Title

Improving Patient Safety in Public Hospitals: Developing Standard Measures to Track Medical Errors and Process Breakdowns

Journal

Journal of Patient Safety, 17(8)

ISSN

1549-8417

Authors

Ackerman, Sara L
Gourley, Gato
Le, Gem
et al.

Publication Date

2021-12-01

DOI

10.1097/pts.0000000000000480

Peer reviewed

HHS Public Access

Author manuscript

J Patient Saf. Author manuscript; available in PMC 2022 December 01.

Published in final edited form as:

J Patient Saf. 2021 December 01; 17(8): e773–e790. doi:10.1097/PTS.0000000000000480.

Improving Patient Safety in Public Hospitals: Developing Standard Measures to Track Medical Errors and Process Breakdowns

Sara L. Ackerman, PhD[1], **Gato Gourley, MSc**[2], **Gem Le, PhD**[2], **Pamela Williams**[2], **Jinoos Yazdany, MD**[3], **Urmimala Sarkar, MD**[2]

[1]Department of Social and Behavioral Sciences, School of Nursing, University of California San Francisco

[2]Division of General Internal Medicine, Department of Medicine, Zuckerberg San Francisco General Hospital, University of California San Francisco

[3]Department of Medicine, University of California San Francisco

Abstract

Objective—To develop standards for tracking patient safety gaps in ambulatory care in safety net health systems.

Methods—Leaders from five California safety net health systems were invited to participate in a modified Delphi process sponsored by the Safety Promotion Action Research and Knowledge Network (SPARKNet) and the California Safety Net Institute (SNI) in 2016. During each of the three Delphi rounds, the feasibility and validity of 13 proposed patient safety measures were discussed and prioritized. Surveys and transcripts from the meetings were analyzed to understand the decision making process.

Results—The Delphi process included eight panelists. Consensus was reached to adopt 9 out of 13 proposed measures. All 9 measures were unanimously considered valid, but concern was expressed about the feasibility of implementing several of the measures.

Conclusions—Although safety net health systems face high barriers to standardized measurement, our study demonstrates that consensus can be reached on acceptable and feasible methods for tracking patient safety gaps in safety net health systems. If accompanied by the active participation key stakeholder groups, including patients, clinicians, staff, data system professionals, and health system leaders, the consensus measures reported here represent one step towards improving ambulatory patient safety in safety net health systems.

Keywords

Quality of Care/Patient Safety (Measurement); Ambulatory/Outpatient Care; Uninsured/Safety Net Providers

Corresponding author: Dr. Sara Ackerman, Assistant Professor, University of California San Francisco, Department of Social and Behavioral Sciences, 3333 California St., Suite 455, San Francisco, CA 94118, UCSF Box 0612, San Francisco, CA 94143, (415) 476-3964, Fax (415) 476-6552, sara.ackerman@ucsf.edu.

Introduction

Patient safety, defined by the Institute of Medicine (IOM) as "the prevention of harm to patients,"[1] and by the Agency for Healthcare Research and Quality (AHRQ) as "freedom from accidental or preventable injuries produced by medical care"[2] has emerged as a primary focus of the health care quality movement.[3,4] Since the 1999 publication of IOM's widely read report, *To Err is Human*, major strides have been made in addressing individual and systemic causes of medical error.[5,6] However, patient safety research has largely focused on adverse events in hospitalized patients, while less is known about the epidemiology and causes of medical error in ambulatory (outpatient) settings.[7,8] Emerging research suggests that patient safety gaps are a significant problem in ambulatory care.[9–11] Knowledge about the types and causes of medical error in ambulatory settings is needed not only because the majority of medical care occurs on an outpatient basis, but also because the ambulatory environment differs substantially from hospital settings—suggesting the need for tailored monitoring and quality improvement efforts.[12]

Patient safety problems in ambulatory care are most often related to diagnosis, medication safety, referrals, care transitions, and testing.[8,13–15] Studies of adverse events in these areas have suggested that outpatient diagnostic errors may affect 1 in 20 U.S. adults[16] and that over 7% of patients are routinely not informed of an abnormal test result.[17] Fragmentation of care has been identified as a major cause of patient safety gaps.[18] However, medical error estimates to date across ambulatory care settings have been highly variable due to heterogeneous definitions and study methods.[19–22] Understanding and improving patient safety in ambulatory settings will require a foundation of agreed-upon definitions and measurements to assess the frequency, type and causes of medical error.

Safety net health care systems, which provide care for low-income, uninsured, and under-insured patients, may have the most to gain from the development and use of such standards. These health systems operate under resource constraints that can make medical errors and process breakdowns more likely, and their performance on existing quality measures has been worse than in other settings.[15,23–25] Understanding the relative prevalence and severity of errors and other patient safety gaps can help these health systems devise strategies to monitor gaps and improve performance.[26]

Recent adoption of electronic health records (EHRs), enabled by federal health reform and financial incentives,[17,27] has facilitated the routine generation of data that can support efforts to prevent or mitigate adverse events and improve patient safety.[28,29] We sought to leverage health information technology resources, and the input of quality improvement experts, to identify priority patient safety measures for California's public hospitals, with a long-term goal of using consensus measures to identify, understand and address patient safety gaps in ambulatory settings.

