Methods for improving efficiency in quality measurement: the example of pain screening

T.G.K. BENTLEY[1,2,3], J. MALIN[1,2,3], S. LONGINO[1], S. ASCH[1,2,3], S. DY[4] AND K.A. LORENZ[1]

[1]Veteran's Health Administration of Greater Los Angeles, Los Angeles, CA, USA, [2]RAND Corporation, Santa Monica, CA, USA, [3]University of California, Los Angeles, Los Angeles, CA, USA, and [4]Johns Hopkins School of Public Health, Baltimore, MD, USA

Address reprint requests to: Tanya G.K. Bentley, Partnership for Health Analytic Research, LLC, Beverly Hills, CA 90212, USA. Tel: +1-310-858-9555; Fax: +1-310-858-9552; E-mail: tbentley@pharllc.com

Abstract

Objective. Collecting unnecessary data when assessing quality of care wastes valuable resources. We evaluated three approaches for estimating quality-measure adherence and determined minimum visit data required to achieve accurate estimates.

Design. We abstracted medical records for calculating physician-level pain screening rates as: visit-specific, using single-visit data for each patient; visit-level average, using data for all patients and visits; and patient-level average, using data from a subset of patients and visits.

Setting. VA Greater Los Angeles Health-care System, 2006.

Participants. One hundred and six patients with Stage IV solid tumors.

Intervention. Pain screening at every medical encounter, measured by a 0–10 numeric rating scale and reported to the national Medicare insurance program under a 'pay-for-reporting' program.

Main Outcome Measures. Amount of visit data needed to reach the smallest 95% confidence interval (CI) and stable pain screening estimates.

Results. Pain screening occurred at 22% (23/106; 95% CI: 14–30%) of initial visits and 50% (8/16; 95% CI: 25–75%) of single visits. Across all visits, screening adherence averaged 34% when estimated at the visit-level precision and 30% at the patient level. Maximum patient-level precision was reached at visit 4 (95% CI: ±8%) and visit level at visit 14 (95% CI: ±6%). Using patient-level and visit-level approaches, estimates stabilized at visits 8 and 11, respectively, and reached within 1 percentage point of the steady-state value at visits 4 and 9.

Conclusion. To address low-pain screening among cancer patients, an oncology pain screening measure may be most efficiently evaluated with data from a sample of patients and visits. This approach may be valid for visit-level quality measures in other settings.

Keywords: quality of care, quality assessment, pain screening, methodology, efficiency

Introduction

As many health-care systems strive to improve the quality of care while controlling costs, there is an increasing international focus on paying health-care providers for their performance on indicators of quality care [1–3]. The USA is moving rapidly toward integrating quality measurement routinely in health-care provision as an important feature of recent health reform. In 2006, the Centers for Medicare and Medicaid Services established the Physician Quality Reporting Initiative (PQRI) [4, 5] to encourage physician practices to report on the quality of care provided to Medicare beneficiaries. Physicians participating in PQRI must report results of at least three quality measures for at least 80% of eligible encounters in order to receive incentive payments equaling 2.0% of their allowed Medicare Physician charges. In the USA, under PQRI, physicians are rewarded for having reported, without any expectation, that they achieved specific clinical outcome targets.

Pay for performance initiatives require substantial efforts, especially for data collection. Performance measures that evaluate the process of care, often described as quality indicators, specify an eligible patient population and then the care that these patients should receive. The metric is a ratio where the denominator identifies the patient population to whom the care should be provided and the numerator describes the care that should be provided. For an eligible patient, the indicator could have a value of 1 ('pass') or 0 ('not pass'); for the population, the result would be a proportion, with a range of values from 0 to 1. While most quality indicators have addressed the care for which a patient is eligible, only once during the measurement interval (e.g. the prior year), some important aspects of care need to be addressed more often and thus a given patient may be eligible more than once. In the USA, for example, of the 153 quality measures included in PQRI, 8 measures apply to every patient visit. Assessment of pain in patients with cancer is one such measure:

> Percentage of patient visits [*provided to Medicare beneficiaries*], regardless of patient age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy in which pain intensity is quantified. [4]

Physicians choosing to use this measure to satisfy their PQRI participation can use claims or registry data to report pain screening at every visit [4, 5], calculated for a 6- or 12-month period as the proportion of eligible patient visits in which pain screening occurred:

$$\text{PQRI\_pain screening rate}$$
$$= \frac{\#\text{visits for which the pain level quantified}}{\#\text{visits for cancer patient during reporting interval}}$$

