SPOkes in the wheel: Structure, Process, and Outcomes of healthcare. An examination of the quality of the relationships among indicators of hospital and general practitioner performance

Ogbu, U.C.

Publication date
2010

Chapter 5

Process to Process correlation: The interpretive validity of indicators

Uzor C. Ogbu
Onyebuchi A. Arah
Liset van Dijk
Dinny H. DeBakker
Karien Stronks
Gert P. Westert

Submitted

ABSTRACT

Objective: To determine the empirical relationship between theoretically related process indicators using antibiotic prescribing guideline adherence among general practices.

Data Source: One hundred and eighteen general practices participating in the Netherlands Information Network of General Practice from 2002 to 2005.

Study design: In this retrospective study, we used Pearson's correlation coefficient to examine the pairwise practice-level association between guideline adherence rates for episodes of bacterial skin infection (30,757), acute throat pain (28,544), sinusitis (39,648), and urinary tract infections (75,300). We used three-level multivariate multilevel analysis to study the association between practice-level adherence rates of any three adherence indicators, and likelihood to adhere to the fourth guideline.

Principal findings: The correlation between performance levels of the prescribing indicators ranged from negative (-0.29) to positive (0.57). The odds ratios and 95% CI from the multilevel analyses ranged from a minimum of 0.97 (0.97 – 0.98) to a maximum of 1.02 (1.02 – 1.04) for the various practice-level adherence measures.

Conclusions: The relatively weak relationships between the four process indicators indicate that they may not be measuring the same underlying construct. These findings raise questions about the interpretation of performance results based on existing quality indicators.

INTRODUCTION

Quality of care is measured using frameworks consisting of various performance indicators.[1-3] In an environment where payment is becoming value based and performance results become more accessible to the public, performance indicators will come under greater scrutiny. Central to the successful interpretation of organizational or physician performance in healthcare is an understanding of the link between the measured indicator and what they represent.[4-6] As a result, studies examining the associations of structure and process measures with outcomes have become frequent in the quality literature.[7-13] However, associations between process indicators are rarely empirically examined.

Process indicators are said to "reflect what the provider did for the patient and how well it was done".[14] Within the context of the environment in which they are measured process indicators reflect behavior or attitudes, thus forming amenable targets for quality improvement initiatives. In some performance frameworks, a number of theoretically related process indicators are used to measure the same provider trait. The indicators are selected using a variety of procedures based on achieving a consensus, taking into account existing best practice and indicator attributes such as reliability, validity, and interpretation.[14] If a group of process indicators measure the same underlying construct, performance levels on the individual indicators should display a reasonable degree of correlation with each other. In the absence of such a correlation, conclusions regarding the underlying quality construct they are assumed to measure might be invalid.

The quality of a general practitioners (GPs) prescribing in the Netherlands is measured using process indicators related to antibiotic prescribing guidelines. In this study, we examine the relationship between four antibiotic prescribing guideline adherence rates. Using data from the Netherlands Information Network of General Practice (LINH), we try to determine if practice-level performance is correlated, and if the four indicators can be taken as a measure of the same underlying trait related to a GPs prescribing quality.

METHODS

Netherlands Information Network of General Practice (LINH)

LINH was first established in 1992 as a referral database. It has since expanded to become an electronic data source for diagnosis, morbidity, prescribing, and referral information on 340,000 patients and 180 GPs across the Netherlands.[15] Patients in LINH are representative of the Dutch population with respect to age and gender. LINH general practices are representative for all Dutch GPs, except for solo practices being slightly underrepresented.[16] Participating practices have electronic medical records coupled with a decision support systems.[17] Diagnoses are coded according to International

Classification of Primary Care (ICPC) standards, and prescriptions recorded using Anatomical Therapeutic Chemical (ATC) codes.[18,19]

LINH was the primary source of clinical data for the Dutch National Survey of General practice (DNSGP-2).[20] This survey, carried out in 2001, evaluated the quality of GP care in the Netherlands using national guidelines developed by the Dutch College of General Practitioners (NHG). Our study uses diagnosis and prescribing information from LINH for the years 2002 to 2005, focusing on adult patients, thus excluding patients younger than 18 years old.

