WORKING
PAPER

Prioritizing Patient Safety Outcomes Measures:

Results of an Expert Consensus Process

DONNA O. FARLEY, MICHAEL D. GREENBERG,
AMELIA M. HAVILAND, SUSAN LOVEJOY

WR-601-AHRQ

July 2008

Prepared for the Agency for Healthcare Research and Quality

HEALTH

PRIORITIZING PATIENT SAFETY OUTCOMES MEASURES: RESULTS OF AN EXPERT CONSENSUS PROCESS

Donna O. Farley, Michael D. Greenberg, Amelia M. Haviland, Susan Lovejoy

RAND Patient Safety Evaluation Center
The RAND Corporation

July 2008

This work was supported by contract No. 290-02-0010 from the
Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services.

PREFACE

In 2000, the U.S. Congress mandated the Agency for Healthcare Research and Quality (AHRQ) to take a leadership role in helping health care providers reduce medical errors and improve patient safety. As part of its patient safety research and development initiative, AHRQ contracted with RAND September 2002 to serve as its patient safety evaluation center. The evaluation center is responsible for performing a longitudinal, formative evaluation of the full scope of AHRQ's patient safety activities, and for providing regular feedback on process and outcomes to support continuing improvement of the initiative over the evaluation period.

This report presents the results of a modified Delphi consensus process carried out as part of the patient safety evaluation, with the purpose of identifying priority patient safety outcome measures for use in monitoring and evaluating progress in improving patient safety. The Delphi process was conducted in 2006, with the participation of 47 individuals who are national clinical and research experts in patient safety issues.

Through this process, the participating experts identified three groups of important patient safety outcome measures for inpatient care, and they also identified measures addressing safety outcomes in ambulatory care and long-term care settings. The identification of these measures represents the first step in a larger process, with needed follow-up work identified for further refinement of measures, decisions regarding which measures should be used for differing applications, and additional validation research to ensure the measures are documented to be valid for the relevant applications.

The contents of this report will be of interest to national and state policy makers, health care organizations and clinical practitioners, patient advocacy organizations, health researchers, and others with responsibilities for ensuring that patients are not harmed by the health care they receive.

This work was sponsored by the Agency for Healthcare Research and Quality, Department of Health and Human Services, under contract No. 290-02-0010, for which James B. Battles, Ph.D. serves as project officer.

This work was conducted in RAND Health, a division of the RAND Corporation. A profile of RAND Health, abstracts of its publications, and ordering information can be found at www.rand.org/health.

TABLE OF CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

We gratefully acknowledge the participation of the many individuals who participated in the modified Delphi process for identification of priority patient safety outcome measures. More than 60 individuals responded to our initial invitation to participate in the process, of whom 47 participated in the rating of the importance of candidate measures. The perseverance of those who continued throughout the entire process has helped to ensure that the important patient safety outcome measures identified reflect a strong consensus among experts in the field. Through their active engagement in both the rating process and the final conference call discussions, we also were able to identify a wide range of issues involved in measuring patient safety outcomes effectively.

Our AHRQ project officer, James Battles, has been instrumental in guiding the conceptual formation and execution of the evaluation, including the identification of this consensus process as a priority for identifying outcome measures that are priorities for use in our outcome evaluation work. We also thank our RAND colleagues Scott Ashwood, Brian Carroll, Stacy Fitzsimmons, and Michelle Platt for their indispensable contributions to the successful conduct of the modified Delphi process and the assembly of materials to support that process.

SECTION 1. INTRODUCTION

Since 2000, the Agency for Healthcare Research and Quality (AHRQ) has served as the lead agency in carrying out the national patient safety research and implementation initiative funded by the U.S. Congress. AHRQ has undertaken a comprehensive strategy for supporting expansion of knowledge about patient safety epidemiology and effective practices, and identifying and disseminating the most effective practices for use in the U.S. health care system.

AHRQ contracted with RAND in September 2002 to serve as the evaluation center for its patient safety initiative. The evaluation center is responsible for performing a longitudinal evaluation of the full scope of AHRQ's patient safety activities and providing regular feedback to support the continuing improvement of this initiative. The evaluation also has used both qualitative and quantitative methods to assess overall initiative impacts, outcomes, and the diffusion of adoption of safe practices by health care providers.

The patient safety evaluation team has been assessing trends in patient safety measures in the product evaluation – the analysis of the effects of the patient safety initiative on stakeholders, safety practices, and patient outcomes. Ideally, this evaluation should be able to work with an already established, comprehensive set of outcome measures to estimate effects of the patient safety initiative. Instead, we are using a limited number of available measures, while also "pushing the envelope" by testing analyses of other newly developed measures. Thus far, we have performed trend analyses on several of the AHRQ patient safety indicators (PSI), as well as selected measures developed by two of the patient safety reporting demonstration projects.

The absence of a comprehensive set of national patient safety outcome measures that are generally accepted in the health care community continues to hinder the ability to track patient safety performance. Numerous sets of measures have been developed by various agencies and private organizations, including AHRQ. However, little progress has been made thus far in achieving an agreed-upon set of safety measures for use in a national system for monitoring patient safety performance. In addition, virtually all of the measures established to date apply to the hospital inpatient care setting. There are relatively few safety measures for long term care and no measures for ambulatory care.

The NQF consensus process for review and endorsement of measures offers the potential to move toward establishment of a generally accepted set of health care measures. However, the only patient safety measures that NQF has addressed thus far are the 27 serious reportable events that are intended to form the basis for a mandatory state-based reporting systems. The Hospital Quality Alliance and Ambulatory Quality Alliance also have adopted measures using similar processes, but these measures focus primarily on quality of care.

USE OF A DELPHI PROCESS TO IDENTIFY PATIENT SAFETY MEASURES

The Delphi panel technique is a method for obtaining expert consensus ratings or forecasts, by drawing on input from a panel of independent experts over two or more rounds of ratings. Experts are provided with materials to rate or evaluate in each round of the process. After each round, a facilitator provides anonymous feedback to the experts regarding the group's ratings and comments. The object of a Delphi process is to allow consensus in ratings to emerge among the experts, based on feeding back their anonymous responses over successive rounds of the process.

Typically, several iterative rating rounds are conducted for a given set of issues, until the participants' ratings reach an acceptable level of consensus.

To address the need for clear identification of priority patient safety outcome measures, a modified Delphi process was conducted in 2006 as part of our evaluation work, with the goal of identifying measures that would be candidates for national patient safety outcome measures. The stated aim of this Delphi process was to identify valid patient safety measures that address the most important types of adverse events for patients, and to assess how ready the measures are for use in a national monitoring system. The measures identified are intended to be used for the following purposes:

- To evaluate the impact of the national patient safety initiative on patient outcomes
- To be included in the national network of patient safety databases as part of the Patient Safety Organization program established by the Patient Safety and Quality Improvement Act of 2005
- To be tracked in an ongoing, national-level monitoring system on patient safety performance and issues
- To be used by providers in their local safety monitoring and performance benchmarking.