Reducing the reporting burden for quality assessment is an important goal. On the one hand, it could improve physicians' acceptance and subsequent utilization of visit-level quality measures [6]. In addition, minimizing the time and effort required for quality assessment might allow more time and effort for quality improvement. We therefore evaluated three approaches for pain screening performance measurement. We evaluated whether or not pain screening had occurred:
(1) when considering a sample of eligible visits (one single visit per eligible patient),
(2) when considering all eligible visits, and
(3) when averaging the pain screening score among all eligible patients.
The three measurement approaches have different implications for the effort required to report, and there is a trade-off between the reporting effort and the precision that can be obtained. Our objective was to compare the three measurement approaches to learn which approach would produce the greatest precision [the smallest confidence interval (CI)] for the smallest number of visits given a fixed sample size, and to determine the point at which adding data from additional visits would not modify the quality score

estimate. Specifically, we sought the measurement approach that had the smallest CI with the smallest number of visits.

Methods

Patient cohort

We identified all patients from the VA Greater Los Angeles Health-care System (VAGLAHS) with newly diagnosed Stage IV solid tumors in 2006 [7–10]. We included patients with at least 1 month of survival from the date of diagnosis and at least one or more outpatient cancer-related physician visits during a 3-month study period. The study was approved by the VAGLAHS IRB [10].

Data collection

Data elements for the quality measure were abstracted from patients' computerized medical records by trained nurse abstractors. We included all visits received by patients from cancer specialists, palliative care clinicians or primary care providers during the 3 months after initiating cancer treatment (surgery, chemotherapy or radiation therapy). Abstractors determined for every visit whether physicians documented the 'fifth vital sign', a quantitative pain assessment tool incorporating a 0–10 numeric rating scale. Inter-rater reliability—a measure of internal validity—for pain scores was excellent (kappa = 0.81) [11].

Analyses

We calculated the rate of performance of pain screening using three methods: (i) sample of eligible visits ('visit-specific rate'); (ii) all eligible visits ('cumulative visit rate'); and (iii) average score for eligible patients ('patient average rate'). While all three methods measure occurrence of pain screening, they have different measurement requirements and provide different perspectives on the quality of provided care. The visit-specific pain-screening rate can be estimated using data for a single visit for each eligible patient during the reporting period and provides an estimate that would be obtained through a random cross section of patient visits. The cumulative visit and patient average rates both require data for all patients over a specified number of physician visits during the time interval of interest, but use the data differently to calculate average scores with different weightings of opportunities for providing care. Specifically, in the cumulative visit rate, each visit is treated as a discrete, independent event and patients with more visits contribute more to the quality measure score, such an approach values higher screening rates overall, regardless of how that screening is distributed across patients. In contrast, the patient average rate gives all patients equal weight regardless of their number of visits, so a single pain screening for a patient with few visits has more impact on the quality measure score than would a single screening for a patient with many visits; this approach values consistency across patients over consistency across visits.

Figure 1 provides an overview of the calculation methods for each of these three approaches. Visit-specific rates were calculated as the proportion of patients screened for each consecutive visit (visit = 1, 2, ..., n) during the study period, such that the rate for visit 1, for example, equaled the number of patients with screening at their first visit divided by the number of patients with at least one visit. Cumulative visit rates were calculated as *cumulative* proportions of patients screened on consecutive 1st–nth visits. For example, to estimate the cumulative visit rate for the third visit, we averaged three visit-specific rates—those for visits 1, 2 and 3; for that of the nth visit, we averaged rates for all visits, as done in the PQRI measure. Finally, we calculated patient averages—our third approach—in two steps: first, we measured individual patients' cumulative rates, and then we averaged those patient-specific estimates over all patients. Patients who had fewer visits than the target visit number for that calculation had data for all available visits included (e.g. if the estimate were based upon a maximum of four visits and a patient only had three visits during the study period, only data from these three visits were included).

Box 1. Calculations for three approaches used to estimate pain screening rates

Visit-specific rates were calculated as:

$$\frac{\text{Among all physicians, \# patients screened at each patient's n}th\text{ visit}}{\text{Among all physicians, \# patients with 'n' \# of visits}}$$

Cumulative visit rates were calculated as:

$$\frac{\text{Among all physicians, total \# visits at which pain screening occurred}}{\text{Among all physicians, total \# visits}}$$

Patient average rates were calculated in two steps:
(1) Calculate patient-specific visit cumulative rates as:

$$\frac{\text{Among all physicians, \# visits at which pain screening occurred for each patient}}{\text{Among all physicians, \# eligible physician visits for each patient}}$$

(2) Calculate the average of patient-specific cumulative rates across all patients.