Prescribing Guidelines/Indicators

NHG used an evidence- and consensus-based approach to develop 70 guidelines. Based on the guidelines, 139 indicators were formulated to measure performance, and were included in DNSGP-2. These indicators cover a wide range of conditions and include structure, process, and outcome measures. This study uses four of the process indicators related to antibiotic prescribing for bacterial skin infections (ICPC S76, S84, S96), acute throat pain (ICPC R21, R72, R76), sinusitis (ICPC R75), and urinary tract infections (ICPC U71). These are common conditions used as prescribing indicators in DNSGP-2 and accounted for the bulk of antibiotic prescribing in the Netherlands.[21]

The guidelines advocate the restrictive prescribing of antibiotics when the diagnosis is a bacterial skin infection, acute throat pain, or sinusitis.[22] The restrictions are in place to prevent the development of antibiotic resistant strains of bacteria. The guideline for urinary tract infections (UTI) advocates nitrofurantoin, or trimethoprim as a first choice drug.[23]

Study variables

Outcome

Adherence to guidelines, described as a dichotomous variable for each disease episode. For the restrictive guidelines, bacterial skin infections, acute throat pain, and sinusitis, this meant not having an antibiotic prescribed. For the first-choice guideline, urinary tract infections, this meant an initial prescription of nitrofurantoin or trimethoprim.

Guideline adherence rates for practices were determined using the number of disease episodes in which the guideline was adhered to as the numerator and the practices total number of disease episodes as the denominator. This was expressed as a percentage. Adherence rates were estimated each year for every practice reporting in that year. A disease episode was defined as starting on the day of the first appearance of the diagnosis in the GP records and counting a predetermined number of days onwards. During the intervening period, all consultations and prescriptions were attributed to that episode although the GPs actions during the first consultation were used to assess adherence. Episodes were defined as lasting for a 14-day period for bacterial skin infections, a 10-day period for acute throat pain, a 21-day period for sinusitis, and a 10-day period for urinary tract infections. The selection of these durations was based on the approximate duration of antibiotic course/illness.

Patient-level covariates

We included patient characteristics that may affect the prescribing decision. We included the age of the patient at the time of consultation, gender, type of insurance (public or private) – as a proxy for socioeconomic status, and episode number. The episode number reflects a running sequential count of the number of disease episodes for which a patient visited the same GP with the same diagnosis over the four-year study period. In the Dutch health system, GPs act as gatekeepers and maintain geographically determined lists. Thus, each resident is on the list of one GP in their neighborhood who attends to all their health needs.

Practice-level covariates

We included the number of full time equivalents (FTE) as a proxy for practice workload, type of practice – single, duo, group practice or health center, and year of consultation as a categorical variable. Practice size and type of practice are related to workload and amount of administration required which have been identified factors in prescribing decisions. The variable FTE represents the number of full-time hours put in by employees of the practice. In using it as a proxy for practice workload, we take into account the patient–population ratio and the likelihood that practices with a high daily patient load would require additional staff to handle administration. We also included the practices adherence rate per indicator, calculated as stated above.

Statistical analysis

We pooled the data across the four years, and carried out univariate analysis showing summary statistics for the total population by indicator. Using Pearson's correlation coefficient, we carried out pairwise correlations of the average performance of each practice over the four years for all four indicators.

The data have a three-level hierarchical structure (disease episodes nested in patients and patients nested in practices), which may violate the assumption of independence of outcomes.[24] We analyzed it using multivariable multilevel logistic regression, which adjusts for the correlation between outcomes. We built separate models for each of the indicators, modeling the episode level determination of adherence against practice-level performance on other indicators and the patient- and practice-level covariates. We built the models in two stages, first including the practice-level adherence rates for the indicators not being studied (for example, acute throat pain, sinusitis, and urinary tract infections when the likelihood of adhering to the guideline for bacterial skin infections is being modeled), and then including all other patient and practice-level covariates. We included episode number as a quadratic term because graphical assessment indicated that its relationship with adherence was non-linear. As a sensitivity analysis, we redefined the length of a disease episode, reducing the duration by 50% and repeated our analysis. In addition, we repeated our analysis with the study population restricted to solo practices, in effect examining the performance of individual physicians. Results were expressed as odds ratios (OR) and 95% confidence intervals (CI).

To quantify the variation in outcome we estimated both the intraclass correlation coefficient (ICC) and the median odds ratio (MOR) using the level specific variance estimates obtained from the multilevel model. The ICC expresses the proportion of the variation in outcome between practices that is attributable to the different hierarchical levels. The MOR expresses this heterogeneity on an odds ratio scale. The MOR represents the median increase in odds of receiving guideline appropriate treatment when two individuals with the same covariates are compared, with one admitted to a hospital in a cluster with a lower likelihood of adherence, in this case poorly performing practices.[25,26] These measures yielded practice- and patient-level estimates of heterogeneity. We present only the practice-level measures in the results.