The participants in this modified Delphi process were the lead researchers for the AHRQ-funded patient safety projects, including the FY 2000-01 projects, Challenge grants, Partnerships in Patient Safety grants, and health information technology value grants. These individuals were invited to participate and to suggest others who also could provide expertise to the measure rating process. The 47 individuals who participated through the importance rating step are listed in Appendix A.

The process consisted of the following four steps, each of which is described further in Section 2 of this report:

Round 1 – Participants reviewed a draft list of candidate patient safety measures, commented on the measures, and suggested others to add to the list.

Round 2 – Participants rated each candidate measure based solely on the importance of the patient safety issue it represents.

Round 3 – Participants rated the validity of each measure in the subset of measures identified as highly or moderately important, taking into consideration available evidence on the measure's validity and measurability.

Round 4 – A teleconference was held at the end of the process, for group discussion by the Delphi participants of the ratings generated and issues raised during the consensus process.

The results of the two sets of ratings on importance and validity were used to identify three sets of measures for the inpatient setting: (1) a set of important and well-validated measures that are close to ready for use, (2) a set of other highly important measures that are not yet well validated, and therefore, are priorities for additional validation research, and (3) other important measures that are not well validated. In addition, Delphi participants identified candidate measures for patient safety outcomes in the ambulatory care and long term care settings, almost

all of which require substantial additional development and testing before being ready for use in monitoring.

The RAND evaluation team framed the rating methodology and the Delphi process in which it was used, and we also provided technical and administrative support for participants as the Delphi process was carried out. This included compilation of initial candidate measures, preparation of rating forms, provision of background documentation on candidate measures, compilation and reporting of rating results after each round, and facilitation of the web meeting discussion. Results and comments from the participants were fed back to them following each step of the process.

ORGANIZATION OF THIS REPORT

The remainder of this report describes the methods that were employed in conducting the expert consensus process and presents the results that emerged from the process, including identification of three groups of outcome measures judged important that are at differing stages of validation and, therefore, readiness for use in the field. Section 2 describes in some detail the methodology we used, including the preparation and instructions to the Delphi participants for each of four steps carried out. Section 3 presents the results of the ratings of importance for the candidate measures identified for consideration, as well as the results of the validity ratings for those measures that emerged from the previous rating step as being important, i.e., representing important patient safety issues. Finally, Section 4 provides a synthesis of the overall results of both rating steps, identifies the three groups of important measures that emerged from the Delphi process, and discusses a number of issues that the expert participants identified as requiring attention before any of the measures can be used with confidence for national monitoring or other purposes.

The report includes a number of appendices that identify the participants in the Delphi process and provide backup detail on the results of each step and comments submitted by the participants. All of these items are provided to ensure that AHRQ has full information on the many specific thoughts, concerns and suggestions offered by the Delphi participants, for consideration as work continues on the refinement of these and other patient safety outcome measures.

SECTION 2. METHODS

The Delphi process builds consensus among experts on a topic by using an iterative process in which individual judgments are elicited from participants in two or more rating rounds, with the results of each round being provided to the individual participants for consideration in subsequent cycles of independent judgments. To elicit expert judgments on candidate patient safety outcomes measures, we undertook a modified Delphi process.

The process we used was a "modified" Delphi process because (1) the expert participants performed only one round of ratings of the importance of each of a set of candidate measures, rather than iterative rounds, and (2) the process ended with a teleconference among participants to enable them to discuss the rating results and their experiences in the process. Only one round of importance ratings was necessary because we were seeking to identify a limited number of the most important measures from a large set of candidate measures; we selected only those measures that received very high median importance ratings that reflected the strongest agreement among the participants We included the teleconference group discussion in the process because we felt it would enable the participants to identify areas of agreement and disagreement with respect to key methodological and implementation issues raised, in addition to their judgments regarding the patient safety outcome measures that emerged from their work.

We engaged the participants in a four-step modified Delphi process, each of which is described in the remainder of this section:

1. Participating experts were presented with an initial list of candidate patient safety outcomes measures and asked them to comment and to nominate additional measures for consideration.
2. Participants were asked to rate the measures on the criterion of the importance of each candidate measure, working with an expanded and revised list of measures that we provided them based on feedback from the first round.
3. For a subset of the candidate measures that the participants had rated has important, they were asked to rate each measure on the criterion of how well validated it was.
4. A conference call was convened with the participants, in which we sought their reactions to, and comments on, the results of the Delphi process.

INITIAL PREPARATION

With the goal of engaging in the Delphi process as many as possible of the leaders in patient safety practice and research in the U.S., we decided to invite the broad set of principal investigators of patient safety projects funded by AHRQ to participate. These included leaders of the patient safety projects funded in FY 2000 and FY 2001, the Challenge grants, Health Information Technology Value grants, and the Partnerships in Patient Safety grants.

We sent e-mails to the primary investigators of each project, which described our plans for the Delphi process and invited them to participate. We also asked them to nominate other patient safety experts, whom we then invited to participate in a second wave of e-mails. A total of 92 individuals agreed to participate. These included 73 out of 138 principal investigators contacted directly, plus 19 of 37 additional people they nominated to participate.

We then compiled an initial list of 174 candidate measures, drawing from several major taxonomies of patient safety outcomes. These included the Patient Safety Indicators (PSI) developed by the Agency for Healthcare Research and Quality, the National Quality Forum Serious Reportable Adverse Events in Healthcare (NQF), the National Surgical Quality Improvement Program measures (NSQIP), the New York Patient Occurrence Reporting and Tracking System categories (NYPORTS), the Medicare Patient Safety Monitoring System measures (MPSMS), and the Utah/Missouri Adverse Event Classes (UT/MO). The initial candidate list included measures based on data from chart abstraction, adverse event reporting systems, and health care claims. We also added several measures based on concepts developed by staff at RAND. For a small number of measures that were included in more than one of the source taxonomies in our initial candidate list, we dropped the duplicates.

ROUND 1: ESTABLISHING SETS OF CANDIDATE MEASURES

It was important to ensure that the candidate measures subjected to rating represented key patient safety issues, without being constrained by current data availability. All the candidate measures that RAND identified were based on available data, which often were not nationally representative. Therefore, we elicited suggestions from the Delphi participants for additional measures or measure concepts that covered safety issues not addressed by any of the original candidate measures. Data for many of the candidate measures they suggested indeed were not currently available. Therefore, eventual success in using those new measures identified as important will depend on the feasibility of obtaining the data necessary to estimate them and to assess their validity.

In the first Delphi round, we sent an e-mail to the participants in which we distributed the list of candidate measures (including definitions for each measure) and asked them to do two things:

1. Review and comment on the measures in the candidate list; and
2. Offer suggestions for any additional measures, or sets of measures, that they felt should be added to the candidate list, in particular measures related to patient safety outcomes for HIT-enabled care, ambulatory care, and long-term care.

Participants were given 2 weeks in which to complete this task, and were sent two reminder e-mails prior to the deadline.

Feedback on this task was provided by 52 participants, either offering comments regarding the measures in the initial list or suggesting additional measures to add to the list. Several comments addressed the way in which RAND organized and presented the candidate list of measures; others responded to assumptions that RAND made in describing particular measures. See Appendix B for the participants' comments on the initial list of measures.