We estimated screening rates with these three approaches for the 106 eligible patients and up to 20 visits per patient. We calculated 95% CIs and identified the lowest visit number (e.g. numbered from each patient's first to last visit, up to the 20th visit) for which the smallest CI was achieved.

We also determined the number of visits required for the point estimate to equal the final estimate—labeled here as 'steady state'—as well as the visit at which the point estimate was within 1% of the final estimate of the pain screening rate.

Results

Table 1 describes the characteristics of the study population ($n = 106$), which consisted of mostly white, older (>65 years of age), unmarried men with Stage IV cancer and treated with chemotherapy and/or radiation therapy. Figure 2 shows the distribution of study patients receiving 1–20 outpatient physician visits during the 3-month study period. Results showed that 75% of patients received 4 visits within 3 months, 46% received 8 and 29% received 11.

Using chart-abstracted documentation of the 'fifth vital sign' as evidence of pain screening, we estimated that 77% of patients (82/106) were screened for pain during at least one outpatient physician visit during the study period. Figure 3 shows screening rates and 95% CIs for visits 1–20 for the three calculation methods: visit-specific (panel a); visit-level cumulative average (panel b); and patient-level cumulative average (panel c). The Appendix reports the number of eligible physician visits and patients, the number of physician visits with pain screening, and method-specific pass rates and 95% CIs.

The rate of screening for pain at patients' first visits was 22% (95% CI: 13–30%; $n = 106$). Across the first 5 visits, the visit-specific rate ranged from 22 to 38% and across visits 6 through 20, as fewer and fewer patients had more visits, this rate ranged from 13 to 50%, with declining precision as the visit number increased and fewer patients remained in the analysis. Visit cumulative average and patient average rates were similar in their final estimates of 30 and 34%, respectively, although they differed in the number of visits required before estimates reached maximum precision or steady state. Maximum precision of $\pm 6\%$ for visit cumulative average rate was reached at visit 14, while that for the patient average rate was $\pm 8\%$ at visit 4. The steady-state visit cumulative average rate of 34% was reached at visit 11, whereas that for the patient-average rate (30%) was reached at visit 8. Estimates reached within 1 percentage point of steady-state value at visits 9 and 4 for visit cumulative and patient average approaches, respectively.