All analyses were carried out using STATA version 11.0 (StataCorp, College Station, TX).

RESULTS

Our study population consisted of 30,757 disease episodes of a bacterial skin infection among 16,006 patients, 28,544 disease episodes of acute throat pain among 21,368 patients, 39,648 disease episodes of sinusitis among 25,527 patients, and 75,300 disease episodes of a urinary tract infection among 34,198 patients. Table 1 describes the characteristics of the various study populations and practices. The average age ranged from approximately 40 years among patients with acute throat pain to 59 years among patients with a urinary tract infection. For each diagnosis, there were more females than males with the largest disparity occurring in the urinary tract infection population (87.4% female). Patients were publicly insured in over 70% of cases. The average number of distinct disease episodes per patient ranged from 1.6 for acute throat pain to 3.2 for urinary tract infections. The total number of GP practices reporting was 118 for all indicators except acute throat pain (117). The number reporting in each year varied from 76 to 111 because of changes in LINH membership. The practices averaged 2.1 FTE's. Solo practices accounted for over 50% of all practices. Solo practices accounted for 35.6% (10,952) of the bacterial skin infection episodes, 35.4% (10,093) of the acute throat pain episodes, 29.2% (11,584) of the sinusitis episodes, and 33.1% (24,892) of the UTI episodes.

Table 1. Characteristics of patients and practices

	Bacterial skin infections (N = 30,757)	Acute throat pain (N = 28,544)	Sinusitis (N = 39,648)	Urinary tract infections (N = 75,300)
Total patient episodes, n (practices reporting)				
2002	8,485 (111)	8,504 (110)	12,088 (111)	21,957 (111)
2003	8,167 (93)	7,458 (94)	10,561 (94)	18,406 (94)
2004	6,838 (75)	5,904 (75)	8,148 (75)	16,266 (75)
2005	7,267 (76)	6,678 (76)	8,851 (76)	18,671 (76)
Age, mean (sd)	42.0 (20.6)	40.4 (16.3)	45.0 (14.9)	58.8 (21.6)
Gender, n (%)				
Male	12,539 (40.8)	11,012 (38.6)	13,085 (33.0)	9,512 (12.6)
Female	18,218 (59.2)	17,532 (61.4)	26,563 (67.0)	65,788 (87.4)
Insurance, n (%)				
Public	22,270 (72.4)	20,841 (73.0)	27,777 (70.1)	57,675 (76.6)
Private	8,487 (27.6)	7,703 (27.0)	11,871 (29.9)	17,625 (23.4)
Episode per patient, mean (range)	3.4 (1, 62)	1.6 (1, 26)	2.2 (1, 39)	3.2 (1, 56)
Total number of practices reporting	118	117	118	118
FTE, mean (range)	2.1 (1, 5.5)	2.1 (1, 5.5)	2.2 (1, 5.5)	2.1 (1, 5.5)
Adherence rates, mean (sd)				
2002	74.3 (9.3)	64.1 (15.3)	39.7 (16.8)	60.5 (14.8)
2003	76.5 (9.2)	65.7 (15.3)	42.1 (16.1)	59.6 (13.9)
2004	75.6 (8.8)	62.3 (17.5)	40.8 (16.2)	60.9 (10.6)
2005	75.3 (8.6)	67.0 (11.6)	41.8 (16.7)	58.7 (12.5)
Practice type, n (% of practices)				
Solo	64 (54.2)	63 (53.9)	64 (54.2)	64 (54.2)
Duo	27 (22.9)	27 (23.1)	27 (22.9)	27 (22.9)
Group practice	20 (17.0)	20 (17.1)	20 (17.0)	20 (17.0)
Health Center	7 (5.9)	7 (6.0)	7 (5.9)	7 (5.9)

The Pearsons correlation coefficient for the association between each of the practice-level adherence rates for the four indicators is displayed in table 2. The results include 112 practices with the majority of the coefficients significant at the 0.05 level. The correlation between the restrictive guidelines were all positive, those with the first-choice guideline were negative.

