less time than they would in an actual clinical encounter).

KNOWLEDGE GAPS

Although vignettes have been used extensively, a number of questions remain about how the methodology should best be applied. Research on general physician surveys likely offers a few valuable lessons, but additional research is warranted to provide further guidance on how a vignette survey should be constructed to measure physician decisions and the effects of system-level and practice-level factors.

Formal Validation Outside Primary Care Settings. Research has formally validated clinical vignette methodology using an open-ended instrument in primary care settings and for a few conditions (see Appendix B).^{3,5–7,27} Rigorous testing of vignettes in a number of different settings, across specialty types, and using a range of vignette designs, would add much value to the body of evidence. Also valuable would be research on the ability of vignettes to capture the influence of particular contextual factors, including financial incentives, on physician decisions at the point of care.

Number of Vignettes. Of critical importance is the number of clinical vignettes needed within a single instrument to characterize physician behavior reliably over a given dimension of care, which may be as narrow as a single clinical scenario. Although it is clear that vignette responses of individual physicians should not be interpreted as representative of practices beyond those explicitly measured by the vignette,³ the evidence offers few insights on the relationship between the accuracy of a vignette survey and the number of vignettes that are used within the instrument. While multiple vignettes are likely to be needed,⁴ including too many vignettes in one instrument may reduce the quality of responses.²⁸

Use of Closed-Ended Vignettes. As previously discussed, closed-ended vignettes may bias responses for some clinical scenarios by limiting them to a list of suggested options from which respondents may choose. More research is needed on the topic of vignettes for which the problems of a closed-ended structure emerge or are most severe. Knowing more about approaches to counter this bias would also be valuable for constructing an effective vignette survey.

Question Format. Despite multiple options for vignette question format (for example, dichotomous, multiple choice, and Likert scale), research into the implications of each has generally been limited to opinion surveys. A direct comparison of the various approaches would provide insights for future vignette surveys. Understanding which question type is ideal for a given scenario or clinical decision, for instance, would inform the development of vignette survey instruments, enhancing their validity.

CONCLUSIONS

Researchers interested in better understanding the causes of variation in physicians' clinical decisions at the point of care can choose from a range of approaches, including medical record abstraction, claims analysis, use of standardized patients, or clinical vignettes. Although clinical vignettes may not be appropriate for measuring variations in all clinical decisions, their use has a number of advantages over the currently feasible alternatives, including the ability to control for differences in case mixture, avoid challenges posed by incomplete or inaccurate patient data, and feasibly generate a large sample size. Given these potential near-term advantages and the relatively limited research into the best practices for administering a paper-based, closed-ended vignette survey, further research into vignette methodology would be worthwhile. In addition to further validation tests across specialties and settings, analyses that examine vignette question design, psychometrics, and ability to accurately capture the

influence of contextual factors on physician decisions will be particularly valuable.

Acknowledgements: This publication is derived from work supported under a contract with the Agency for Healthcare Research and Quality (AHRQ) (HHSP23320095642WC/HHSP23337033T). However, this publication has not been approved by the agency. We would like to acknowledge our technical expert panel and consultants for their feedback on an earlier version of this manuscript: Steven Haist, Eric Holmboe, John Peabody, Robert Reid, Meredith Rosenthal, Mark Schwartz, and Jane Sisk. We also are grateful to Emily Carrier, Myles Maxfield, and Paul Beatty for their comments. An earlier version of this manuscript was presented at the Collecting Data on Physicians and their Practices technical expert panel meeting on 23 July 2014.

Conflict of Interest: The authors have no conflicts of interest.

Corresponding Author: Lara Converse, MA; Mathematica Policy Research, Washington, DC, USA (e-mail: lconverse@mathematica-mpr.com).

APPENDICES

Appendix A: Selected Examples of Clinical Vignette, Medical Record Abstraction, and Claims Data Analyses

Clinical Vignettes. Finkelstein, J, Lozano P, Shulruff R, et al. Self-reported physician practices for children with asthma: Are national guidelines followed? *Pediatrics*. 2000;106(4):886–96.

Haggstrom D, Klabunde C, Smith J, Yuan G. Variation in primary care physicians' colorectal cancer screening recommendations by patient age and comorbidity. *J Gen Intern Med*. 2012;28(1):18–24.

Landon B, Reschovsky J, Reed M, Blumenthal D. Personal, organizational, and market level influences on physicians' practice patterns: Results of a national survey of primary care physicians. *Med Care*. 2001;39(8):889–905.

Williamson I, Bengt S, Moore M, Kumar S, Cross M, Little P. Acute sinusitis: Which factors do FPs believe are most diagnostic and best predict antibiotic efficacy? *J Fam Pract*. 2006;55(9):789–96.

Medical Record Abstraction. Horn D, Koplan K, Senese M, et al. The impact of cost displays on primary care physician laboratory test ordering. *J Gen Intern Med*. 2014;29(5):708–14.

McGlynn E, Asch S, Adams J, et al. The quality of health care delivered to adults in the United States. *New Engl J Med*. 2003;248(26):2635–45.