Discussion

The goal of this study was to identify the most efficient sampling strategy for performance measures that assess care that is supposed to be provided at every patient visit, in this case screening for pain in patients with cancer. Estimates of pain screening adherence depended on the calculation approach and number of visits considered for
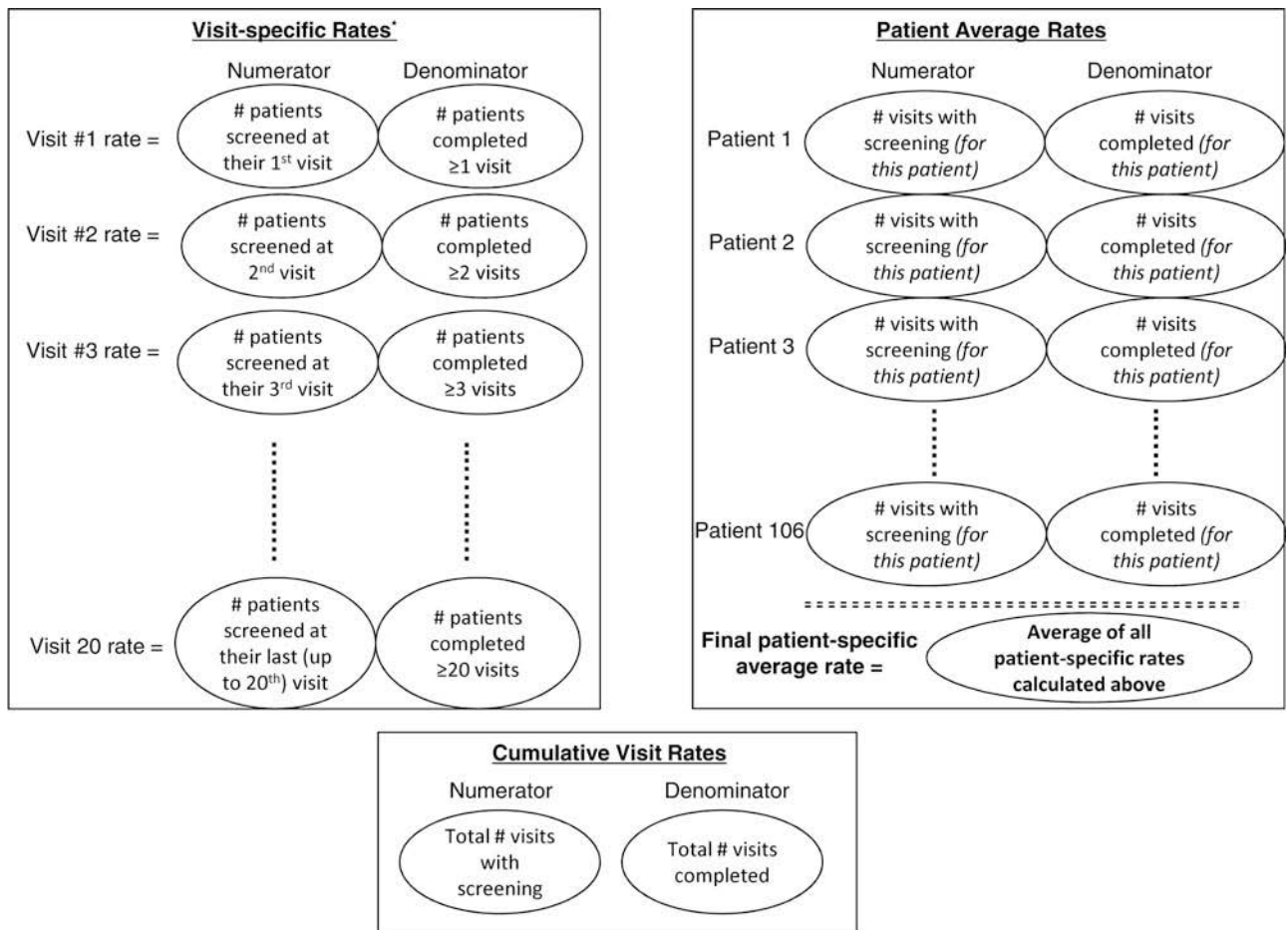
Figure 1 Overview of the calculation of the visit-specific rate, cumulative visit rate and patient-average rate for measuring physician adherence to pain screening.

each approach, with rates ranging from 22 to 34% of specific visits and averaging between 30 and 34% of all visits. Using a visit-specific approach and measuring at only the first visit—a method used by the American Society of Clinical Oncology's commonly used measure [12]—screening adherence was 22% (95% CI: ± 8%). When evaluating across all visits with visit-level or patient-level cumulative average approaches, estimates were similar but still low. Those estimated at the patient level were most precise and stable with fewer data points required than for visit-level averages. With the majority of patients failing to receive consistent pain screening, our results suggest that there is substantial room for improvement in evaluating cancer pain, even in a system with a long-standing pain screening policy [13]. In addition, oncology pain screening in a performance measurement program may be most efficiently evaluated by collecting data from a sample of patients—rather than from an entire patient cohort—and from four consecutive visits per patient—

rather than from all visits had by each patient—during a specified reporting period.

Alternative reporting mechanisms are possible. For example, instead of reporting performance data directly on claims for 80% of eligible patients, physicians may choose to use a certified registry and report on a sample of 30 consecutive patients. By limiting the sample size, this approach is likely to reduce the measuring and reporting burden, although the smaller samples may impact the reliability of results. Currently, data accuracy has not been of concern because participating physicians receive incentive payments regardless of their scores, as long as they successfully report the data. However, if in the future physicians receive incentive payments only when meeting or exceeding a specified quality-of-care benchmark, measurement accuracy will be paramount. Reliability and efficiency will likewise be of concern if individual physician-practice scores become publicly available. Our approach of determining minimum visits necessary for accurate pain screening evaluation could thus be useful to focus data collection and balance reliability and efficiency.