The results of the adjusted multivariable logistic multilevel analyses are displayed in table 3. The unadjusted models (not shown) indicated significant associations between all the indicators, except between bacterial skin infections and urinary tract infections. On adjustment for the patient- and practice-level covariates, the pattern of results persisted for all indicators. The odds ratio for the three practice-level adherence rates included as dependent variables in each model was also calculated to represent the effect of a 10-unit (10%) increase in the adherence rate. The corresponding measures of heterogeneity

Table 2. Pairwise correlation of average practice-level guideline adherence rates for 112 practices

	Bacterial skin infections	Acute throat pain	Sinusitis	Urinary tract infections
Bacterial skin infections	1.00			
P-value	-			
Acute throat pain	0.52	1.00		
P-value	<0.001	-		
Sinusitis	0.52	0.57	1.00	
P-value	<0.001	<0.001	-	
Urinary tract infections	-0.18	-0.27	-0.29	1.00
P-value	0.06	<0.01	<0.01	-

were as follows: bacterial skin infections (ICC 0.04; MOR 1.82), acute throat pain (ICC 0.14; MOR 2.03), sinusitis (ICC 0.14; MOR 2.38), and urinary tract infections (ICC 0.15; MOR 2.38).

Patient- and practice-level covariates that displayed an association with the guideline adherence rate included the age and gender of the patient, and the source of insurance, a proxy for socioeconomic status. The history between the patient and GP for each diagnosis, characterized by the episode number, had a significant impact on the decision making process. Practice type was associated with adherence for a number of conditions. Practice workload was only associated with the adherence for the acute throat pain indicator.

In the sensitivity analysis (Appendix 3, Table A1), we reduced the duration of a defined episode thus increasing the total number of episodes. Thus, there were 32,399 episodes of bacterial skin infections, 29,496 episodes of acute throat pain, 40,619 episodes of sinusitis, and 82,551 episodes of urinary tract infections. The results of this analysis did not differ significantly from the main analysis.

In the subgroup analysis (Table 4), restricted to solo practices the odds ratios for the indicators did not differ from the analysis including all practices. However, the median odds ratios (bacterial skin infections 2.18; acute throat pain 2.41; sinusitis 2.85; urinary tract infections 2.59) were higher than in the main analysis.

DISCUSSION

Using process indicators related to adherence to antibiotic prescribing guidelines for four common conditions, this study shows relatively weak correlation between practice-level performance, and the limited predictive effect of adherence rates to an antibiotic prescribing guideline on the odds of adhering to other antibiotic prescribing guidelines in our set.

Table 3. Multilevel logistic regression analysis of adherence to antibiotic prescribing guidelines for all four indicators

	Bacterial skin infections	Acute throat pain	Sinusitis	Urinary tract infections
Guideline				
Bacterial skin infections	-	1.03 (1.03 – 1.04)	1.03 (1.02 – 1.04)	0.99 (0.99 – 1.00)
10% increase	-	1.30 (1.21 – 1.41)	1.33 (1.24 – 1.43)	0.96 (0.91 – 1.01)
Acute throat pain	1.02 (1.01 – 1.02)	-	1.02 (1.01 – 1.02)	0.99 (0.99 – 0.99)
10% increase	1.18 (1.13 – 1.24)	-	1.17 (1.12 – 1.21)	0.92 (0.90 – 0.95)
Sinusitis	1.01 (1.01 – 1.02)	1.02 (1.01 – 1.02)	-	0.98 (0.98 – 0.99)
10% increase	1.12 (1.07 – 1.18)	1.19 (1.13 – 1.24)	-	0.85 (0.83 – 0.85)
Urinary tract infections	0.99 (0.99 – 1.00)	0.99 (0.98 – 0.99)	0.97 (0.97 – 0.98)	-
10% increase	0.93 (0.88 – 0.99)	0.88 (0.84 – 0.93)	0.76 (0.73 – 0.80)	-
Age	0.96 (0.96 – 0.97)	1.03 (1.03 – 1.03)	1.00 (0.99 – 1.00)	1.00 (0.99 – 1.00)
Gender				
Male	Reference	Reference	Reference	Reference
Female	2.67 (2.38 – 2.99)	1.07 (0.99 – 1.16)	0.93 (0.87 – 1.00)	6.41 (5.97 – 6.90)
Insurance				
Public	Reference	Reference	Reference	Reference
Private	0.91 (0.80 – 1.02)	0.85 (0.78 – 0.93)	0.90 (0.83 – 0.96)	1.04 (0.99 – 1.10)
Episode	1.09 (1.06 – 1.11)	1.26 (1.21 – 1.32)	1.71 (1.66 – 1.77)	0.79 (0.78 – 0.80)
Episode squared	1.00 (1.00 – 1.00)	-	0.90 (0.98 – 0.99)	1.01 (1.01 – 1.01)
FTE	1.04 (0.83 – 1.31)	0.69 (0.54 – 0.87)	1.11 (0.83 – 1.49)	1.04 (0.81 – 1.33)
Practice type				
Solo	Reference	Reference	Reference	Reference
Duo	0.75 (0.52 – 1.08)	1.73 (1.16 – 2.58)	0.94 (0.57 – 1.55)	1.31 (0.90 – 1.89)
Group	0.71 (0.44 – 1.16)	2.67 (1.63 – 4.39)	0.85 (0.47 – 1.54)	1.43 (0.94 – 2.19)
Health Center	0.72 (0.41 – 1.26)	2.84 (1.63 – 4.94)	0.95 (0.51 – 1.78)	1.87 (1.18 – 2.96)
Year of admission				
2002	Reference	Reference	Reference	Reference
2003	1.02 (0.90 – 1.16)	0.89 (0.80 – 0.99)	0.72 (0.66 – 0.78)	1.21 (1.14 – 1.28)
2004	1.06 (0.92 – 1.22)	0.69 (0.62 – 0.77)	0.62 (0.56 – 0.68)	1.36 (1.27 – 1.45)
2005	0.98 (0.85 – 1.14)	0.90 (0.80 – 1.00)	0.60 (0.54 – 0.66)	1.43 (1.33 – 1.53)
Intraclass correlation coefficient	0.04	0.14	0.14	0.15
Median odds ratio (95% credible intervals)	1.82 (1.60 – 2.03)	2.03(1.80 – 2.26)	2.38 (2.06 – 2.71)	2.38 (2.06 – 2.70)