Claims Data Analysis. Rosenthal M, de Brantes F, Sinaiko A, et al. Bridges to Excellence—Recognizing high-quality care: Analysis of physician quality and resource use. *Am J Manag Care*. 2008;14(10):670–677

Weiner J, Parente S, Garnick D, et al. Variation in office-based quality: A claims-based profile of care provided to Medicare patients with diabetes. *JAMA*. 1995;273(19):1503–1508.

Appendix B: Validity and Reliability of Clinical Vignettes

Vignettes should demonstrate validity—both content validity (the scenario presented in the vignette appropriately captures the researcher's question of interest) and construct validity (for example, a vignette intending to measure physician ordering behavior should be capable of measuring this concept).²⁹ When studying physician decisions, however, the most important consideration is criterion validity—that is, when given a clinical vignette with attributes similar to those of a standardized patient (the gold standard), physicians should report the same choices in the vignette as they would make when with the standardized patient. High criterion validity suggests that a vignette accurately reflects actual physician decisions, and consequently, that the results of the clinical vignette can be reasonably interpreted as “observations” of respondents' decisions.

Before 1990, reports on the criterion validity of vignettes were scarce. Jones et al. (1990) attempted to determine the capacity of written case simulations to predict actual clinical behavior, as measured by standardized patients or expert chart review. In a literature review of articles using written vignettes, only a small subset of their 74 identified articles (15 %) contained an evaluation of the vignettes' criterion validity, and among these, only two studies used a design that permitted complete assessment of criterion validity.³⁰

Since the 1990s, however, a small number of additional studies have explicitly examined vignettes' ability to measure actual physician decisions accurately (although validity in these cases did not necessarily indicate the presence of evidence-based care). In a study with a small sample size, Sandvik (1995) assessed the criterion validity of open-ended and closed-ended case vignettes focused on the treatment of urinary incontinence. When vignettes featured checklists for responses, physicians claimed more actions than they had actually performed on similar patients they had seen in the recent past. When offered the opportunity to provide an open-ended response, respondents showed no difference in their claimed actions versus their actual actions.²⁷

More recently, Peabody et al. (2004) sought to determine if open-ended, computerized clinical vignettes could accurately measure physicians' outpatient care delivery with standardized patients and to analyze how well vignettes performed compared with medical chart abstraction. Criterion validity was higher with vignettes than with medical chart abstraction. During treatment of standardized patients, physicians met 73 % of the a priori determined criteria for quality of care. These same physicians met 68 % of the criteria when completing the comparable clinical vignette and 63 % with medical chart abstraction. The findings were similar regardless of disease condition (depression, chronic obstructive pulmonary disease, diabetes, or vascular disease), patient complexity, level of physician training, or type of health care system, and the authors concluded that vignettes are a valid assessment of

clinical quality.⁷ Moreover, in a 2000 study, Peabody et al. concluded that “vignettes appear to be a valid and comprehensive method that directly focuses on the process of care provided in actual clinical practice.”³

A comparably structured pair of studies also validated the use of clinical vignettes for measuring physician decisions to offer preventive care to patients making their first visits to a primary care clinic. Although discrepancies between the two varied by type of preventive care decision, overall performance on clinical vignettes compared favorably to performance as measured by standardized patients.^{5,6}

The close correspondence in these studies between the patients portrayed in the vignettes and the standardized patients who actually consulted the doctors suggests that, at least in some instances, vignettes have strong criterion validity. Whether these results can be generalized across specialties and a large number of clinical scenarios is unknown, however.

Although Peabody et al. (2004),⁷ Dresselhaus et al. (2004),⁶ and similar studies have validated clinical vignettes for measuring quality of care for particular chronic conditions and preventive services, caution in generalizing these results to all vignettes across all types of point-of-care decisions is important. As a case in point, Mohan et al. (2014) found that case vignettes did not perform well in predicting emergency medicine clinicians' actual decisions in triaging patients in the emergency department.³¹ The discrepancy may be explained by differences in practice setting and patient acuity and/or the use of medical chart abstraction rather than standardized patient reports as the standard for comparison.