Table 1 Study patient characteristics

Total patients (*n*)	106
Age [mean (SD) years]	66 (10)
Gender (% male)	98
Race (% white)	62
Marriage (% married)	25
Cancer type (%)	
Head and neck	23
Lung/bronchus	23
Prostate gland	17
Colon	9
Oesophagus	6
Pancreas	5
Other[a]	17
Stage at diagnosis (%)	
Stage IV with metastases	88
Treatment received (%)	
Chemotherapy and radiation therapy	35
Chemotherapy alone	12
Radiation therapy alone	17
History of (%)	
Mental illness	29
Depression	15
Post-traumatic stress disorder	10
Schizophrenia/psychosis	3
Dementia	2
Substance abuse or alcoholism	20
Serious medical comorbidities	19
Median survival (months)	12
Median follow-up (months)	25
Outpatient physician visits	
# patients with outpatient physician visits	106
Mean (SD; max) # physician visits per eligible patient	10 (7; 32)
Total # outpatient physician visits	982

[a]Bladder, rectum, liver/intrahepatic biliary tract, stomach, kidney, breast, gallbladder, ureter.
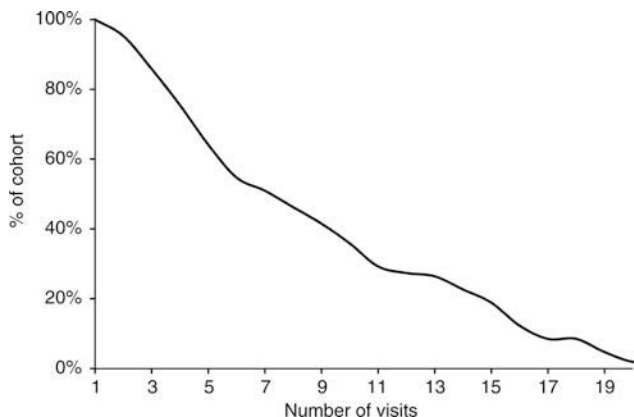

Figure 2 Number of outpatient physician visits achieved by study patients within 3 months of the first physician visit.

The approach is likely also valid for other visit-level quality measures, but further research is needed to characterize the most efficient measurement approaches for different visit-level measures.

Our results should be viewed in light of several limitations. First, our cohort included veterans newly diagnosed with Stage IV cancer. To the extent that the reproducibility of our results depends on factors such as distribution of visits, reasons for cohort attrition (e.g. death vs. seeking out a new provider due to poor pain management), the phase of cancer illness or other patient characteristics, our results may not be generalizable. Second, the VA has an integrated electronic medical record; and although we used a data abstraction approach that has also been previously used with traditional paper medical records, there may be differences in documentation that could limit the generalizability of our results.

While the results of this analysis can guide policy-makers and clinicians in developing and evaluating pain screening quality measures, they also raise questions relevant to informing quality improvement. For example, do physicians find visit- or patient-level feedback most helpful for assessing and targeting improvement? If physicians need only report screening for patients' initial visits, would practice patterns change to reflect reporting requirements rather than true patient need, with quality deteriorating at subsequent visits? And will policy-makers need to consider additional incentives, practice tools or structural practice changes to facilitate implementation of routine pain screening? [6]

Alternative approaches to quality measurement need not be limited by data collected over only a short time period, although system unresponsiveness could limit the value of shorter assessment cycles. For example, Queensland Health uses statistical control techniques to evaluate frequent events and deviation from usual performance [3], allowing for continuous triggers of improvement based on deviation from expected norms. Although such an approach could be valuable when applied to pain management (e.g. by evaluating deviations from usual pain levels achieved), it would only be successful if the time required to respond to a complaint of severe pain were brief. Even very frequent assessment of quality care measures will be of no value in health systems that have not actuated a tightly linked quality improvement arm.