METHODS

Setting

Based at Zuckerberg San Francisco General Hospital (ZSFG) and the University of California, San Francisco, the Safety Promotion Action Research and Knowledge Network (SPARKNet) was launched in 2015 with collaborators from five publically funded health systems in California that provide services for ethnically and linguistically diverse patient populations in both urban and rural settings. SPARKNet's primary goals are to: 1) examine the epidemiology of patient safety in ambulatory care settings in the safety net, including disparities in patient safety gaps across patient populations; 2) gain insights into the root causes of medical errors and other gaps in patient safety; and 3) develop a toolkit of patient safety monitoring methods.

For the study reported here, SPARKNet partnered with the California Healthcare Safety Net Institute (SNI), a non-profit organization that provides training and assistance in quality improvement strategies and patient safety measure development for California's public hospitals and clinics. The aim was for SPARKNet collaborators to reach consensus on a set of measures to assess (a) whether patients have been notified of actionable test results and (b) whether patients with high-risk conditions are being monitored. We chose these two specific domains of safety because of extensive evidence of related safety vulnerabilities in outpatient care and evidence of subsequent harm to patients.[8,14,20,30] Data obtained with these measures could then be used to develop routine patient safety monitoring methods, identify the root causes of safety gaps, and develop quality improvement initiatives.

Delphi Consensus Process

From January through February, 2016, we used a modified Delphi process to obtain expert opinions and reach consensus on a set of patient safety measures to be used with EHR-based data in safety net health systems. The Delphi method involves multiple rounds of questionnaires in which expert opinion is first solicited, then aggregated and de-identified for use in subsequent rounds. It is important to emphasize that the Delphi approach does not aim to develop consensus through recruitment of a representative sample. Rather, it focuses on eliciting opinions from a purposive sample of participants with relevant expertise, and can be particularly helpful when evidence to support a practice or set of practices is contested or lacking.[31] The method has previously been used for the development of patient safety monitoring guidelines in ambulatory settings.[32,33]

Our three-round Delphi process began with the selection of 13 patient safety measures by the principal investigator of SPARKNet, in consultation with the Chief Medical Officer at SNI (Table 1; see Appendix B for initial list). The measures were drawn from those proposed by the National Quality Forum (NQF), and by the Public Hospital Redesign and Incentives in Medi-Cal (PRIME) program, which ties federal Medicaid funding to the achievement of metrics associated with improvements in the delivery and cost-effectiveness of care.[34]

Representatives from all five SPARKNet health systems were invited to participate in the Delphi panel. All individuals invited were responsible for PRIME implementation at their institution and/or had demonstrated expertise in patient safety measure development.

Rounds 1 and 2—Rounds 1 and 2 took place during an in-person meeting at ZSFG. The chief medical officer of SNI first explained each measure to the group, followed by a round-table discussion of each measure. For Round 1, each panelist was asked to anonymously rate the validity and feasibility of each measure on a nine-point Likert scale,[1,20] with 1 being definitely not valid/feasible and 9 being definitely valid/feasible. Validity and feasibility were defined through a set of existing questions developed for AHRQ (Center for Health Policy 2011), that were presented to panelists (Table 3). An open-ended comments section was also included for panelists to qualify their votes and/or add their own measures for discussion. During a break in the meeting, mean, minimum and maximum scores were calculated for each measure. The results were reported back to panelists to prompt discussion of the rationale for a high or low validity or feasibility score for specific measures.

After the Round 1 discussion, panelists rated the feasibility and validity for each measure a second time. The results of the second round were emailed to the group shortly after the in-person meeting, in the form of a table with each measure's validity and feasibility rankings listed, ordered by validity ranking.

Round 3—Approximately one month after Rounds 1 and 2, a one-hour conference call was held with panelists to review the results of Round 2 voting. The aim was to reach consensus on a final list of measures through discussion and consideration of concerns about measures' validity and feasibility.

RESULTS

Participants

A total of eight individuals participated in the modified Delphi process, including six SPARKNet collaborators, a nationally recognized expert in measure development based at UCSF, and SNI's chief medical officer (Table 2). Two panelists (US and JY) are co-authors of this article. Participants had a response rate of 100% (n = 8) in Round 1, 88% (n = 7) in Round 2, and 100% (n = 8) in Round 3.

Delphi Process Results

After Round 1, the panel unanimously decided to eliminate two of the 13 proposed measures because they were determined to be redundant (Table 1). Several additional measures were proposed by panelists during Round 1 but did not receive enough support to proceed to the next round (Appendix B).

In Round 2, panelists ranked 10 of 11 measures with high validity, and 6 of 11 with high feasibility, scores (7 or higher out of 9). Despite the high validity scores, panelists expressed concern about whether some measures could be interpreted and tracked in a standardized fashion. For example, one measure aimed to identify the number of individuals

on warfarin who received an abnormal international normalized ratio (INR) test and received appropriate and timely follow-up care (measure #3, retained). At least one panelist noted that standardized deployment of this measure requires a clear definition of appropriate follow-up care, and that variable definitions could undermine the validity of the measure. Limiting the definition of appropriate follow-up to a repeated INR test, a panelist explained, would enhance validity and make measurement more feasible in participating health systems.