We incorporated many of the participants' suggestions in revising and reformatting the list. We also added a number of new measures to the candidate list based on the participants' suggestions. Our goal in this stage of the process was to be inclusive in expanding the list of candidate measures based on the suggestions of the Delphi participants.

ROUND 2: RATINGS OF IMPORTANCE FOR CANDIDATE MEASURES

The second Round of the Delphi process was begun approximately 3 weeks after completing the first Round, with e-mail contact to the 67 individuals who participated in Round 1 or expressed interest in continuing to participate in the ratings. The Round 2 e-mail contained instructions for completing ratings of the importance of each candidate measure. Attached to the e-mail were (1) a revised list of candidate measures with supporting documentation on measure specifications, and (2) full feedback of the participants' comments from Round 1, with no identification of the individuals submitting each comment, to retain anonymity. The full set of the candidate measures and supporting definitions and references is available upon request in a separate document.

The final list of candidate measures consisted of 501 measures, including 118 measures suggested by the Delphi participants.[1] The new measures added came from a number of sources and taxonomies, including the Nursing Sensitive Performance Measures (NSPM), the Society of Thoracic Surgeons Measures (STS), the National Quality Forum Endorsed Mortality Measures (NQFM), the Wisconsin Medical Injury Prevention Program measures (WI), the National Quality Forum Endorsed Home Health Care Measures (HHC), National Quality Forum Endorsed Nursing Home Measures (NH), and measures from the Surgical Care Improvement Project (SC). The Delphi participants also suggested a number of additional "measurement concepts" – ideas for patient safety outcomes measures that were not formally defined in any existing taxonomy. These measurement concepts were also included in the candidate measures.

In Round 2, the participants were asked to rate the importance of each of the measures on a 9 point scale, where 1 = not important and 9 = most important. Participants were given the following instructions regarding how they should consider the concept of "importance" for a patient safety outcome measure:

> *For this round of the Delphi process, we would like you to rate each of the candidate measures based solely on its ability to represent an important patient safety outcome. "Important" outcomes are those that strongly involve some combination of the following features:*
>
> - *Health care origin: An adverse effect or event resulting from health care that should not occur with effective systems and practices.*
> - *Patient Harm: A strong likelihood for harm to patients*
> - *Frequency: An adverse outcome that is at risk of occurring frequently.*
>
> *In rating the importance of the candidate measures, please focus on the basic outcome concept that is represented by each measure. Additional measurement issues, of which there are many (as highlighted in participants' comments), can be addressed later, specifically for measures that represent the most important outcome concepts identified by the Delphi process.*
>
> *Please rate the importance of each candidate measure based on your own judgment regarding: (1) How important the patient outcome concept is (see our criteria above for considering importance), as well as (2) the relative strength of each of the similar measures in representing that concept. If similar measures receive equal importance ratings, they will be retained if rated*

---

[1] Of the 501 measures in the final candidate measure set, some were duplicates of measures developed for a specific setting (usually inpatient care), as Delphi participants identified their applicability in other settings. These measures are considered to be concepts in the new settings that differ from the source measures because modifications are likely to be required to adapt the measures precisely to new setting(s).

highly, and selection among similar measures will be done in later rating or discussion stages of the process.

Participants were asked to complete their importance ratings using a web-based survey instrument, but they were also given the option of completing a hardcopy version of the instrument. Respondents were also invited to comment on the measures and share their opinions regarding this rating task.

For ease of navigation on the survey instrument, the 501 candidate measures were divided into several content groups based on relevant settings (e.g., Inpatient Surgical Measures, Inpatient Non-Surgical Measures, Ambulatory Measures, Long Term Care Measures), and then into subordinate functional areas within the groups (e.g., cardiac, pulmonary, etc.).[2] Respondents were invited to choose which content groups and functional areas they wanted to rate for importance, and they were asked to try to rate all measures within each of the groups and functional areas that they chose.[3]

Respondents were given 3 weeks to complete their importance ratings. A total of 47 participants rated importance for some or all of the candidate measures.

ROUND 3: RATINGS OF VALIDITY FOR IMPORTANT MEASURES

The third Round of the Delphi process was begun approximately one month after completing the second Round. We sent an e-mail to the 47 individuals who participated in the importance ratings in Round 2, which provided feedback on the Round 2 results along with instructions for performing the validity ratings for measures identified as "important" patient safety outcome measures in the ratings performed in Round 2.

The Round 2 results, which were attachments to the Round 3 e-mail, included (1) median importance ratings for the individual candidate measures, (2) summary statistics on the distributions of the importance ratings across groups of measures, and (3) a compilation of the participants' comments from the importance rating process. Summary statistics for the importance ratings are presented in the Results Section below.

Based on the importance rating results from Round 2, 106 of the 501 candidate measures were identified as important patient safety outcome measures, and the validity ratings were performed for these measures. Details of the methods and cutoff points used to identify these 106 measures are presented in the Results Section below.

In Round 3, the participants were asked to rate the validity of each of the measures on a 9 point scale, where 1 = not valid or not validated and 9 = highly valid. For measures in the long-term care and ambulatory content groups (for which little empirical validation literature was

---

[2] Note that we divided the measures into these groupings based partly on their definitions and on common sense, and partly based on the input we received in Round 1 from the Delphi participants.

[3] A number of candidate measures were included in *more than one content group* on the survey form. Some measures, for example, were separately included for rating both in the Inpatient Non-Surgical group, and in the Ambulatory group. Both for rating and statistical purposes, we dealt independently with measures that were included in more than one group in the survey form, even where those measures otherwise shared the same definition. Thus, although the Revised List contained 292 unique measures based on their definitions, there were 501 distinct measures once we allocated each measure to relevant content groups (and corresponding clinical settings).

available), and for "measure concepts" (for which no empirical validation literature was available), we offered the option for participants to designate a measure as "not ready for rating."

Extensive supporting documentation summarizing the existing scientific literature on validity (or lack thereof) for each of the measures was provided in attachments to the validity rating instructions in the Round 3 e-mail (See Appendix C). The measures identified as important include three distinct types of measures – event reporting measures, measures based on administrative data, and measures based on patient health record data. In the introduction to the validity package, we stated that different criteria should be involved in assessing the validity of each type of measures, as follows:

Event reporting measures: How completely the relevant events are being reported.

Administrative measures: How accurately the counts or rates estimated with these measures match those that would be obtained using data from patient health records.

Chart-Based Measures: How consistently data is abstracted out of the health records, as measured by inter-rater reliability.