We also observed that practices modified their behavior over time, as they developed a 'history' with their patients. The variable, episode number, along with its squared term indicate a nonlinear relationship between number of disease episodes and guideline adherence. After a few occurrences, physicians were less likely to adhere to guidelines

Table 4. Multilevel logistic regression analysis of adherence to antibiotic prescribing guidelines for solo practices (Subgroup analysis)

	Bacterial skin infections	Acute throat pain	Sinusitis	Urinary tract infections
No. of disease episodes	10,952	10,093	11,584	24,892
No. of patients	5,789	7,525	7,793	11,904
No. of practices	64	63	64	64
Guideline				
Bacterial skin infections	-	1.03 (1.02 – 1.05)	1.03 (1.02 – 1.04)	0.99 (0.98 – 1.00)
10% increase	-	1.40 (1.24 – 1.58)	1.31 (1.17 – 1.46)	0.92 (0.86 – 0.99)
Acute throat pain	1.02 (1.01 – 1.02)	-	1.01 (1.00 – 1.02)	0.99 (0.99 – 0.99)
10% increase	1.18 (1.10 – 1.27)	-	1.10 (1.03 – 1.17)	0.89 (0.86 – 0.93)
Sinusitis	1.01 (1.01 – 1.02)	1.01 (1.00 – 1.02)	-	0.99 (0.99 – 1.00)
10% increase	1.14 (1.06 – 1.23)	1.11 (1.03 – 1.18)	-	0.94 (0.90 – 0.98)
Urinary tract infections	0.99 (0.98 – 1.00)	0.98 (0.98 – 0.99)	0.98 (0.98 – 0.99)	-
10% increase	0.91 (0.82 – 1.01)	0.85 (0.78 – 0.92)	0.84 (0.78 – 0.90)	-
Intraclass correlation coefficient	0.06	0.13	0.17	0.19
Median odds ratio (95% credible intervals)	2.18 (1.70 – 2.64)	2.41 (1.94 – 2.89)	2.85 (2.16 – 3.54)	2.59 (2.02 – 3.15)

the next time they diagnosed the same patient with the same condition. This finding, akin to clinical intuition, is not surprising but is often neglected in assessing physician performance. Future studies of guideline adherence should take into account the effect of the physician-patient relationship.