REFERENCES

1. **Institute of Medicine.** Crossing the quality chasm: A new health system for the 21st Century. Washington, D.C.: National Academy Press; 2001.
2. **McGlynn E, Asch S, Adams J, et al.** The quality of health care delivered to adults in the United States. *New Engl J Med.* 2003;248(26):2635–2645.
3. **Peabody J, Luck J, Glassman P, Dresselhaus TR, Lee M.** Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *JAMA.* 2000;283(13):1715–1722.
4. **Veloski J, Tai S, Evans A, Nash D.** Clinical vignette-based surveys: A tool for assessing physician practice variation. *Am J Med Qual.* 2005;20(3):151–157.
5. **Dresselhaus T, Peabody J, Lee M, Wang MM, Luck J.** Measuring compliance with preventive care guidelines: Standardized patients, clinical vignettes, and the medical record. *J Gen Intern Med.* 2000;15(11):782–788.
6. **Dresselhaus T, Peabody J, Luck J, Bertenthal D.** An evaluation of vignettes for predicting variation in the quality of preventive care. *J Gen Intern Med.* 2004;19(10):1013–1018.
7. **Peabody J, Luck J, Glassman P, et al.** Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Ann Intern Med.* 2004;141(10):771–780.
8. **Dresselhaus TR, Luck J, Peabody JW.** The ethical problem of false positives: A comparison of standardized patients and the medical record. *J Med Ethics.* 2002;28(5):291–294.
9. **Pham T, Roy C, Mariette X, Lioté F, Durieux P, Ravaud P.** Effect of response format for clinical vignettes on reporting quality of physician practice. *BMC Health Serv Res.* 2009;9(128).
10. **Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P.** Assessing the accuracy of administrative data in health information systems. *Med Care.* 2004;42(11):1066–1072.

11. **Song Z, Safran D, Landon B, et al.** The 'Alternative Quality Contract' in Massachusetts, based on global budgets, lowered medical spending and improved quality. *Health Aff.* 2012;31(8):1885–1894.
12. **Luck J, Peabody J.** Using standardised patients to measure physicians' practice: Validation study using audio recordings. *BMJ.* 2002;325(7366):679–682.
13. **Lutfey KE, Campbell SM, Renfrew MR, Marceau LD, Roland M, McKinlay JB.** How are patient characteristics relevant for physicians' clinical decision making in diabetes? An analysis of qualitative results from a cross-national factorial experiment. *Soc Sci Med.* 2008;67(8):1391–1399.
14. **Accreditation Council for Continuing Medical Education.** Supporting maintenance of certification with CME. Available at: <http://www.accme.org/education-and-support/video/interview/supporting-maintenance-certification-cme>. Accessed March 6, 2014.
15. **Kadivar H, Goff BA, Phillips WR, Andrilla CH, Berg AO, Baldwin LM.** Guideline-inconsistent breast cancer screening for women over age 50: A vignette-based survey. *J Gen Intern Med.* 2014;29(1):82–89.
16. **Landon B, Reschovsky J, Reed M, Blumenthal D.** Personal, organizational, and market level influences on physicians' practice patterns: Results of a national survey of primary care physicians. *Med Care.* 2001;39(8):889–905.
17. **Green A, Carney D, Pallin D, et al.** Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *J Gen Intern Med.* 2007;22(9):1231–1238.
18. **Mosca L, Linfante A, Benjamin E, et al.** National study of physician awareness and adherence to cardiovascular disease prevention guidelines. *Circulation.* 2005;111(4):499–510.
19. **Bradburn N, Sudman S, Wansink B.** Asking questions: The definitive guide to questionnaire design—for market research, political polls, and social and health questionnaires. San Francisco: Jossey-Bass; 2004:117–145.
20. **Dillman D, Smyth J, Christian L.** Internet, mail, and mixed-mode surveys: The tailored design approach. Hoboken: Wiley; 2009:65–68.
21. **Krosnick J.** Response strategies for coping with the cognitive demands of attitude measures in surveys. *App Cognitive Psych.* 1991;5(3):213–236.
22. **Jurges H, Winter J.** Are anchoring vignettes' ratings sensitive to vignette age and sex? *Health Econ.* 2011;22(1):1–13.
23. **American Board of Internal Medicine Foundation.** Choosing wisely partners' announcement press release. Available at: <http://www.choosing-wisely.org/choosing-wisely-partners-announcement-press-release-december-14-2011/>. Accessed July 3, 2014.
24. **Federation of State Medical Boards and National Board of Medical Examiners.** United States Medical Licensing Examination. Available at: <http://www.usmle.org/>. Accessed February 9, 2015.
25. **Veloski J, Rabinowitz H, Robeson M, Young PR.** Patients don't present with five choices: An alternative to multiple-choice tests in assessing physicians' competence. *Acad Med.* 1999;74(5):539–546.
26. **Rethans J, van Boven C.** Simulated patients in general practice: A different look at the consultation. *BMJ.* 1987;294(6575):809–812.
27. **Sandvik H.** Criterion validity of responses to patient vignettes: An analysis based on management of female urinary incontinence. *Fam Med.* 1995;27(6):388–392.
28. **Mohan D, Rosengart MR, Farris C, Fischhoff B, Angus DC, Barnato AE.** Sources of non-compliance with clinical practice guidelines in trauma triage: a decision science study. *Implement Sci.* 2012;7(103).
29. **Cronbach LJ, Meehl PE.** Construct validity in psychological tests. *Psychological Bulletin.* 1955;52:281–302.
30. **Jones T, Gerrity M, Earp J.** Written case simulations: do they predict physicians' behavior? *J Clin Epidemiol.* 1990;43(8):805–815.
31. **Mohan D, Fischhoff B, Farris C, et al.** Validating a vignette-based instrument to study physician decision making in trauma triage. *Med Decis Making.* 2014;24(2):242–252.