Given the panel's consensus that all measures but one were highly valid, Round 2's discussion focused on feasibility. Panelists described the challenges of (a) identifying measures' "denominator" – or the number of patients during a defined time period who were at risk, or eligible for, the event to be measured, and (b) obtaining the data needed for specific measures at participating health systems. For example, the panel unanimously agreed that estimating the proportion of patients with chronic pain on long-term opioid therapy, and registered in Prescription Drug Monitoring Programs (PDMP) would be very difficult to implement because enrollment in PDMPs is not consistently documented at participating health systems. Panelists agreed to eliminate this measure (#8).

However, consensus on feasibility was not as easily reached for other measures, with some sites reporting more challenges obtaining necessary data than others, as well as mismatches between the importance of safety-related topics and health systems' ability to measure them. For example, systems that referred patients to multiple independent subspecialty practices anticipated difficulty tracking referral responses from outside facilities, such as mammography results. Feasibility concerns also focused on health IT infrastructure and capacity, with some sites lacking interoperability among electronic systems, making measures that incorporate two different types of data—such as laboratory data and encounter data—more resource intensive. Finally, a lack of clinician motivation to document events tied to specific measures was reported, particularly for measures that were not understood as directly linked to health outcomes (see Table 4 for panelist quotes about feasibility).

During the Round 3 discussion, panelists unanimously eliminated one measure that was ranked lowest on both feasibility and validity, and two measures that were ranked 2nd and 3rd lowest on feasibility. The panel also decided to separate one approved measure into two measures, with the goal of ensuring that all measures were consistent with those recommended by PRIME. Consensus was achieved for a final list of nine measures (Table 1). After Round 3, SPARKNet developed data extraction protocols to guide use of the patient safety measures at all five collaborating medical centers. Results from this phase of the project will be reported in a future publication.

DICUSSSION

Our modified Delphi process evaluated standardized measures that could be used to track patient safety gaps in two ambulatory care processes: 1) notifying patients of actionable test results; and 2) monitoring patients with high-risk conditions. Several rounds revealed broad consensus about the importance of nearly all proposed measures, and some disagreement about the feasibility of at least half the measures—with concerns focused on (a) the

challenges of translating an important patient safety concern into a standardizable measure and (b) IT and human resources-related barriers to producing, obtaining and sharing required data. By the final round, the panel unanimously agreed to adopt nine measures.

The consensus measures reported here represent one step towards improving ambulatory patient safety in safety net health systems. Patient safety experts have long championed better measurement as integral to improvement.[28,29] However, the proliferation of quality metrics has also added tremendous time and cost burden to health care systems, especially in safety net health systems plagued by proliferating data silos. Current electronic health records and data management infrastructure do not permit efficient measurement of clinically relevant measures. The trade-off between more feasible but "messy" measures and precise, labor-intensive measures is universal, but is particularly acute in settings with fragmented health IT systems and scant resources for additional IT personnel. These barriers are compounded by the persistent challenge of identifying measures that front-line clinicians will accept as valid and beneficial to patients. Nonetheless, as payment mechanisms in the U.S. health care system move toward an emphasis on "value" rather than "volume," participation in self-auditing to protect against payment cuts is obligatory. A strong measurement and quality improvement infrastructure may prove critical to the financial viability of these health care systems.

Although measurement is broadly assumed to be a necessary step toward higher quality medical care, reductions in medical errors and process breakdowns will not be achieved simply through standardized measurement. Indeed, the consensus measures reported here will not lead to improved patient safety without the engagement of all stakeholders: patients, clinicians, staff, data system professionals, and health system leaders. Establishing and communicating shared expectations, and identifying mismatched expectations, will be as essential as accurate measurement for understanding the reasons for safety gaps and devising strategies to mitigate them.

Efforts to transform the delivery of health care through the PRIME program point to both potential strengths and weaknesses of the performance targets developed here. The proposed targets overlap considerably with those required by PRIME, and feasibility was accounted for. Therefore, safety net health systems are likely to have built-in incentives and capacity to track their efforts to reach these targets. On the other hand, resource-limited safety net health systems may be reluctant to pursue new performance targets in an era of increasing measurement burden. Other study limitations include the small number of participating panelists, although participants represented five health systems that are broadly representative of California's safety net in terms of patient population, information technology systems, and population density.