This study has a number of strengths and limitations. It was limited by the lack of information about severity at presentation, which is an exception to prescribing guidelines not captured in administrative reviews. However, we did not expect that more severe patients would concentrate in specific practices. The variable FTE does not specify the role of the employee but the correlation between the size of a practices patient list, and the number of GPs and the need for administrative work makes it a good approximation of practice workload. We were unable to identify individual physicians in the larger practices but we performed a subgroup analysis restricted to solo practices with similar results. It has been noted that the organizational culture of healthcare facilities has an influence on behavior. Physicians in one practice are more likely to behave more like each other.[27] This has been found to be true for prescribing behavior, and for hospital characteristics such as length of stay, and even among practitioners performing the same procedures in separate practices.[28] Studies of prescribing guideline adherence are often criticized for being restricted to filled prescriptions, incomplete follow-up, early refills/medication changes, use of over-the counter (OTC) replacements, the duration of patient eligibility for treatment, disease identification, and presence of contraindications.[29] Our approach to defining disease episodes, unavailability of OTC antibiotics in the Netherlands and the nature of the research question (focus on initial physician behavior) negates the impact of these criticisms. The data used in this study were collected in practices with good access to guidelines and diagnostic aids. The extensive use of electronic medical records, including notation of dispensed prescriptions, and continuity of care were beneficial for patient follow-up. This electronic medical record system was the setting for the evaluation of the electronic prescribing system in the Netherlands.[30] Another critique of studies of guideline adherence relates to factors that influence prescribing behavior. A qualitative study of Dutch GPs cited lack of agreement and external factors as the most significant barriers to adherence to guidelines in general practice.[31] These findings echo those of a previous study examining adherence to practice guidelines.[32] The external factors identified have limited influence on the decision whether to prescribe antibiotics. A lack of agreement was assessed by examining the variation and levels of adherence among practices. The conditions selected are common and have guidelines that were used in a national performance assessment among Dutch GP practices. Antibiotic prescribing is restricted in the Netherlands; it has very low prescribing rates compared to other countries.[33]

The use of a process indicator to summarize quality of care is dependent on the assumption that the indicator in question captures an unmeasured characteristic of the practice that is responsible for the performance. The same principle is applied when several indicators are used to measure the same domain, as is the case for prescribing quality. However, we observed pairwise correlations between adherence rates to antibiotic prescribing guidelines among our practices that were weaker than would be expected if they were measuring the same characteristic. The associations between the restrictive and first choice guidelines were particularly weak. A general lack of agreement with a guideline, often cited as reason for poor adherence rates, is a factor that might influence the correlations. In our study, we observed that between 4% and 14% of the variation in adherence to each of the guidelines is attributable to differences between practices as indicated by the ICC for each model. The high MOR for each guideline further indicates marked differences in a patients odds of receiving guideline appropriate care at different practices. These observed differences were even stronger when examining the behavior of individual physicians in solo-practices. The relatively high adherence rates and large residual variation attributed to differences between practices reduces the likelihood that the weak association is a result of a general lack of agreement with the selected guidelines.[34] The large residual variation attributable to differences between practices, including physician behavior is highlighted in the subgroup analysis.

The guidelines in this study were developed using an evidence-based and consensus model, which appears to yield good clinical practice measures linked to desired outcomes.[35-37] The mere existence of guidelines is enough to reduce some variation in behavior between practices[38], thus improving the quality of care for individual patients. On a societal level, the restrictive prescribing of antibiotics reduces the risk of antibiotic resistant bacteria.[39] Unfortunately, performance indicators have not always been applicable as measures of overall quality.[40-42] While collectively the guideline-based indicators we studied are intended to reflect a providers prescribing quality, we did not observe a strong association between adherence to one guideline and adherence to another. This was in spite of the fact that all of the guidelines measured address one related field, antibiotic prescribing. The negative association between adhering to a first choice guideline and restrictive guideline shows a dichotomy identified in the literature.[43] This previously demonstrated dichotomy already raises red flags as to what the indicators are actually measuring. The individual indicators included in our set capture a practices performance regarding the corresponding condition but do not appear, on their own or collectively, to adequately represent a facilities prescribing quality beyond these conditions. This highlights an important distinction between performance indicators and their use as quality indicators. The observed performance indicates that the physician is 'good' at the specified task but its interpretation as a marker of wider quality may be limited. The use of process indicators to measure quality of care has been useful for individual conditions. Failure to give advice on smoking cessation or prescribing aspirin for patients with a myocardial infarction, are direct measures of the poor quality of care received for myocardial infarction. However, once these measures are elevated beyond these conditions, their interpretation becomes complicated. In an editorial, Press attempted to differentiate "quality improvement measures" meant for continuous internal consumption from "accountability measures" intended to be used as report cards.[44] O'Brien and Peterson expressed the conundrum succinctly, after highlighting the potential and realized usefulness of process indicators, by adding the caveat "...to be effective, providers must properly interpret these performance reports.....".[45] As a wider audience uses these performance frameworks, the same message applies. It can be argued that this indicator set captures different aspects of prescribing quality but the problem still arises that we do not then know which aspects. Thus, the interpretation of the individual measures and performance is open to discussion.