Participants were given the following instructions regarding how they should consider the concept of "validity" for a patient safety outcome measure and the approach they should take for rating the validity of each measure:

When assessing the validity of each measure being considered, we ask you to address the following standard:

*** *For the type of event the measure is intended to represent, how well does the measure accurately capture the frequency of actual events that are occurring during clinical care processes?*

To draw a conclusion on measure validity, the following logical steps should be taken:

1. *Determine the extent to which evidence regarding the validity of a measure is available in the published scientific literature.*
2. *If such published evidence is available, assess what the evidence says about how accurately the measure captures the underlying actual events.*
3. *In the absence of such published evidence (or limited evidence availability), determine the extent to which you feel it is appropriate to use clinical judgment and a consensus rating process to judge the validity of the measure.*

You will be facing the following two major issues in determining measure validity:

1. *Many of the measures in the established measure sets have not been validated regarding their ability to accurately capture the frequency of actual events, as has been specified in the validity standard we are asking you to address. As you will see from the validity materials RAND has compiled, for some measures we could not find published validation data, and for others, the validation is of a limited nature.*
2. *For the measures identified as candidates for ambulatory care and long-term care settings, only a few measure sets currently exist for long-term care (e.g., NQF Endorsed Nursing Home Measures) or for ambulatory care. All the inpatient-based measures identified by the Delphi participants as potentially applying to one of these settings must be considered concept measures in these settings, because the original development and testing work for these measures was done using inpatient data.*

With this context, we suggest that you be conservative in rating the validity of the short list of important measures. We suggest the following approach:

+++ *For the inpatient surgical and non-surgical measures, use the full rating scale, reserving higher ratings for measures you feel are well validated and found to be valid, and using the lower ratings for measures that you feel have either been found to not be valid or require more validation work.*

++ *For the ambulatory care and long-term care measures, in recognition that many of them are neither fully developed nor validated, we have added to the rating scale a response option "not ready for rating", for your use for measures you feel are not developed sufficiently to even merit rating.*

All participants were asked to complete their validity ratings using a web-based survey instrument, but they were also given the option of completing a hardcopy version of the instrument. Text boxes were provided in the survey form, which participants could use to record comments on the measures and relevant measurement issues, and to share their opinions on the validity rating process.

Respondents were given approximately 2 weeks to complete their validity ratings for Round 3. A total of 30 participants rated validity for some or all of the 106 measures that the Delphi participants had identified as important measures in Round 2.

ROUND 4: CONCLUDING CONFERENCE CALL

The Round 4 conference call was held about 4 weeks after the end of Round 3. Shortly before completing Round 3, we invited the Delphi participants who had performed ratings in Round 2 to participate in the conference call (the participants in Round 3 were a subset of the Round 2 participants). The stated purpose of the call was to give participants an opportunity to discuss the results of the Delphi process, to identify key issues regarding use of the measures, and to offer recommendations to AHRQ regarding next steps for use of the measures or for additional research needed to strengthen them.

Three weeks after completing Round 3, we sent an e-mail to all of the participants who had participated in Delphi Round 2. We thanked them for their hard work, provided the results of the validity rating step with supportive documentation, and presented an agenda for discussion during the call.

Several materials were provided as attachments to the conference call e-mail, which presented specific results from the validity ratings as well as aggregated results from the Delphi process as a whole. These included a summary of the results of the validity rating step, detailed tables showing validity scores for each of the candidate measures, combined importance and validity results with median ratings and 20th percentile ratings for each, a summary of participant comments regarding the validity rating step, the final set of three groupings for the important measures designed to guide future development and validation work, and a summary of overall results of the entire Delphi process.

The following agenda was established for the conference call:

1. Introductions of participants in the call
2. Review of results from the importance ratings and validity ratings
3. General feedback from participants on the rating process and results
4. Discussion of the draft summary of results from the Delphi process

5. Other specific discussion items
    a. The groupings of important inpatient measures based on priority for future work
    b. Key measurement issues raised by participants for the inpatient measures
    c. Approach to newly identified measures for ambulatory care and long-term care
    d. Choices of specific inpatient items that are similar to other less highly rated measures

Sixteen Delphi participants were able to participate in the conference call. RAND staff facilitated the discussion, working with the following list of discussion questions that was distributed to the Delphi participants as an attachment to the e-mail:

General Topics for Discussion:

1. *What are your general reactions to the outcome measures that were identified as the most important in the Delphi rating process? What feedback do you have on the list of lessons from this process that RAND identified?*
2. *The validity rating step in the process was undertaken to develop some additional perspective that could guide priority setting for future development work on the important measures, even while recognizing the limited amount of validation evidence available. We know this was a difficult step for you, and we expected to see low validity ratings. This step did help refine priorities, and it also highlighted how far we have to go before measures are well validated. What are your general reactions to your experience in performing the validity ratings and the results of that process?*
3. *What general message should we communicate to AHRQ regarding the current knowledge of validity for existing patient safety outcome measures and implications for future research priorities?*

Discussion of Next Steps for Measures Rated as Important

4. *Based on the results of your ratings in the Delphi process, we have identified two groups of measures that merit the first attention for additional research to validate them and refine measurement specifications and procedures. What are your reactions to our suggested groups of measures? How might these groupings be revised?*
5. *Several specific themes regarding measurement issues emerged during the Delphi process, on which we would like to hear your discussion as a group:*
    - *The lack of denominators for many of the measures*
    - *Variations in use of measures by different users*
    - *Risk adjustment issues*
    - *Low rates for measures addressing low-frequency events*
    - *Coding ambiguities for measures based on administrative data*
6. *We plan to characterize the ambulatory care and long-term care measures as starting points for future measure development for these health care settings. This approach is reinforced by your low validity ratings for them, as well as the substantial percentages of people who said it was premature to rate them. What guidance do you have regarding what to say about these measures in the report?*
7. *To inform our report on the Delphi results, we would like additional information on choices made among measures during the importance rating step. For the measures listed in Table 1 (below), why were some measures rated more highly than alternative similar ones?*

Delphi participants were invited to respond to these questions in an open-ended way, and to engage each other in reviewing the results of the process. RAND staff took notes during the conference call, and a summary set of notes describing the participants' responses to each of the above questions was distributed to the participants about a week following the conference call.

SECTION 3. RESULTS

In this section are presented the results of the two key steps of the Delphi process in which the participants rated the importance of the candidate outcome measures the validity of the subset of those measures that they identified as important. We start by presenting the numbers and characteristics of the individuals who participated in each Delphi step. Then we present the results of the importance ratings, followed by the results of the validity ratings. In discussion of the importance ratings results, we describe the method RAND used to identify the subset of measures that were rated most highly by the Delphi participants, and we present tables listing those measures for each health care setting addressed in this work.

Using the combined results of the importance and validity ratings, three groups of important inpatient measures were identified that are at differing states of readiness for use in national-level monitoring and epidemiological tracking. These groups, along with the method used to identify them, are presented in Section 4 of this report.

The priority measures identified by the Delphi participants also included 12 measures for ambulatory care (most of them addressing outpatient surgery) and 23 measures for long term care. All but a few of the measures for these two settings represent outcomes in concept only. That is, development of technical specifications and empirical testing of the measures remain to be done before the measures would be ready for use, and the data systems needed to apply measures may not exist.

PARTICIPATION IN THE DELPHI PROCESS

At the start of the Delphi process, we contacted a total of 92 individuals with invitations to participate in the process. As shown in Table 1, participation rates were high in Steps 1 and 2 of the process, but decreased somewhat in Step 3 (validity rating). The lowest participation occurred in Round 4, in part due to scheduling conflicts for this group of very busy individuals. Those interested in participating in the call provided us with their time availability, and we selected the date to get the maximum participation possible. However, only 16 out of the 22 individuals who provided us with their schedules for the call were available at the selected time.