CONCLUSION

Although the nine performance targets developed in this study were intended for use in safety net health systems, they could also be used for efforts to improve patient safety in a wider array of ambulatory settings. If found to be both feasible and valid, information about health systems' ability to meet these targets would provide important knowledge about the

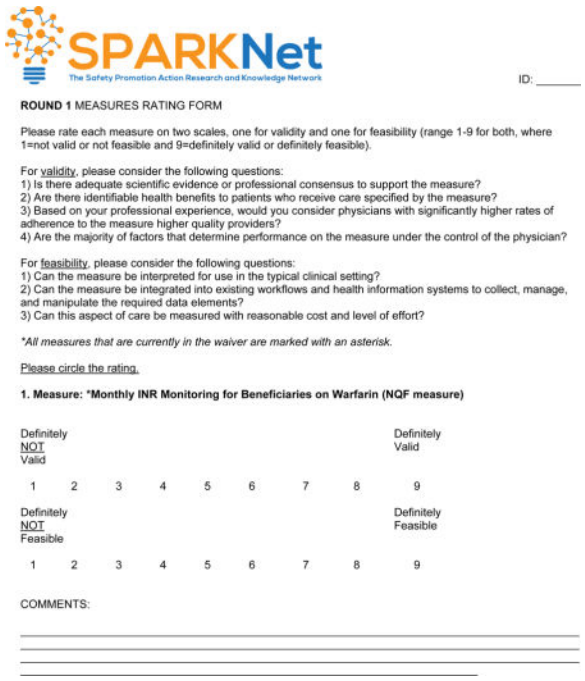
current state of outpatient safety in the U.S., as well as a foundation for testing targeted interventions to reduce medical errors and improve health outcomes.

Acknowledgments

We would like to thank the SPARKNet members, especially the California Public Hospital Systems, who generously shared their time and expertise with us.

Sources of Support: Agency for Healthcare Research and Quality Award Number: R01HS024426; Agency for Healthcare Research and Quality Award Number R24HS022047

Appendix A



SPARKNet
The Safety Promotion Action Research and Knowledge Network

ID: _____

ROUND 1 MEASURES RATING FORM

Please rate each measure on two scales, one for validity and one for feasibility (range 1-9 for both, where 1=not valid or not feasible and 9=definitely valid or definitely feasible).

For validity, please consider the following questions:
1) Is there adequate scientific evidence or professional consensus to support the measure?
2) Are there identifiable health benefits to patients who receive care specified by the measure?
3) Based on your professional experience, would you consider physicians with significantly higher rates of adherence to the measure higher quality providers?
4) Are the majority of factors that determine performance on the measure under the control of the physician?

For feasibility, please consider the following questions:
1) Can the measure be interpreted for use in the typical clinical setting?
2) Can the measure be integrated into existing workflows and health information systems to collect, manage, and manipulate the required data elements?
3) Can this aspect of care be measured with reasonable cost and level of effort?

**All measures that are currently in the waiver are marked with an asterisk.*

Please circle the rating.

1. Measure: *Monthly INR Monitoring for Beneficiaries on Warfarin (NQF measure)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

2. Measure: Proportion of patients who were on warfarin and received an abnormal INR test result

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

3. Measure: Proportion of those who were on warfarin and received an abnormal INR test result and received appropriate follow up in the appropriate time period

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

4. Measure: *Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and at least one serum potassium and a serum creatinine therapeutic monitoring test in the measurement year (NQF measure)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

5. Measure: Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and had at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

6. Measure: Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year, received at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year) and received appropriate follow-up (repeated test)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

7. Measure: *Percentage of patients age 18 years and older diagnosed with chronic pain with functional outcome goals documented in the medical record (NQF measure)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

8. Measure: *Proportion of Patients with chronic pain is on long term opioid therapy who are checked in Prescription Drug Monitoring Programs (PDMP)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

9. Measure: *Closing the referral loop: receipt of specialist report: Percentage of patients with referrals, regardless of age, for which the referring provider receives a report from the provider to whom the patient was referred.

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

10. Measure: The percentage of members 50–75 years of age who had appropriate screening for colorectal cancer. IF a patient has an abnormal test result THEN there should be evidence of receipt of appropriate follow-up for abnormal CRC screening

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

11. Measure: *Medication reconciliation - Percentage of patients aged 65 years and older discharged from any inpatient facility (e.g. hospital, skilled nursing facility, or rehabilitation facility) and seen within 30 days following discharge in the office by the physician providing on-going care who had a reconciliation of the discharge medications with the current medication list in the medical record documented.

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

12. Measure: Proportion of women 21–64 years of age received one or more Pap tests to screen for cervical cancer AND received an abnormal result (any type of abnormal result – ASCUS, HSIL, ASIL) AND evidence of appropriate follow-up (Have either a colposcopy or repeat PAP within 6 months) - (adapted from NQF 0032)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

13. Measure: Percentage of women with mammogram showing BIRADS score (see codes below) and received the recommended action taken:

- **If BIRADS not equal to 1 or 2:**
  - **BIRADS = 0 – Percent with recall for additional images or comparison with prior mammograms within 30 days**
  - **BIRADS = 3 – Percent with 6 month follow up**
  - **BIRADS = 4 or 5 – Percentage of women who received the recommended breast biopsy within 14 days**

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

New Measure: _____

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

New Measure: _____

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

SPARKNet
The Safety Promotion Action Research and Knowledge Network

ID: _____

ROUND 2 MEASURES RATING FORM

Please rate each measure on two scales, one for validity and one for feasibility (range 1-9 for both, where 1=not valid or not feasible and 9=definitely valid or definitely feasible).

