Methodology Matters

"Methodology Matters" is the title of a new feature in the Journal: a series of intermittently appearing articles on methodology. Other articles on such topics as Confidence Intervals are planned. Suggestions from readers of additional topics they would like to see covered in this series are welcome.

III Validity Review of Performance Measures

ANN G. LAWTHERS

Center for Quality of Care Research and Education, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

Glossary of Terms

Algorithm. A rule of procedure, or set of instructions, containing conditional logic for solving a problem or accomplishing a task.

Criterion (plural: criteria). A statement or standard of care by which data concerning clinical events experienced by individuals are classified in order to create a performance measurement.

Flag ("flagged"). A mark or code indicating that review finds a criterion has not been met.


1) Instrument that sums data about the health care given to a population to create a rate or score for average performance.
2) Instrument that measures the extent to which a health-care provider delivers clinical services that are appropriate to the patient's condition, provides them safely, competently, and in an appropriate time frame, and achieves desired outcomes in terms of those aspects of patient health and patient satisfaction that can be affected by clinical services (also: Clinical Performance Measure).

A structured process for evaluating the validity of performance measures is essential if the rates generated by these measures are to be relied upon for large-scale programmes to improve clinical care. In this evaluation process a panel of objective clinicians who were not involved in developing the performance measures examines the criteria, coding instructions, and coding form. The panel's job is to determine the degree to which a sample of patient records containing treatment elements that have been flagged for nonconformance to certain criteria are truly nonconformant. Modification of the performance measure according to the recommendations of the panel provides an important means of reducing measurement error by eliminating obvious sources of false positive findings.

This article describes the procedures followed and the materials used to validate the performance measures developed for the project to Develop and Evaluate Methods for Promoting Ambulatory Care Quality (DEMPAQ), a research project contracted by the Health Care Financing Administration of the United States government [1-3]. DEMPAQ is described in Fig. 1. In the DEMPAQ project, record review is conducted by entering data directly from patient ambulatory records into a computer system. Instructions for abstracting the data are displayed on the computer monitor.
DEMPAQ
A Project to Develop and Evaluate Methods for Promoting Ambulatory Care Quality

DEMPAQ was a three-year demonstration project to develop tools to review physician office care for the U.S. Government's Peer Review Organization (PRO) program. PROs are organizations of physicians in active practice who contract with the U.S. government to review quality of care in the Medicare program, a program that serves persons over 65 years of age.

The project emphasized education and feedback to physicians and concentrated on the quality of care rendered in office settings. The study population was a random sample of primary care physicians with an office practice who treat Medicare patients in the states of Maryland, Iowa, and Alabama, U.S.A.

DEMPAQ developed two review methods: a review of medical records and an instrument for creating profiles from a national data base of medical care billing information. The following is a summary of the methods used for review of medical records.

The record review focused on functions the physician typically performs in patient encounters: ordering tests, prescribing drugs, performing procedures, making new diagnoses, following up on abnormalities, and offering preventive or screening care. Rates of satisfactory performance are averaged across the patients treated by this physician to yield a profile for the physician. The rates are generated by abstracting data from medical records and applying to them computerized algorithms, one for each physician function.

Personal data are reported to each provider along with the corresponding rates for their peers. This educational component is critical to the philosophy of the project. The occasion of this feedback is used as an opportunity to solicit comments from the physicians whose performance was reviewed about the usefulness and value of the review.

A critical element in the project is that input from physician organizations has been sought in both the design and evaluation phases of the project. The record review criteria have been submitted to physicians representing state and national physicians' organizations for critique and comment before implementing the profiling or record review. Following the feedback phase, the same physician organizations are asked to evaluate the educational value and usefulness of the data.

In addition to developing the record review, DEMPAQ has undertaken extensive evaluation of both the validity and reliability of the review technique. The patient record data are also compared to that contained in the HCFA1500* to assess whether claims data can be reliably used to assess quality.

* Administrative database used in the US federal government's program of health care for the elderly.
METHODS

Training the panel

Validity reviewers are provided with a manual of procedures describing the purpose of the performance review and the panel's role in critiquing the performance measure. They are given a two-page review worksheet for each element of care that has been flagged for reconsideration. Page 1 of the worksheet (Fig. 2) poses the review questions, and Page 2 (Fig. 3) provides details of the review criteria and the abstraction process that produced this flag. Reviewers are instructed that the objective of examining patient records for validity review is not to make a determination of health care quality for a particular case but to focus on the ability of the medical review criteria and abstracting instructions to identify guideline conformance. It is emphasized that the review evaluates the criteria, not the clinician.

Reviewers are instructed to consider the following broad concerns as they examine very specific questions about the clinical content of each criterion:

1. Are the review criteria appropriate for the performance measure? If not, the reviewers are asked to suggest more appropriate criteria so that outdated, inappropriate or ambiguous ones may be replaced with criteria that are more useful.
2. Are the review criteria applicable to the case? If not, the reviewers are asked to give a reason why the criteria do not apply, so that the abstractor's instructions for the review process can be improved.
3. Was the flag (i.e. the coding that the case did not conform to the guideline) justified? If not, the reviewers are asked to give reasons so that, where necessary, additional acceptable alternatives or abstracting rules may be formulated.

