



ELSEVIER

Contents lists available at ScienceDirect

Healthcare

journal homepage: www.elsevier.com/locate/hjdsi

Evaluating the feasibility and utility of translating Choosing Wisely recommendations into e-Measures



Kanaka D. Shetty^a, Daniella Meeker^{a,b,1}, Eric C. Schneider^{c,1}, Peter S. Hussey^{d,1}, Cheryl L. Damberg^{a,*}

^a RAND Corporation, 1776 Main Street, Santa Monica, CA 90401, USA

^b University of Southern California, Los Angeles, CA 90033, USA

^c The Commonwealth Fund, New York, NY, 10021, USA

^d RAND Corporation, Boston, MA, USA

ARTICLE INFO

Article history:

Received 6 April 2014

Received in revised form

8 November 2014

Accepted 16 December 2014

Available online 31 January 2015

Keywords:

Overuse measures

Efficiency measures

e-Measures

Choosing Wisely

Health information technology

ABSTRACT

Background: Efforts to reduce health care spending have focused on reducing use of low-value services, but relatively few performance measures address overuse of care. In 2012, the American Board of Internal Medicine Foundation's "Choosing Wisely" (CW) campaign identified 45 low-value services that clinicians and patients should avoid. Translating these overuse concepts into performance measures could assist in discouraging the use of these services. We assessed the feasibility and utility of converting these recommendations into e-Measures based on data from electronic health records [EHR].

Materials and methods: We used four criteria to evaluate 45 CW recommendations for e-Measure development: (1) feasibility of extracting needed data from EHR systems meeting Meaningful Use Stage 2 standards; (2) whether the recommendation's terminology was sufficiently specific for translation into an e-Measure; (3) scientific evidence supporting the recommendation; and (4) impact on reducing resource use.

Results: Only six of the 45 CW recommendations were deemed feasible for e-Measure development. Thirty-two recommendations require data elements unlikely to be found in current EHR systems; eight of 45 recommendations do not use sufficiently specific terminology.

Conclusions: Improved capture of clinical information in EHRs and greater specificity of clinical terminology are required to advance these overuse concepts into standardized e-measures.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Hundreds of billions of dollars may be spent on services with little clinical benefit.^{1,2} To address this problem, efforts to identify and encourage physicians to reduce the use of low-value services are beginning to emerge.^{3–5} Professional practice recommendations, performance measures, and clinical decision support tools are various approaches being considered to discourage the delivery of overused services, such as imaging in non-specific lower back pain.^{5–7} Prior experience with professional practice guideline generation has revealed that expert consensus is just one step in the difficult process of changing clinical practice.⁸ Converting professional practice recommendations into performance measures that are tied to audit and feedback, public reporting, or

financial incentives may be an effective complement to professional practice guidelines for discouraging the use of low-value services.^{3,9} However, the development and deployment of overuse measures have faced a number of challenges, most notably that their construction often depends on complex clinical information, including patient behaviors and physical exam findings, to define circumstances under which some commonly overused procedures might have substantial value.¹⁰

Although a variety of organizations including the Office of the National Coordinator for Health Information Technology (ONC) and the Centers for Medicare and Medicaid Services (CMS) seek to advance the development and use of measures of overuse constructed using EHR data,¹² few valid, evidence-based measures of overuse of services have been developed. e-Measures—performance measures specified using data from electronic health record [EHR] systems—may facilitate measurement of overuse of services because of the potential for capturing detailed clinical information during the patient encounter which could be used to construct

* Corresponding author. Tel.: +1 310 393 0411; fax: +1 310 451 7085.

E-mail address: damberg@rand.org (C.L. Damberg).

¹ Tel.: +1 310 393 0411.

measures. A 2009 study by Roth et al. highlighted challenges with constructing quality measures based on data available in EHRs, particularly data corresponding to disease-specific history and physical exam findings which were difficult to access.¹¹ Since the Roth study was published, the adoption of EHRs has grown and greater requirements are being placed on the meaningful use of data contained in EHRs,¹³ raising the prospect that such data may be increasingly available to facilitate decision making by clinicians and to construct more clinically complex performance measures.

In 2012, the American Board of Internal Medicine Foundation (ABIMF) launched the Choosing Wisely (CW) initiative, a voluntary effort by specialty societies to identify commonly-used, low-value services; each of nine specialty societies produced a list of five low-value services.^{6,14} The leaders of the CW initiative hoped that the combination of strong evidence, professional society support, and public promotion would encourage clinician-patient discussions regarding appropriate care and thereby reduce use of low-value services.^{4,6,15–17}

The CW recommendations were never intended for use as performance measures. However, the National Quality Forum's (NQF) Measure Application Partnership sees measures developed from Choosing Wisely as a near-term opportunity to address overuse of services¹⁸ and many payers are actively evaluating how to translate the CW recommendations into measures of overuse in an effort to improve affordability of health plan products. Given significant interest in advancing the development of overuse measures and the potential for leveraging more clinically-enriched data within EHRs to construct measures, our study uses the CW recommendations to illustrate the types of issues that would need to be considered to advance development of overuse concepts into e-Measures.