Table 1
Number and Backgrounds of Participants in Each Step of the Modified Delphi Process

		Participant Backgrounds			
Delphi Round	Number Contacted	Number	Percentage of MDs or RNs	Percentage of MDs and RNs with Academic Degrees	Percentage with Academic Degree Only
1. Candidate Measures	92	67	26.9	44.8	28.3
2. Importance Rating	67	47	29.8	42.5	27.7
3. Validity Rating	47	30	40.0	36.7	23.3
4. Conference Call	47	16	37.5	37.5	25.0

The distribution of clinicians and PhDs did not change considerably across the Delphi steps. Between Round 2 and Round 3, there was a relative loss of individuals with research but no clinical degrees, resulting in a greater percentage of individuals with clinical degrees only

participating in validity rating steps. (See Appendix A for the list of participants in the importance rating step.)

RESULTS OF THE IMPORTANCE RATINGS

In the first round of rating, 47 participants rated some or all of the 501 candidate patient safety measures. Presented in Table 2 are the numbers of measures in the five major categories, and within each category, the modal number of raters and the range of the minimum to maximum number of raters for each item in the category.

Table 2
Numbers of Measures and Ratings for the Importance Rating Step

Patient Safety Measure Categories	Number of Measures	Modal Number of Raters (Range)
Surgical Inpatient	106	34 (31-37)
Surgical and Non-surgical Inpatient	24	35 (34-36)
Non-surgical Inpatient	135	36 (33-37)
Ambulatory	94	22 (20-22)
Long Term Care	142	13 (11-14)
Total	501	47

Highlights of Rating Results

- Overall, the surgical inpatient measures were rated highest for importance by the Delphi participants. Ratings for these measures ranged from 3 to 9, peaking at 7 and 8, and 33 percent of the measures were identified as either highly or moderately important. The percentages identified as important from the other two groups were 21 percent for the inpatient measures relevant to both surgical and non-surgical care and 23 percent for the non-surgical inpatient measures.
- The most highly rated measures were four surgical inpatient measures, all of which had median ratings of 9 and 20th percentile ratings of 8. All of these measures addressed procedural errors during surgery, e.g., wrong patient, wrong body part, wrong procedure.
- Several of the surgery measures rated as important included alternative measures for the same concepts, which will require further deliberation to determine which measures are the most appropriate to represent each concept.
- For non-surgical inpatient measures, the top measures fell into three categories: medication error with serious consequences (death, near-death, disability, or permanent harm), criminal activity or security breach with serious consequences (sexual assault, rape, impersonation of care giver, assault, wrongful infant discharge), and severe pressure ulcers acquired within a healthcare facility.
- What did not get rated as important also is of interest. For surgical inpatient measures, these include a number of risk-adjusted mortality measures for specific procedures, numerous specific types of nosocomial infections, a variety of post-surgical adverse events and negative outcomes, and accidental cuts or punctures.

- The non-surgical inpatient measures that were not rated as important include many drug-specific adverse events and poisonings, physical security problems (e.g., falls, bed entrapment), medical conditions and mortality where faulty medical monitoring may be implicated (e.g., patient harm from hypoglycemia, anticoagulation, over-hydration, drug psychoses), and a variety of events that relate to general vigilance (e.g., protecting against suicides, burns, timely follow-up of test results).
- The more inclusive measures appeared to be more likely to be rated as important than any similar measures that applied specifically to one aspect of the relevant issue. For example, medication errors with serious patient consequences were rated highly but numerous specific types of medication errors were not.
- All but one of the candidate ambulatory measures rated as most important were related to ambulatory surgery, with similar priorities identified for the surgical inpatient ratings, or to adverse events in medication management. Only one measure – failure to follow up test results – was related to general outpatient care. None of the rehospitalization measures were rated highly, although three of these measures had median ratings of 7. No candidate measures addressing the normal office visit aspects of outpatient care were available for consideration in this rating process.
- The long-term care measures rated as important include medication errors, security events (e.g., rape and sexual assault), vigilance problems (pressure ulcers and burns during bathing), and administration of ABO incompatible blood (which was not rated as important for inpatient surgical measures). These results are quite similar to those for the non-surgical inpatient measures identified as important.
- The majority of the measures in all categories that were rated as important were those that are most clearly attributable to an error, such as wrong-site surgery or blood stream infections associated with catheters. Other measures that address complications or other negative outcome that suggest iatrogenic involvement, such as 30-day mortality after CHF or MI, tend not to be rated as important.

Ratings Results and Identification of Important Measures

We show here the distributions of the median[4] ratings for all the measures in each of the five sets. The first three sets for inpatient measures are shown in Figure 1, and the final two sets, *Ambulatory Care* and *Long Term Care*, are shown in Figure 2. Overall, the *Surgical Inpatient* measures received the highest ratings. The *Surgical and Nonsurgical Inpatient* measures also were given fairly high ratings but the distribution is more compact. The median ratings for the other three sets of measures are more widely distributed, particularly for *Ambulatory Care*.

Our goal in the importance rating step was to identify the most important measures. We selected measures with a median rating of 8 or higher (on the 1 to 9 scale) as the most important

[4] The median is the middle data value when the data are listed in order from smallest to largest (if there are an even number of values and the two middle values differ, the median is defined to be the mean of these two values). Thus half of the participants rated a measure at the median or higher.

measures.[5] The median provides some information about the level of agreement in the ratings, since a median of 8 or higher indicates that at least half of the raters scored the measure with an 8 or a 9, the top two levels of importance. These measures moved to the next round of the Delphi process where their validity was rated.

We then identified a subset of these measures for which participants had particularly high levels of agreement. The subset of measures we label as 'High' are those which, in addition to having median ratings of 8 or higher, also have 20th percentile ratings of 7 or higher. This indicates that 80 percent or more of the raters gave the measure a rating of 7 or higher. The rest of the important measures are designated as 'Moderate' and these measures, while having a median rating of 8 or higher, have 20th percentile ratings lower than 7, indicating that a full fifth of the raters judged the measures as lower than a 7 on the importance scale.

Table 3 shows the number and percentage of the candidate measures rated as important for each of the measure categories. Tables 4 through 8 list the 'High' and 'Moderate' important measures for each health care setting. In these tables, the "High" important measures are highlighted in light grey, and measures for which there was some disagreement among raters are highlighted in a darker grey in the 20th percentile column (see discussion on disagreement below).