For validity, please consider the following questions:
1) Is there adequate scientific evidence or professional consensus to support the measure?
2) Are there identifiable health benefits to patients who receive care specified by the measure?
3) Based on your professional experience, would you consider physicians with significantly higher rates of adherence to the measure higher quality providers?
4) Are the majority of factors that determine performance on the measure under the control of the physician?

For feasibility, please consider the following questions:
1) Can the measure be interpreted for use in the typical clinical setting?
2) Can the measure be integrated into existing workflows and health information systems to collect, manage, and manipulate the required data elements?
3) Can this aspect of care be measured with reasonable cost and level of effort?

**All measures that are currently in the waiver are marked with an asterisk.*

Please circle the rating.

1. Measure: *Monthly INR Monitoring for Beneficiaries on Warfarin (NQF measure)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

2. Measure: Proportion of patients who were on warfarin and received an abnormal INR test result

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

3. Measure: Proportion of those who were on warfarin and received an abnormal INR test result and received appropriate follow up in the appropriate time period

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

4. Measure: *Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and at least one serum potassium and a serum creatinine therapeutic monitoring test in the measurement year (NQF measure)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

5. Measure: Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and had at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

6. Measure: Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year, received at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year) and received appropriate follow-up (repeated test)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

7. Measure: *Percentage of patients age 18 years and older diagnosed with chronic pain with functional outcome goals documented in the medical record (NQF measure)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

8. Measure: *Proportion of Patients with chronic pain is on long term opioid therapy who are checked in Prescription Drug Monitoring Programs (PDMP)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

9. Measure: *Closing the referral loop: receipt of specialist report: Percentage of patients with referrals, regardless of age, for which the referring provider receives a report from the provider to whom the patient was referred.

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

10. Measure: The percentage of members 50–75 years of age who had appropriate screening for colorectal cancer. IF a patient has an abnormal test result THEN there should be evidence of receipt of appropriate follow-up for abnormal CRC screening

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

11. Measure: *Medication reconciliation - Percentage of patients aged 65 years and older discharged from any inpatient facility (e.g. hospital, skilled nursing facility, or rehabilitation facility) and seen within 30 days following discharge in the office by the physician providing on-going care who had a reconciliation of the discharge medications with the current medication list in the medical record documented.

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

12. Measure: Proportion of women 21–64 years of age received one or more Pap tests to screen for cervical cancer AND received an abnormal result (any type of abnormal result – ASCUS, HSIL, ASIL) AND evidence of appropriate follow-up (Have either a colposcopy or repeat PAP within 6 months) - (adapted from NQF 0032)

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

1 2 3 4 5 6 7 8 9

COMMENTS:

13. Measure: Percentage of women with mammogram showing BIRADS score (see codes below) and received the recommended action taken:

- **If BIRADS not equal to 1 or 2:**
  - **BIRADS = 0 – Percent with recall for additional images or comparison with prior mammograms within 30 days**
  - **BIRADS = 3 – Percent with 6 month follow up**
  - **BIRADS = 4 or 5 – Percentage of women who received the recommended breast biopsy within 14 days**

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9

Definitely NOT Feasible — Definitely Feasible

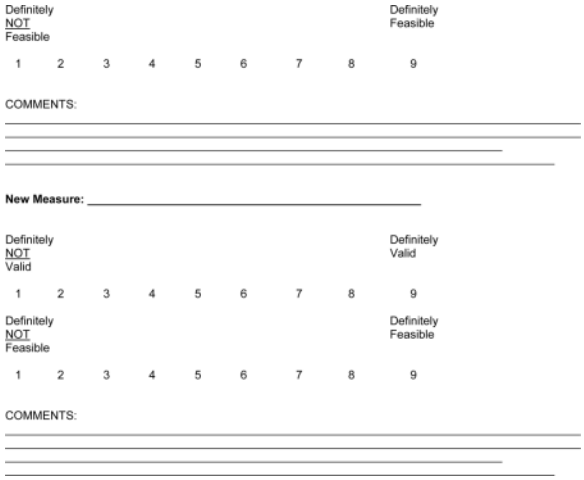
1 2 3 4 5 6 7 8 9

COMMENTS:

New Measure: _____

Definitely NOT Valid — Definitely Valid

1 2 3 4 5 6 7 8 9



Appendix B

Initially Proposed Measures

1. Monthly INR Monitoring for Beneficiaries on Warfarin (NQF measure)
2. Proportion of patients who were on warfarin and received an abnormal INR test result
3. Proportion of those who were on warfarin and received an abnormal INR test result and received appropriate follow up in the appropriate time period
4. Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and at least one serum potassium and a serum creatinine therapeutic monitoring test in the measurement year (NQF measure)
5. Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and had at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year)
6. Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year, received at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year) and received appropriate follow-up (repeated test)
7. Percentage of patients age 18 years and older diagnosed with chronic pain with functional outcome goals documented in the medical record (NQF measure)
8. Proportion of Patients with chronic pain is on long term opioid therapy who are checked in Prescription Drug Monitoring Programs (PDMP)

9. Closing the referral loop: receipt of specialist report: Percentage of patients with referrals, regardless of age, for which the referring provider receives a report from the provider to whom the patient was referred.