Review worksheet, page 1

In addition to having the manual of instructions, DEMPAQ reviewers also receive the patient's medical record and the two-page review worksheet described above (Figs 2 and 3). At the top of page 1 of the review worksheet is a 2- by 4-inch self-adhesive label (Fig. 4) generated by the computer on the basis of the case reviews stored within it. The label bears information identifying the case and the flagged item to be reviewed. Flags are generated each time a review finds that the practitioner did not meet a criterion for one of the indicators associated with the item.

The entries on the label have the following meanings:

Flag. The code for the criterion (e.g., D1, T3)
Code. The code of the clinical item (e.g., hypertension [HTN], diabetes mellitus [DM], hematocrit [HCT], electrocardiogram [ECG]).
Function. One of the six physician activities assessed by DEMPAQ:
- prescribing drugs
- making new diagnoses
- ordering tests
- performing procedures
- following up on abnormalities
- offering preventive or screening care
Item. A particular clinical service related to this activity; e.g., for ordering tests, one of the clinical items is a hemoglobin test.
Date. The date of the visit note in which the item was observed by the abstractor.
Reason. A brief remark describing the reason for the flag (e.g., "Actions not taken to monitor diagnosis appropriately," "No appropriate follow-up of abnormal test results").
Reference. Additional information for some flags that assists in reviewing the flagged item.

Below the label is a list of possible evaluations of the flag (Fig. 1). After reviewing the patient record with the assistance of the information on page 2 of the worksheet, reviewers are asked to select the one response that fits most closely with their evaluation of the flagged item and to explain their response with comments. The same choice of responses is offered for all flags:

1. The criterion flag was justified; the criterion is appropriate.
2. The criterion could not be met in this circumstance because of patient factors. The reviewer is asked to describe the patient factors under "Comments".
3. The criterion could not be met in this circumstance because of practice factors. The reviewer is asked to describe the practice factors under "Comments".

Downloaded from https://academic.oup.com/intqhc/article-abstract/8/3/299/1845261 by guest on 25 December 2018
The patient care episode shown on the label above was flagged for physician review because it failed to meet the DEMPAQ record review criteria. On the attached page you will find the review format and the criteria used by the initial PRO reviewer. Please review the medical record for the item of care shown above, and indicate below your own judgment about the item in this case.

CONSIDER: Are the review criteria appropriate for this indicator? Are the review criteria applicable to the case at hand? Was the flag justified in this case?

Choose only the one response that most closely fits your opinion. Use the spaces below and the back of this page to give your comments.

___ 1. The criterion flag was justified given the data in the medical record.

___ 2. The criterion could not be met in this circumstance because of patient factors. Describe the patient factors below under "Comments."

___ 3. The criterion could not be met in this circumstance because of practice factors. Describe the practice factors below under "Comments."

___ 4. The criterion should not be met in this case due to extenuating circumstances. Describe the extenuating circumstance below under "Comments."

___ 5. The flag is not justified because of abstractor error. Describe the error under "Comments."

___ 6. The flag is not justified because the abstracting decision rule is incorrect or incomplete. Suggest a revised decision rule below.

___ 7. The flag is not justified because the criterion needs to be revised. Propose a revised criterion for this item below.

Criterion/Coding rule revisions: __________________________________________

Comments: ___________________________________________________________

Is the documentation in this record adequate to conduct a valid quality of care review?
DEMPAQ Review Worksheet - Page 2

Ischemic heart disease

Review question displayed on DRRS† screen

Evidence for a new diagnosis:
Patient ID:   Diag:   Visit date:

Indicate if any of the following evidence for a new diagnosis is in any
note in the medical record:  1 = yes   2 = no

Look in entire medical record.

0 <

Clinical criteria displayed in window on DRS screen:

Evidence establishing a new diagnosis:

1. Cardiac chest pain (worse with exercise, relieved by rest) with ST segment
changes on ECG

OR

2. Positive exercise tolerance test

OR

3. Positive cardiac catheterization

† DEMPAQ Record Review System

FIGURE 3. DEMPAQ peer review worksheet — page 2.

(4) The criterion could not be met in this case
due to extenuating circumstances. The reviewer is
asked to describe the extenuating circumstance under “Comments.”

(5) The flag is not justified because of abstrac-
tor error. The reviewer is asked to describe the
error under “Comments.”

(6) The flag is not justified because the
abstracting decision rule is incorrect. The
reviewer is asked to suggest a revised decision
rule.

(7) The flag is not justified because the
criterion needs to be revised. The reviewer is
asked to propose a revised criterion.
Examples of how to select the proper response are shown in Fig. 5.

Review worksheet, page 2

Page 2 is specific to the flagged clinical item being reviewed. There are many criterion-item combinations (e.g., monitoring a diagnosis of diabetes, follow-up of abnormal creatinine). The information shown on the sample screen on Page 2 of the worksheet is that of the case under review, as given on the label on the first page of the worksheet.