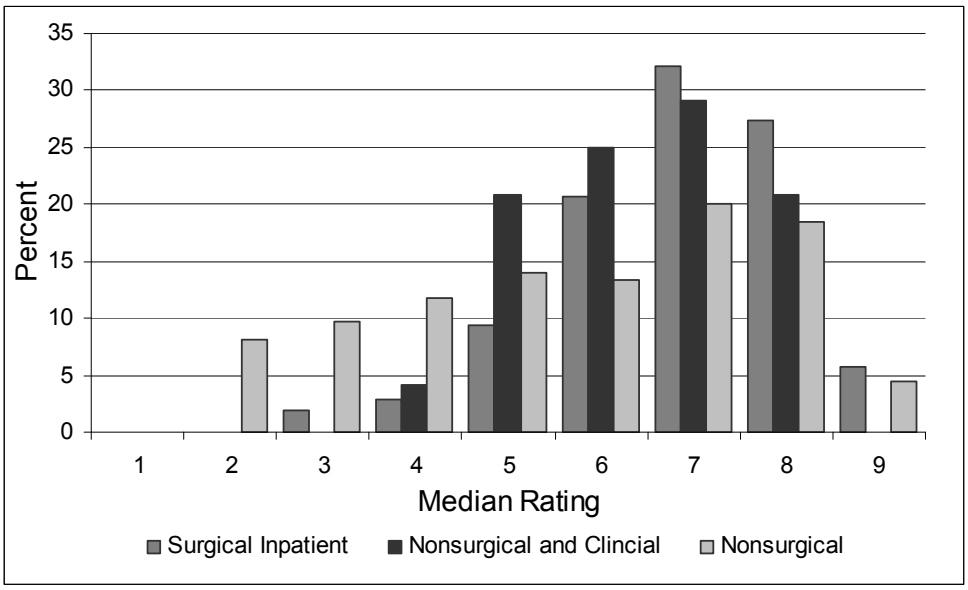

Figure 1 Distributions of Median Importance Ratings for Inpatient Care Measures

[5] Note that the measures that achieved a median rating of less than 8 we label as being of comparatively 'Low' importance. The 'Low' measures are not included in the next round of the Delphi process (examining the validity of highly rated candidate measures).

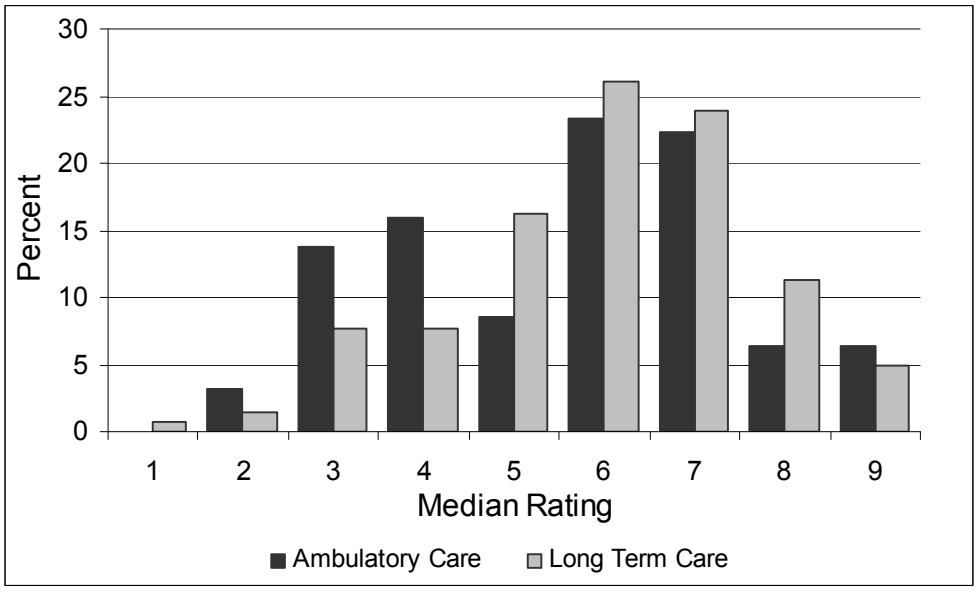

Figure 2 Distributions of Median Importance Ratings for Ambulatory Care and Long Term Care Measures

Table 3
Number and Percentage of Candidate Measures in Each Category That Were Rated As Important by the Delphi Participants

Patient Safety Measure Categories	Number of Measures	Number and Percentage of Candidate Measures Scored as Important		
		High	Moderate	Combined
Surgical Inpatient	106	10	25	35 (33%)
Surgical and Non-surgical Inpatient	24	-	5	5 (21%)
Nonsurgical Inpatient	135	10	21	31 (23%)
Ambulatory	94	10	2	12 (13%)
Long Term Care	142	11	12	23 (16%)
Total	501	41	65	106 (21%)

Table 4
Surgical Inpatient Measures Identified as High or Moderate in Importance Based on Ratings by the Delphi Participants

	Median	20th Percentile	Rating
Cardiac			
NS 20 - Myocardial infarction intra/post-surgery	8	6	Moderate
NS 19 - Cardiac arrest requiring CPR post surgery	8	6	Moderate
STS 6 - Risk-adjusted operative mortality for CABG	8	7	High
STS 7 - Risk-adjusted operative mortality for aortic valve replacement (AVR)	8	6	Moderate
STS 8 - Risk-adjusted operative mortality for mitral valve replacement/repair (MVR)	8	6	Moderate
STS 9 - Risk-adjusted operative mortality for MVR and CABG surgery	8	6	Moderate
STS 10 - Risk-adjusted operative mortality for AVR and CABG surgery	8	6	Moderate
Pulmonary			
SC 5 - Patients diagnosed with post-operative ventilator-associated pneumonia (VAP) during index hospitalization	8	6	Moderate
PE/DVT			
SC - 3 Intra/post-operative pulmonary embolism diagnosed during index hospitalization and within 30 days of surgery	8	5	Moderate
PSI 12 - Postoperative pulmonary embolism or deep vein thrombosis (DVT)	8	6	Moderate
MP 1 - Postoperative venous thromboembolic events	8	5.5	Moderate
Labor & Delivery			
NQF 14 - Maternal harm in labor and delivery of low-risk pregnancy	8	6	Moderate
PSI 17 - Birth trauma - Injury to neonate	8	6	Moderate
PSI 18 - Obstetric trauma - vaginal delivery with instrumentation	8	5	Moderate
PSI 20 - Obstetric trauma - cesarean delivery	8	6	Moderate
Infections			
STS 2 - Deep sternal wound infection rate (for CABG surgical patients)	8	6.5	Moderate
MP 4 - Insertion-site infections associated with central vascular catheters (CVC's)	8	5	Moderate
PSI 13 - Postoperative sepsis	8	6	Moderate
NS 25 - Systemic Sepsis (SIRS) post-surgery	8	6	Moderate
NS 26 - Systemic Sepsis (Sepsis) post-surgery	8	7	High
NS 27 - Systemic Sepsis (Septic Shock) post-surgery	8	7	High
MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	8	6	Moderate
NSPM 7 - Central line-associated bloodstream infection (CLABSI) rate for intensive care unit locations	8	6	Moderate

NSPM 8 - Central line-associated bloodstream infection (CLABSI) rate for birthweight categories in neonatal intensive care unit locations	8	6	Moderate
NS 3 - Organ/Space surgical site infection (SSI)	8	6	Moderate
Procedural			
PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	8	7	High
NQF 4 - Foreign object retention after surgery	8	7	High
NY 3 - Unintentionally retained foreign body	8	7	High
NY 1 - Wrong patient, wrong site surgical procedure	9	7	High
NQF 1 - Surgery performed on the wrong body part	9	8	High
NQF 2 - Surgery performed on the wrong patient	9	8	High
NQF 3 - Wrong surgical procedure performed on a patient	9	8	High
NY 2 - Incorrect procedure or treatment - invasive	9	6	Moderate
Other			
NQF 5 - Intraoperative or immediately postoperative death in an ASA Class 1 patient	8	6	Moderate
NQF 13 - Administration of ABO-incompatible blood	9	6	Moderate