10. The percentage of members 50–75 years of age who had appropriate screening for colorectal cancer. IF a patient has an abnormal test result THEN there should be evidence of receipt of appropriate follow-up for abnormal CRC screening

11. Medication reconciliation - Percentage of patients aged 65 years and older discharged from any inpatient facility (e.g. hospital, skilled nursing facility, or rehabilitation facility) and seen within XX days following discharge in the office by the physician providing on-going care who had a reconciliation of the discharge medications with the current medication list in the medical record documented.

12. Proportion of women 21–64 years of age received one or more Pap tests to screen for cervical cancer AND received an abnormal result (any type of abnormal result – ASCUS, HSIL, ASIL) AND evidence of appropriate follow-up (have either a colposcopy or repeat PAP within 6 months) - (adapted from NQF 0032)

13. Percentage of women with mammogram showing BIRADS score (see codes below) and received the recommended action taken. If BIRADS not equal to 1 or 2:

    - BIRADS = 0 – Percent with recall for additional images or comparison with prior mammograms within 30 days
    - BIRADS = 3 – Percent with 6 month follow up
    - BIRADS = 4 or 5 – Percentage of women who received the recommended breast biopsy within 14 days

Additional Measures Proposed by Delphi Participants

1. Patients on diuretic: f/u of abnormal sodium and creatinine
2. Annual EKG monitoring for corrected QT interval in patients on specific drugs (i.e. methadone)
3. Completed safety check list before pre-specified, high-risk drug dispensation in sub-specialty clinics
4. Documentation of medication in EMR for children in foster care
5. Documentation of high-cost medication in EMR

References

1. Aspden, P, Corrigan, JM, Wolcott, J, Erickson, SM. Patient Safety: Achieving a New Standard for Care. Washington (DC): National Academies Press (US); 2004. Copyright 2004 by the National Academy of Sciences. All rights reserved
2. AHRQ. Chapter 3 Patient Safety | AHRQ Archive. 2011. 2017a

3. Becher EC, Chassin MR. Improving quality, minimizing error: making it happen. Health Aff (Millwood). 2001; 20 (3) :68–81.
4. IOM. To Err is Human: Building a Safer Health System. Summary - To Err is Human 1999 report 274 brief.pdf. 2017b. http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/1999/To-Err-is-Human/To%20Err%20is%20Human%201999%20%20report%20brief.pdf
5. Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. Jama. 1998; 280 (15) :1311–1316. [PubMed: 9794308]
6. Lee GM, Kleinman K, Soumerai SB, et al. Effect of nonpayment for preventable infections in U.S. hospitals. N Engl J Med. 2012; 367 (15) :1428–1437. [PubMed: 23050526]
7. Gandhi TK, Lee TH. Patient safety beyond the hospital. N Engl J Med. 2010; 363 (11) :1001–1003. [PubMed: 20825311]
8. Sarkar U, Wachter RM, Schroeder SA, Schillinger D. Refocusing the lens: patient safety in ambulatory chronic disease care. Jt Comm J Qual Patient Saf. 2009; 35 (7) :377–383. 341. [PubMed: 19634806]
9. Lorincz, C; Drazen, E; Sokol, PE; , et al. Research in Ambulatory Patient Safety 2000 2010: A 10-Year Review. 2011. https://psnet.ahrq.gov/resources/resource/23742
10. Panesar SS, deSilva D, Carson-Stevens A, et al. How safe is primary care? A systematic review. BMJ Qual Saf. 2016; 25 (7) :544–553.
11. Shekelle, PG, Sarkar, U, Shojania, K. , et al. Patient Safety in Ambulatory Settings. Rockville (MD): Agency for Healthcare Research and Quality (US); 2016. AHRQ Comparative Effectiveness Technical Briefs.
12. Wachter RM. Is ambulatory patient safety just like hospital safety, only without the “stat”? Ann Intern Med. 2006; 145 (7) :547–549. [PubMed: 17015874]
13. Bell BG, Campbell S, Carson-Stevens A, et al. Understanding the epidemiology of avoidable significant harm in primary care: protocol for a retrospective cross-sectional study. BMJ Open. 2017; 7 (2) :e013786.
14. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Ann Intern Med. 2006; 145 (7) :488–496. [PubMed: 17015866]
15. Roy CL, Poon EG, Karson AS, et al. Patient safety concerns arising from test results that return after hospital discharge. Ann Intern Med. 2005; 143 (2) :121–128. [PubMed: 16027454]
16. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf. 2014; 23 (9) :727–731.
17. Casalino LP, Dunham D, Chin MH, et al. Frequency of failure to inform patients of clinically significant outpatient test results. Archives of internal medicine. 2009; 169 (12) :1123–1129. [PubMed: 19546413]
18. WHO. WHO | World Alliance for Patient Safety. 10 (29) :43. 2012;
19. Ranji, S. Measuring and Responding to Deaths From Medical Errors | AHRQ Patient Safety Network. 2016. https://psnet.ahrq.gov/perspectives/perspective/221/measuring-and-responding-to-deaths-from-medical-errors?q=ranji
20. IOM. DiagnosticError_ReportBrief.pdf. 2017a. http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2015/Improving-Diagnosis/DiagnosticError_ReportBrief.pdf
21. AHRQ. Measurement of Patient Safety | AHRQ Patient Safety Network. 2017c. https://psnet.ahrq.gov/primers/primer/35/measurement-of-patient-safety. Accessed June 1, 2017
22. AHRQ. Patient Safety in Ambulatory Care | AHRQ Patient Safety Network. 2017b. https://psnet.ahrq.gov/primers/primer/16/patient-safety-in-ambulatory-care. Accessed June 1, 2017, 2017
23. Metersky ML, Hunt DR, Kliman R, et al. Racial disparities in the frequency of patient safety events: results from the National Medicare Patient Safety Monitoring System. Med Care. 2011; 49 (5) :504–510. [PubMed: 21494115]
24. Wallace E, Lowry J, Smith SM, Fahey T. The epidemiology of malpractice claims in primary care: a systematic review. BMJ Open. 2013; 3 (7)