With the exponential development of indicators and the awareness of consumers and payers, quality of care research faces a double burden of measurement and interpretation. In the measurement of overall quality, there is the need to examine whether theoretically related process measures are indeed measuring the same underlying concept. In the absence of evidence of links between process measures, the interpretation of performance should be selective, to avoid misleading consumers of performance information.

References

(1) Arah OA, Klazinga NS, Delnoij DM, ten Asbroek AH, Custers T. Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement. Int J Qual Health Care 2003;15:377-398.

(2) Arah OA, Westert GP, Hurst J, Klazinga NS. A conceptual framework for the OECD Health Care Quality Indicators Project. Int J Qual Health Care 2006;18 Suppl 1:5-13.

(3) ten Asbroek AH, Arah OA, Geelhoed J, Custers T, Delnoij DM, Klazinga NS. Developing a national performance indicator framework for the Dutch health system. Int J Qual Health Care 2004;16 Suppl 1:i65-i71.

(4) Hammermeister KE, Shroyer AL, Sethi GK, Grover FL. Why it is important to demonstrate linkages between outcomes of care and processes and structures of care. Med Care 1995;33:OS5-16.

(5) Shroyer AL, London MJ, Sethi GK, Marshall G, Grover FL, Hammermeister KE. Relationships between patient-related risk factors, processes, structures, and outcomes of cardiac surgical care. Conceptual models. Med Care 1995;33:OS26-OS34.

(6) Shafi S, Parks J, Ahn C et al. Centers for Medicare and Medicaid services quality indicators do not correlate with risk-adjusted mortality at trauma centers. J Trauma 2010;68:771-777.

(7) Isaac T, Jha AK. Are patient safety indicators related to widely used measures of hospital quality? J Gen Intern Med 2008;23:1373-1378.

(8) Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. Health Aff (Millwood) 2007;26:1104-1110.

(9) Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. JAMA 2006;296:2694-2702.

(10) Bradley EH, Herrin J, Elbel B et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. JAMA 2006;296:72-78.

(11) Patterson ME, Hernandez AF, Hammill BG et al. Process of care performance measures and long-term outcomes in patients hospitalized with heart failure. Med Care 2010;48:210-216.

(12) Fonarow GC, Peterson ED. Heart failure performance measures and outcomes: real or illusory gains. JAMA 2009;302:792-794.

(13) Hernandez AF, Hammill BG, Peterson ED et al. Relationships between emerging measures of heart failure processes of care and clinical outcomes. Am Heart J 2010;159:406-413.

(14) Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. Int J Qual Health Care 2003;15 Suppl 1:i5-11.

(15) Netherlands Institute for Health Services (NIVEL). Netherlands Information Network for General Practice. 2009. Available at: www.linh.nl. Accessed 1 September-2009.

(16) Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, van der ZJ. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health 2005;15:59-65.

(17) De Jong JD, Groenewegen PP, Spreeuwenberg P, Westert GP, de Bakker DH. Do decision support systems influence variation in prescription? BMC Health Serv Res 2009;9:20.

(18) Lamberts H, Woods M eds. ICPC, International Classification of Primary Care. New York: Oxford University Press, 1987.

(19) WHO. Anatomical Therapeutic Chemical (ATC) classification index. WHO Collaborating Center for Drug Statistics Methodology; 1993;Oslo.

(20) Schellevis FG, Westert GP. The design of the second Dutch National Survey of General Practice. In: Westert GP, Jabaaij L, Schellevis FG, eds. Morbidity, performance and quality of primary care. Oxford, United Kingdom: Radcliffe Publishing; 2006; p10-18.

(21) Ong DS, Kuyvenhoven MM, van Dijk L, Verheij TJ. Antibiotics for respiratory, ear and urinary tract disorders and consistency among GPs. J Antimicrob Chemother 2008;62:587-592.

(22) Starreveld JS, Zwart S, Boukes FS, Wiersma T, Goudswaard AN. [Summary of the practice guideline 'Sore throat' (second revision) from the Dutch College of General Practitioners]. Ned Tijdschr Geneeskd 2008;152:431-435.

(23) Wiersma TJ, Timmermans AE. [Summary of the 'Urinary tract infections' guideline (first revision) of the Dutch College of General Practitioners]. Ned Tijdschr Geneeskd 2001;145:735-739.