Table 5
Surgical and Non-Surgical Inpatient Measures Identified as High or Moderate in Importance Based on Ratings by the Delphi Participants

	Median	20th Percentile	Rating
NSPM 9 - Ventilator-associated pneumonia (VAP) rate for intensive care unit (ICU) locations	8	6	Moderate
NSPM 10 - Ventilator-associated pneumonia (VAP) rate for birthweight categories in the neonatal intensive care unit (NICU) locations	8	6	Moderate
NY 12 - New acute pulmonary embolism	8	5.5	Moderate
NQF 8 - Air embolism	8	5	Moderate
NQF 16 - Failure to treat hyperbilirubinimia	8	5	Moderate

Table 6
Non-Surgical Inpatient Measures Identified as High or Moderate in Importance Based on Ratings by the Delphi Participants

	Median	20th Percentile	Rating
General Clinical Care			
MP 8 - Hospital-acquired bloodstream infections (BSIs)	8	6	Moderate
NY 20 - Misadministration of radiation or radioactive material	8	6	Moderate
NQF 7 - Patient harm from improperly used device	8	4	Moderate
NY 8 - Malfunction of equipment with harm	8	6	Moderate
NQF 17 - Severe pressure ulcers acquired after admission to a health care facility	8	7	High
NY 4 - Unexpected deaths	8	6	Moderate
PSI 2 - Death in low mortality DRGs	8	6	Moderate
PSI 4 - Failure to rescue	8	5	Moderate
NQF 19 - Patient harm from electric shock	8	6	Moderate
NQF 20 - Delivery of contaminated gas to a patient	8	5	Moderate
NQF 21 - Patient harm from burns	8	6	Moderate
NY 15 - Patient burns	8	6	Moderate
CP 5 - Burns/scalding during bathing	8	5	Moderate
NY 6 - Loss of limb or organ	8	2	Moderate
NQF 11 - Patient suicide, or attempted suicide	8	5	Moderate
CP 7 - Hours of overwhelming pain in terminal patients	8	6	Moderate
Medication/Drug			
CP 15 - Administration of enteral medications/solutions intravenously	8	6	Moderate
NQF 12 - Death or serious disability due to medical error	9	7	High
NY 9 - Medication error occurred that resulted in permanent patient harm	9	8	High
NY 10 - Medication error occurred that resulted in near-death event	8	7	High
NY 11 - Medication error occurred that resulted in patient death	9	7	High
Physical Security			
NQF 23 - Patient harm from use of restraints	8	5	Moderate
NQF 9 - Infant discharged to the wrong person	9	7	High
NY 31 - Infant discharged to the wrong family	9	6.5	Moderate
NY 30 - Infant abduction	9	6.5	Moderate
NQF 25 - Abduction of a patient of any age	8.5	6	Moderate
NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	8	7	High
NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	8.5	7	High
NY 32 - Rape of a patient	8.5	7	High
NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	8	7	High
NY 21 - Crime resulting in death or serious injury	8	5	Moderate

Table 7
Ambulatory Care Measures Identified as High or Moderate in Importance Based on Ratings by the Delphi Participants

	Median	20th Percentile	Rating
AMB NS 3 - Organ/space surgical site infection (SSI)	8	5	Moderate
AMB PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	8	7	High
AMB NQF 4 - Foreign object retention after surgery	8	7	High
AMB NY 1 - Wrong patient, wrong site surgical procedure	9	7	High
AMB NQF 1 - Surgery performed on the wrong body part	9	8	High
AMB NQF 2 - Surgery performed on the wrong patient	9	7	High
AMB NQF 3 - Wrong surgical procedure performed on a patient	9	8	High
AMB NQF 12 - Death or serious disability due to medical error	9	8	High
AMB NY 9 - Medication error occurred that resulted in permanent patient harm	8	8	High
AMB NY 10 - Medication error occurred that resulted in near-death event	8	8	High
AMB NY 11 - Medication error occurred that resulted in patient death	9	8	High
AMB CP 14 - Failure to follow-up test results	8	5	Moderate

Table 8
Long-Term Care Measures Identified as High or Moderate in Importance Based on Ratings by the Delphi Participants

Important Long Term Care Measures	Median	20th Percentile	Rating
LTC MP 4 - Insertion-site infections associated with central vascular catheters (CVC's)	8	3	Moderate
LTC MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	8	5	Moderate
LTC NQF 7 - Patient harm from improperly used device	8	6	Moderate
LTC NY 8 - Malfunction of equipment with harm	8	6	Moderate
LTC NQF 13 - Administration of ABO-incompatible blood	8	7	High
LTC UT 23 - Decubitus ulcer	8	5	Moderate
LTC NQF 17 - Severe pressure ulcers acquired after admission to a health care facility	9	7	High
LTC NH 10 - Average risk residents with pressure ulcers	8	5	Moderate
LTC NQF 21 - Patient harm from burns	8	6	Moderate
LTC CP 5 - Burns/scalding during bathing	8	7	High
LTC NY 7 - Impairment of limb, organ, or body function	8	5	Moderate
LTC NQF 11 - Patient suicide, or attempted suicide	8	5	Moderate
LTC CP 7 - Hours of overwhelming pain in terminal patients	8	6	Moderate
LTC NQF 12 - Death or serious disability due to medical error	9	8	High
LTC NY 9 - Medication error occurred that resulted in permanent patient harm	9	8	High
LTC NY 10 - Medication error occurred that resulted in near-death event	8	8	High
LTC NY 11 - Medication error occurred that resulted in patient death	9	8	High
LTC HHC 11 - Emergent care for improper medication administration, medication side effects	8	5	Moderate
LTC NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	8	7	High
LTC NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	8	7	High
LTC NY 32 - Rape of a patient	9	7	High
LTC NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	9	8	High
LTC NY 21 - Crime resulting in death or serious injury	9	6	Moderate

Measures Rated as Important But With Low Agreement

We show here the distributions of the individual ratings for several measures that were highly rated on importance, but for which there was some disagreement among raters regarding their importance. These measures were included in the validity rating round, and participants were advised that they should be aware there is considerable disagreement about their importance.

These measures fall into two categories. The first category consists of five measures with a median rating of 9 (at least half the raters judged the measure at the highest level of importance), but with a 20th percentile less than 7 (suggesting some disagreement among raters). The distributions of importance ratings for these five measures are shown in Figures 3 and 4. The second set of measures consists of three measures for which the 20th percentile is less than 5, indicating that 20 percent of the raters judged the measure to have low importance. The distributions of ratings for these measures are in Figure 5. These eight measures are also highlighted with dark grey shading in the list of important measures in Tables 4 through 8.