25. Blagev DP, Lloyd JF, Conner K, et al. Follow-up of incidental pulmonary nodules and the radiology report. J Am Coll Radiol. 2014; 11 (4) :378–383. [PubMed: 24316231]
26. Zwaan L, Schiff GD, Singh H. Advancing the research agenda for diagnostic error reduction. BMJ Qual Saf. 2013; 22 (Suppl 2) :ii52–ii57.
27. Bhasale AL, Miller GC, Reid SE, Britt HC. Analysing potential harm in Australian general practice: an incident-monitoring study. Med J Aust. 1998; 169 (2) :73–76. [PubMed: 9700340]
28. AHRQ. Common Formats | AHRQ Patient Safety Organization Program. 2017d. https://pso.ahrq.gov/common. Accessed June 1, 2017
29. Boston MANPSF. Free From Harm: Accelerating Patient Safety Improvement Fifteen Years After To Err Is Human. AHRQ Patient Safety Network; 2015. https://psnet.ahrq.gov/resources/resource/29554
30. Singh H, Thomas EJ, Mani S, et al. Timely follow-up of abnormal diagnostic imaging test results in an outpatient setting: are electronic medical records achieving their potential? Archives of internal medicine. 2009; 169 (17) :1578–1586. [PubMed: 19786677]
31. Fernandes O, Gorman SK, Slavik RS, et al. Development of clinical pharmacy key performance indicators for hospital pharmacists using a modified Delphi approach. Ann Pharmacother. 2015; 49 (6) :656–669. [PubMed: 25780250]
32. Mull HJ, Nebeker JR, Shimada SL, Kaafarani HM, Rivard PE, Rosen AK. Consensus building for development of outpatient adverse drug event triggers. J Patient Saf. 2011; 7 (2) :66–71. [PubMed: 21587117]
33. Tjia J, Field TS, Garber LD, et al. Development and pilot testing of guidelines to monitor high-risk medications in the ambulatory setting. Am J Manag Care. 2010; 16 (7) :489–496. [PubMed: 20645664]
34. PRIME. Public Hospital Redesign and Incentives in Medi-Cal (PRIME) - California Association of Public Hospitals and Health Systems. 2017

Table 1

Patient Safety Measures Under Consideration

Measures for which Consensus was Reached		**Validity Score Rounds 1 + 2 (1–9 scale)**		**Feasibility Score Rounds 1 + 2 (1–9 scale)**		**Final Vote on Inclusion**
1. Monthly INR Monitoring for individuals on Warfarin (NQF measure)		8.30	8.43	7.63	7.86	YES
3. Proportion of those who were on warfarin and received an abnormal INR test result and received appropriate follow up in the appropriate time period		8.13	8.57	6.80	7.23	YES
4. Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and at least one serum potassium and a serum creatinine therapeutic monitoring test in the measurement year (NQF measure)		7.38	7.86	7.13	7.57	YES
6. Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year, received at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year) and received appropriate follow-up (repeated test)		7.63	7.86	6.13	7.43	YES
9. Closing the referral loop: receipt of specialist report: Percentage of patients with referrals, regardless of age, for which the referring provider receives a report from the provider to whom the patient was referred		8.80	8.86	7.00	7.71	YES
10. The percentage of members 50–75 years of age who had appropriate screening for colorectal cancer.*	The percentage of members 50–75 years of age who had appropriate screening for colorectal cancer. IF a patient has an abnormal test result THEN there should be evidence of receipt of appropriate follow-up for abnormal CRC screening.*	8.38	8.86	7.50	8.29	YES (Converted to two-part measure)
13. BIRADS = 4 or 5 – Percent who received the recommended breast biopsy within 14 days.*	Percentage of women with mammogram showing BIRADS score and received the recommended action taken. If BIRADS not equal to 1 or 2: BIRADS = 0 – Percent with recall for additional images or comparison with prior mammograms within 30 days BIRADS = 3 – Percent with 6 month follow up*	8.25	8.29	5.38	5.57	YES (Converted to two-part measure)
Measures Eliminated after Discussion						
2. Proportion of patients who were on warfarin and received an abnormal INR test result		No vote – measure considered redundant				
5. Percentage of patients 18 years of age and older who received a least 180 treatment days of ambulatory medication therapy for a select therapeutic agent (ACE inhibitors) during the measurement year and had at least one abnormal test result (serum potassium and a serum creatinine therapeutic monitoring test in the measurement year)		No vote – measure considered redundant				
7. Percentage of patients age 18 years and older diagnosed with chronic pain with functional outcome goals documented in the medical record (NQF measure)		4.50	4.00	2.75	2.57	NO
8. Proportion of Patients with chronic pain is on long term opioid therapy who are checked in Prescription Drug Monitoring Programs (PDMP)		8.30	7.86	4.38	4.57	NO
11. Medication reconciliation - Percentage of patients aged 65 years and older discharged from any inpatient facility (e.g. hospital, skilled nursing facility, or rehabilitation facility) and seen within 30 days following discharge in the office by the physician providing on-going care who had a reconciliation of the discharge medications with the current medication list in the medical record documented		7.63	7.43	6.38	4.86	NO
12. Proportion of women 21–64 years of age received one or more Pap tests to screen for cervical cancer AND received an abnormal result (any type of abnormal result – ASCUS, HSIL, ASIL) AND evidence of appropriate follow-up (Have either a colposcopy or repeat PAP within 6 months) - (adapted from NQF 0032)		8.13	8.23	5.38	5.79	NO