The abstractor uses an on-screen format on the computer monitor to answer the questions that produced the flag and this format is reproduced on page 2 of the worksheet. It is this generic format of the question that is seen by the initial reviewer when evaluating the criterion.

Below this illustration of the format on the computer monitor, the worksheet shows the text of the criterion used for review of the specific clinical item. By seeing the actual review questions that appear on the data entry screen, the list of possible responses, and the clinical criteria unique to that item, the validity reviewer attempts to reconstruct the rationale for the original review that resulted in a flag. This exercise assists the reviewer in assessing the validity of the criterion.

Review co-ordinator

A review co-ordinator, who is an experienced reviewer, assists the review panelists by preparing the worksheet and the patient record for the review and is present during the review to answer any questions that the clinician reviewer might have. The co-ordinator attaches the computer-generated label to page 1 and selects the appropriate second page for the worksheet to reflect the criterion-item combination specified on the label. Chart preparation involves locating the visit date in the record and the clinical item identified on the worksheet label. Relevant clinical information is highlighted in the patient record to reduce the reviewer’s time spent searching the record. The review co-ordinator is familiar with the abstracting rules used by the original peer review organization reviewer and, therefore, can clarify any questions about the way the original review was conducted. This strategy of using a review coordinator to facilitate the work of the validity reviewer has made the review efficient, allowing the clinician to concentrate on evaluating the criteria.

Examples of validity review evaluations

Figure 5 is an excerpt from the reviewer instruction manual. It shows examples of situations that would result in one of the seven criterion judgments listed on page 1 of the worksheet.

Using the validity review evaluations

The panel staff summarize the validity reviewers’ evaluations and comments. The panelists examine these summaries to determine
1. The criterion flag was justified; the criterion is appropriate.

**Example**
The action required by the criteria is appropriate for the patient in this case. You have searched the specified time period in the patient record, and do not find the action. The review was apparently conducted accurately, and you do not disagree with the criterion.

2. The criterion could not be met in this circumstance because of patient factors.

**Example A**
The action required is appropriate, but is not found in the patient record. The record states that the patient was away from home for the two months during which the action was to have been performed.

**Example B**
The action required is appropriate, but is not found in the patient record. The record states that the patient was diagnosed with terminal lung cancer, making a mammogram unnecessary.

3. The criterion could not be met in this circumstance because of practice factors.

**Example**
The action required is appropriate, but is not found in the patient record. There is evidence in the record that the practice was in the process of converting to a new computerized practice management system, and the procedures for patient recall were disrupted during this time period.

4. The criterion should not be met in this case due to extenuating circumstances.

**Example**
The action required is appropriate, but is not found in the patient record. There is a note that the patient is undergoing psychotherapy and has initiated legal proceedings against her husband following an episode of domestic violence. In this case, the patient’s psychosocial problems supersede the strict time period of the criterion.

5. The flag is not justified because of abstractor error.

**Example A**
The action required is appropriate, but is not found in the patient record. There is a note that the patient refused an exam. This is a coding error. If the reviewer had coded "8" for patient refusal, the item would be scored as an "acceptable alternative," and not flagged.

**Example B**
You found evidence in the record that the action was done. The initial reviewer appears to have seen that visit note because other data from the visit have been used in the review.

6. The flag is not justified because the abstracting decision rule is incorrect.

**Example**
The action specified for monitoring diabetes is examination of the fundus. This is appropriate, and the patient record shows that "HEENT" were examined. The case is flagged because the abstracting rule is that for diabetes monitoring, a fundus exam should be stated explicitly. You believe that HEENT is sufficient evidence of a fundus exam, and do not think the case should be flagged.

7. The flag is not justified because the criterion needs to be revised.

**Example**
You disagree with the clinical content of the criterion as stated on the bottom of page 2 of the worksheet. Give your rationale and suggest a more appropriate criterion.

FIGURE 5. Excerpt from the DEMPAQ reviewer instruction manual.
whether any part of the performance measure should be modified. Suggestions for improving data abstraction forms and instructions are incorporated into the measure.

Suggestions regarding the criteria are handled in several ways. If the validity reviewers identify inconsistencies between the criteria and the guideline on which the performance measure is based, then the criteria may be revised accordingly. However, if the suggestions for changing criteria are contrary to guideline recommendations, and clinician convenience or habit is the reason given for rejecting the guideline, the criteria should not be modified. The panel notes the discrepancy and the number of reviewers who recorded criticism of the criterion. Interpretation of the performance rates for these criteria is influenced by the strength of disagreement with them.

For example, the American Board of Family Practitioners guideline for diabetes management recommends quarterly HgbA1C monitoring. Even though a large proportion of validity reviewers might suggest that a criterion to this effect should be changed to permit serum glucose testing as an alternative, changing the criterion would be inappropriate, because it is based on a published guideline. However, since the reviewers as a group feel so strongly about this point, discussions of performance rates by users of the measure include consideration of the differences between recommended and actual practice.


REFERENCES