(24) Hox JJ. Applied Multilevel Analysis. Amsterdam: TT Publications, 1995.

(25) Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. Am J Epidemiol 2005;161:81-88.

(26) Merlo J, Chaix B, Ohlsson H et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health 2006;60:290-297.

(27) Fattore G, Frosini F, Salvatore D, Tozzi V. Social network analysis in primary care: the impact of interactions on prescribing behaviour. Health Policy 2009;92:141-148.

(28) De Jong JD, Westert GP, Lagoe R, Groenewegen PP. Variation in hospital length of stay: do physicians adapt their length of stay decisions to what is usual in the hospital where they work? Health Serv Res 2006;41:374-394.

(29) Kawamoto K, Allen LaPointe NM, Silvey GM, Anstrom KJ, Eisenstein EL, Lobach DF. Development and evaluation of an improved methodology for assessing adherence to evidence-based drug therapy guidelines using claims data. AMIA Annu Symp Proc 2007;394-398.

(30) Wolters I, van den Hoogen H, de Bakker D. [Evaluatie-invoering Elektronisch Voorschrijf Systeem eindrapport]. Utrecht, Netherlands, NIVEL; 2003

(31) Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. Implement Sci 2009;4:54.

(32) Cabana MD, Rand CS, Powe NR et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999;282:1458-1465.

(33) Cars O, Molstad S, Melander A. Variation in antibiotic use in the European Union. Lancet 2001;357:1851-1853.

(34) Ohlsson H, Merlo J. Is physician adherence to prescription guidelines a general trait of health care practices or dependent on drug type?--a multilevel logistic regression analysis in South Sweden. Pharmacoepidemiol Drug Saf 2009;18:682-690.

(35) Pilling S. History, context, process, and rationale for the development of clinical guidelines. Psychol Psychother 2008;81:331-350.

(36) Rycroft-Malone J. Formal consensus: the development of a national clinical guideline. Qual Health Care 2001;10:238-244.

(37) Silagy CA, Stead LF, Lancaster T. Use of systematic reviews in clinical practice guidelines: case study of smoking cessation. BMJ 2001;323:833-836.

(38) de Jong JD, Groenewegen PP, Spreeuwenberg P, Schellevis F, Westert GP. Do guidelines create uniformity in medical practice? Soc Sci Med 2010;70:209-216.

(39) Arroll B, Kenealy T, Kerse N. Do delayed prescriptions reduce antibiotic use in respiratory tract infections? A systematic review. Br J Gen Pract 2003;53:871-877.

(40) Veninga CC, Denig P, Pont LG, Haaijer-Ruskamp FM. Comparison of indicators assessing the quality of drug prescribing for asthma. Health Serv Res 2001;36:143-161.

(41) Avery AJ, Heron T, Lloyd D, Harris CM, Roberts D. Investigating relationships between a range of potential indicators of general practice prescribing: an observational study. J Clin Pharm Ther 1998;23:441-450.

(42) Andersen M. Is it possible to measure prescribing quality using only prescription data? Basic Clin Pharmacol Toxicol 2006;98:314-319.

(43) van Roosmalen MS, Braspenning JC, De Smet PA, Grol RP. Antibiotic prescribing in primary care: first choice and restrictive prescribing are two different traits. Qual Saf Health Care 2007;16:105-109.

(44) Press I.The measure of quality. Qual Manag Health Care 2004;13:202-209.

(45) O'Brien SM and Peterson ED. Identifying high-quality hospitals: consult the ratings or flip a coin? Arch Intern Med 2007;167:1342-1344.

APPENDIX 3

Table A1. Multilevel logistic regression analysis of adherence to antibiotic prescribing guidelines with the length of a diseases episode halved (Sensitivity analysis)

	Bacterial skin infections	Acute throat pain	Sinusitis	Urinary tract infections
Number of episodes	32,399	29,496	40,619	82,551
Guideline				
Bacterial skin infections	-	1.02 (1.02 – 1.03)*	1.03 (1.02 – 1.03)*	0.99 (0.98 – 1.00)*
Acute throat pain	1.02 (1.01 – 1.02)*	-	1.01 (1.00 – 1.02)*	0.99 (0.99 – 0.99)*
Sinusitis	1.01 (1.01 – 1.02)*	1.02 (1.01 – 1.02)*	-	0.99 (0.98 – 0.99)*
Urinary tract infections	0.99 (0.98 – 1.00)*	0.99 (0.98 – 0.99)*	0.97 (0.97 – 0.98)*	-

* P-value <0.05