J Patient Saf. Author manuscript; available in PMC 2022 December 01.

Patient safety measures considered during Delphi process. (INR) Abnormal international normalized ratio; (ACE) Angiotensin Converting Enzyme; (CRC) Colorectal Cancer Screening; (BIRADS) Breast Imaging Reporting and Data System

Table 2

Delphi Participants

Characteristics of Panelists	n=8
Position	
Special Projects Manager	1
Director, Quality/Risk/Patient Safety[2]	1
Ambulatory Care Medical Director[2]	1
Chief Medical Officer	1
Chief Administrative Officer, Ambulatory Services	1
Associate Professor/General Medicine Clinician[1,2]	1
Associate Professor/Rheumatology Clinician[1]	1
Assistant Professor	1
Academic degrees obtained	
MBA	1
MPH	2
PhD	1
MD/DO	7

[1]Also co-author of this article.

[2]Practicing primary care clinician.

Table 3

Validity and feasibility criteria

Validity	1) Is there adequate scientific evidence or professional consensus to support the measure? 2) Are there identifiable health benefits to patients who receive care specified by the measure? 3) Based on your professional experience, would you consider physicians with significantly higher rates of adherence to the measure higher quality providers? 4) Are the majority of factors that determine performance on the measure under the control of the physician?
Feasibility	1) Can the measure be interpreted for use in the typical clinical setting? 2) Can the measure be integrated into existing workflows and health information systems to collect, manage, and manipulate the required data elements? 3) Can this aspect of care be measured with reasonable cost and level of effort?

SPARKNet measure validity and feasibility criteria considered.

Table 4

Feasibility Concerns

Theme	Illustrative Quote*
Balancing importance of safety-related issue with measurability	*"… one thing that often is important to consider just right at the outset is what the typical data streams are across projects… for example, are there integrated laboratory systems that can be queried across the entire network?"* *"Is the concept important and is it measureable?"* *"All measures should aspire to be electronically reported…"* 9: "*How can we fully measure closing the referral loop? What if a referral email was sent to the physician, but the physician never read it? Is sending the referral email enough to measure closing the referral loop? How would we determine whether the physician actually read the referral email or not? This is too hard to know for sure, so the best we can do is to document that the referral email was sent."* 13: *"…abnormal vs. normal is a discrete value, but numbers are hard to document"* 8: *"Chronic pain measure requires tracking of medications … dispensing data is hardest; prescribing data is also hard."*
System-level barriers to obtaining needed data	*"Looking in claims or EMR for these data, some may fall out because they cannot be uniformly pulled out across systems."* *"There is such a spectrum of systems in place. Especially on the EHR side, some folks who have been on EHR for 10 years and others for one reason or another are all on paper or transitioning. Even those on EHR don't have these measures built in or have up to 60 different informatics systems to pull from. … can this even be done in a consistent way?"* 10: *"The main barrier is outside colonoscopies and getting result into internal EHR."*
Clinician resistance to collecting data that are perceived as not directly linked to health outcomes	7: *"I think other things to consider with some of these monitoring measures, you know, I have physicians screaming at me often because they don't think these things are that important and we invest a large amount of money into them; for example, can we actually produce the percent of patients with adverse events related to these drugs and have any data showing that monitoring impacts those episodes? …the closer measures are to outcomes, the more likely [physicians] are to participate…[it] we have to think about will it grab people's attention and get people interested"* 7: *"Their [clinicians'] perspective is they want to move closer to outcomes. Their perspective is they are interested in documenting goals … there has been no testing or proof that anyone can do this in clinical practice because they could not convince if it was useful … [It's a] high risk measure…so many confounding factors; it is really, really challenging."*

Feasibility concerns discussed during Delphi consideration process.

*Quotes that are labeled with a number are specific to the following proposed measures: 7-Chronic pain; 8-opiod; 9-Referral; 10-CRC; 13-BIRADS