In summary, we identified an efficient quality assessment strategy for pain screening among cancer patients, demonstrating that quality assessment may be optimally evaluated for a consecutive patient sample over a limited number of physician visits. Our results suggest that there is substantial room for improvement in cancer pain screening, and that while programs like PQRI can help to establish standards for pain assessment and documentation to ensure that pain is recognized and treated promptly, measurement will be challenging unless practices standardize routine pain assessment and documentation. Because optimizing data-collection efficiency lowers cost and improves quality improvement feasibility, the approach we have used should be considered and further evaluated across a broad range of settings and measures.
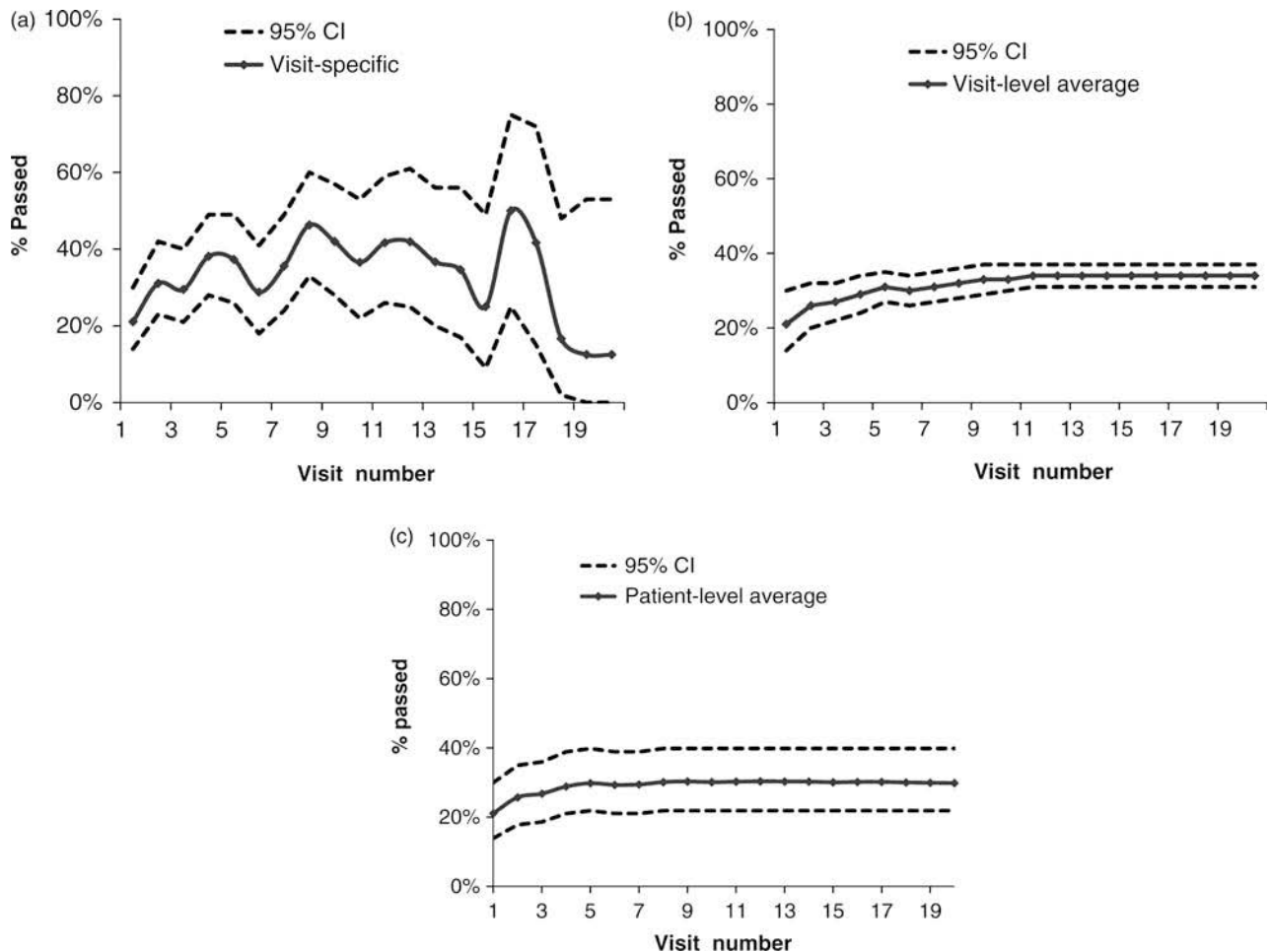
Figure 3 Outpatient pain screening rates among cancer patients: visit-specific rate (**a**), cumulative visit average rate (**b**) and patient average rate (**c**). *The significant increase in variability in pain screening rates across visits for the visit-specific rate largely reflects the decreasing sample size as declining precision as the visit number increased and fewer patients remained in the analysis (see Appendix for details).

Funding

References

1. Epstein Arnold M. Paying for performance in the United States and abroad. *N Engl J Med* 2006;**355**:406−8.

2. Doran T, Fullwood C, Gravelle H *et al*. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;**355**:375−84.

3. Duckett SJ, Collins J, Walker K. An improvement focus in public reporting: the Queensland approach. *Med J Aust* 2008;**189**:616−7.