2. Methods

We evaluated the feasibility of translating the 45 CW recommendations into e-Measures using four criteria adapted from the NQF's proposed approach for pre-screening measure concepts: (1) the feasibility of extracting needed data elements from claims data and/or EHR systems meeting Meaningful Use (MU) Stage 2 certification standards; (2) whether the recommendation's terminology was sufficiently specific for translation into an operational e-Measure; (3) scientific evidence supporting the recommendation; and (4) impact on reducing resource use and spending.¹⁹ We assessed each recommendation against the four evaluation criteria to determine whether the recommendation met ("passed") or did not meet each criterion ("flagged"). For each criterion, we tallied the number of recommendations that were flagged. Members of the research team independently assessed each recommendation against each of the criterion and arrived at final decisions after discussing the ratings with the entire research team. Differences in assessments were resolved through group discussion and review of additional data sources/documentation. We computed Cohen's kappa statistics to measure inter-rater agreement.

The 45 CW recommendations address 34 distinct concepts, in that similar recommendations were independently issued by several different specialty societies. For example, the American Academy of Family Physicians recommendation #1 (AAFP1) and the American College of Physicians recommendation #2 (ACP2) represent the same concept because both advise against doing imaging in non-specific back pain.

We used each CW statement and information contained in the literature references cited in the CW documentation to evaluate the overuse measure concepts against the criteria.

The CW documentation for each recommendation generally referenced guidelines that cited randomized controlled trial (RCT) evidence. Where such evidence was lacking or was published before 2009, we searched for more recent evidence. For some recommendations, the specialty societies provided limited evidence on clinical efficacy, cost-effectiveness, or population prevalence; in those cases, we conducted scoping reviews to identify relevant systematic reviews, cost-effectiveness studies, and RCTs. When we could not locate formal cost-effectiveness analyses, we evaluated potential impact by estimating the size of the population affected and per episode costs.

3. Analysis

Two researchers on our study team evaluated each concept for feasibility of data capture (Criterion 1). The accessibility of data elements in structured EHR data was assessed by a clinical informaticist (DM) experienced in analysis of data from multiple MU-certified EHRs and a clinician (KS) who is an experienced EHR user. We assessed whether the potential measure's components (denominator, numerator, and exclusions) used data elements that would be present in standard claims data or in EHR systems meeting (but not exceeding) Meaningful Use Stage 2 (MU-2) certification standards.^{12,20,21} We flagged CW recommendations as infeasible for immediate e-Measure development if they required one or more important data elements beyond what might be present in claims data or EHR data systems meeting MU-2 certification standards.

Two clinicians (KS and ES) assessed each recommendation to determine whether it contained a series of clearly defined component elements that could be transformed into an unambiguous, easily-understood set of rules (Criterion 2).²² For example, the hypothetical statement, "Do not give nonsteroidal anti-inflammatory drugs to patients with heart failure" can easily be translated into a denominator (patients with heart failure) and numerator (use of nonsteroidal anti-inflammatory drugs [NSAIDs]). On the other hand, a statement advising clinicians to "avoid" prescribing NSAIDs to patients with heart failure would have the same denominator and numerator as the statement above, but the term "avoid" suggests the need for additional clinical specificity.

Two clinicians (KS and ES) assessed recommendations on the strength of their supporting scientific evidence (Criterion 3) to identify those recommendations that would likely be excluded by measure development expert panels as having weak or equivocal evidence.¹⁹ We flagged recommendations as failing to meet this criterion if they were primarily supported by an expert panel with only limited or mixed clinical evidence.

Finally, two clinicians (KS and ES) used the documentation and literature scans described above to assess the expected impact of the recommendation on improving efficiency or resource use (Criterion 4). We flagged recommendations with low impact potential—those addressing low-incidence and/or low cost interventions—that likely would be rejected by measure development panels.¹⁹

We deemed as the strongest candidates for e-Measure development those CW recommendations without serious concerns on any of the four evaluation criteria.

Inter-rater agreement for each criterion and overall was (1) feasibility: 0.90, (2) specificity: 0.62, (3) scientific evidence: 0.66, (4) impact: 0.43, and overall assessment: 0.55.

Download English Version:

<https://daneshyari.com/en/article/514920>

Download Persian Version:

<https://daneshyari.com/article/514920>

[Daneshyari.com](https://daneshyari.com)