4. Centers for Medicare and Medicaid Services. *2009 PQRI Measures List*. Baltimore, MD: Department of Health and Human Services, 2009.

5. Centers for Medicare and Medicaid Services. *2009 Physician Quality Reporting Initiative Implementation Guide*. Baltimore, MD: Department of Health and Human Services, 2009.

6. Cleeland CS, Reyes-Gibby CC, Schall M *et al*. Rapid improvement in pain management: the Veterans Health Administration and the institute for healthcare improvement collaborative. *Clin J Pain* 2003;**19**:298−305.

7. Dy S, Asch S, Naeim A *et al*. Evidence-based standards for cancer pain management. *J Clin Oncol* 2008;**26**:3879−85.

8. Dy S, Lorenz K, Naeim A *et al*. Evidence-based recommendations for cancer fatigue, anorexia, depression, and dyspnea. *J Clin Oncol* 2008;**26**:3886−95.

9. Lorenz K, Dy S, Naeim A *et al*. Quality measures for supportive cancer care: the cancer quality-ASSIST Project. *J Pain Symptom Manage* 2009;Epub.

10. Dy SM, Lorenz K, O'Neill SM *et al*. Cancer quality-ASSIST supportive oncology quality indicator set: feasibility, reliability, and validity testing. *Cancer* 2010;**116**:3267–75.

11. Landis J, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.

12. Jacobson JO, Neuss MN, McNiff KK *et al*. Improvement in oncology practice performance through voluntary participation in the Quality Oncology Practice Initiative. *J Clin Oncol* 2008;**26**:1893–8.

13. de Rond M, de Wit R, van Dam F. The implementation of a Pain Monitoring Programme for nurses in daily clinical practice: results of a follow-up study in five hospitals. *J Adv Nurs* 2001;**35**:590–8.

Appendix

Outpatient pain screening rates with three methods: visit-specific, cumulative visit and patient-level cumulative average[a]

Visit #	Visit-specific pain screening		# eligible (# screened)	Cumulative visit: % screened (95% CI)	Patient average: % screened (95% CI)
	# eligible (# screened)	% screened (95% CI)			
1	106 (23)	22 (13, 30)	106 (23)	22 (13, 30)[b]	22 (13, 30)
2	102 (32)	31 (23, 42)	208 (55)	26 (20, 32)[b]	26 (18, 36)
3	95 (28)	29 (21, 40)	303 (83)	27 (22, 32)[b]	27 (19, 37)
4	84 (32)	38 (28, 49)	387 (115)	29 (24, 34)	29 (19, 39)
5	75 (28)	37 (26, 49)	462 (143)	31 (27, 35)	30 (21, 42)
6	66 (19)	29 (18, 41)	528 (162)	30 (26, 34)	29 (18, 41)
7	59 (21)	36 (24, 49)	587 (183)	31 (27, 35)	29 (18, 42)
8	54 (25)	46 (33, 60)	641 (208)	32 (28, 36)	30 (18, 44)
9	50 (21)	42 (28, 57)	691 (229)	33 (29, 37)	30 (18, 45)
10	41 (15)	37 (22, 53)	732 (244)	33 (30, 37)	30 (16, 46)
11	36 (15)	42 (26, 59)	768 (259)	34 (31, 37)	30 (16, 48)
12	31 (13)	42 (25, 61)	799 (272)	34 (31, 37)	30 (14, 48)
13	30 (11)	37 (20, 56)	829 (283)	34 (31, 37)	30 (15, 49)
14	26 (9)	35 (17, 56)	855 (292)	34 (31, 37)	30 (14, 52)
15	20 (5)	25 (9, 49)	875 (297)	34 (31, 37)	30 (12, 54)
16	16 (8)	50 (25, 75)	891 (305)	34 (31, 37)	30 (11, 59)
17	12 (5)	42 (15, 72)	903 (310)	34 (31, 37)	30 (10, 65)
18	12 (2)	17 (2, 48)	915 (312)	34 (31, 37)	30 (10, 65)
19	8 (1)	13 (0, 53)	923 (313)	34 (31, 37)	30 (3, 65)
20	8 (1)	13 (0, 53)	931 (314)	34 (31, 37)	30 (3, 65)

[a]% Screened = # passed/# eligible per indicator, per physician visit.
[b]$P < 0.05$ in z-test of proportions comparing the pass rate for physician visit × to that of the overall cumulative pass rate for all physician visits.