Table 9
Measures Rated As Important But With Some or Strong Disagreement Among Delphi Participants

ID Number	Measure Name
Some Disagreement	
NQF 13	Administration of ABO-incompatible blood
LTC NY 21	Crime resulting in death or serious disability
NY 2	Incorrect procedure or treatment - Invasive
NY 30	Infant abduction
NY 31	Infant discharged to the wrong family
Disagreement	
NQF 7	Patient harm from improperly used device
NY 6	Loss of limb or organ
LTC MP 4	Insertion-site infections associated with central vascular catheters (CVCs)

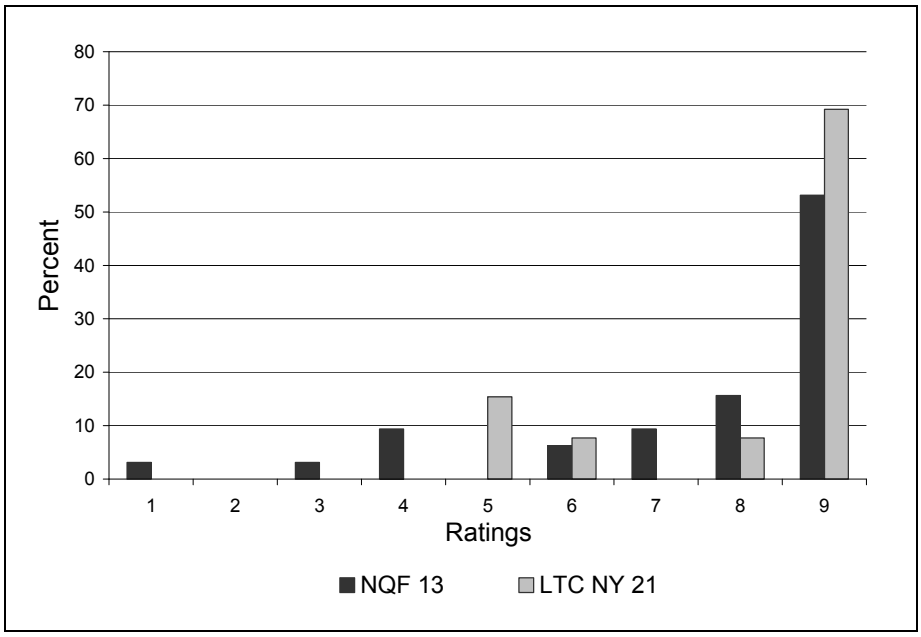

Figure 3 Distributions of Ratings for Highly Rated Measures With Some Disagreement Among Raters (#1)

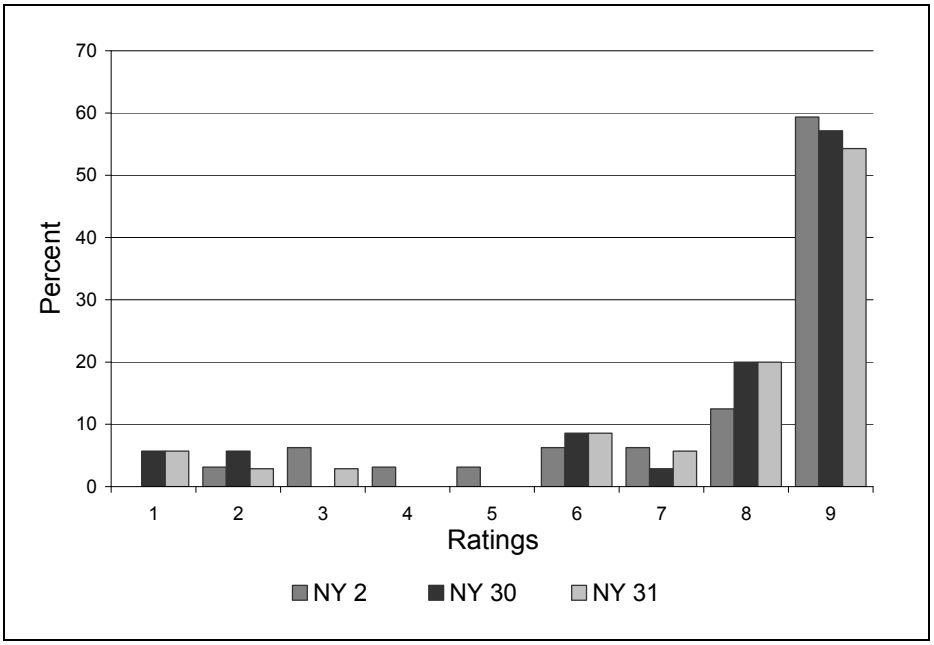

Figure 4 Distributions of Ratings for Highly Rated Measures With Some Disagreement Among Raters (#2)

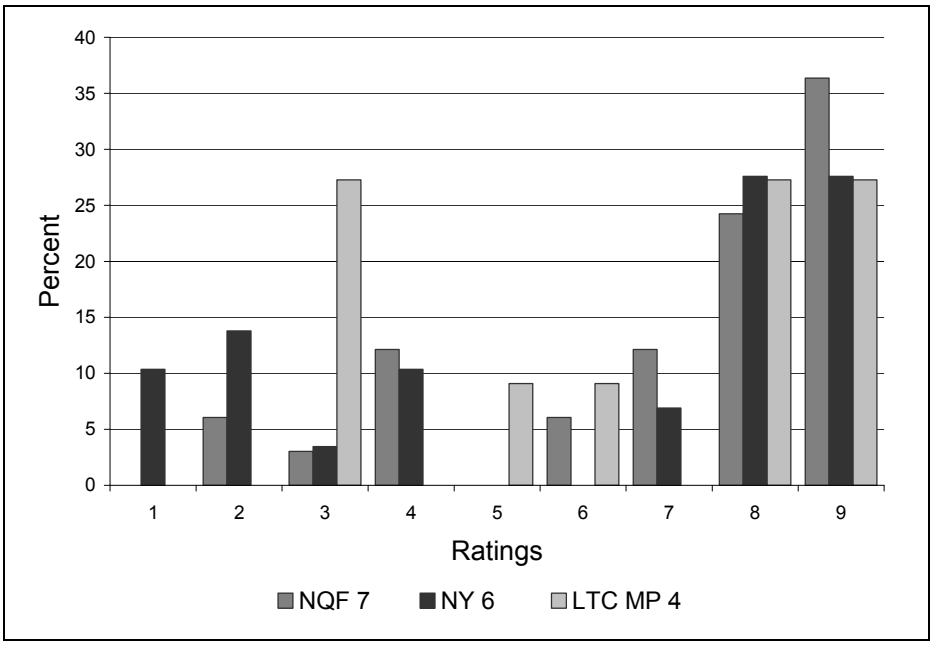

Figure 5 Distributions of Ratings for Highly Rated Measures With Disagreement Among Raters

Full Documentation of Importance Rating Results

Presented in Appendix D are summary statistics for the ratings for each candidate measure that was included in the Delphi importance rating task. The full documentation includes the number of raters, median rating, 20th percentile rating, mean absolute deviation[6], mean, and standard deviation of the ratings for each candidate measure. This documentation is ordered by major categories of measures (e.g., surgical inpatient, ambulatory, etc.), and then by functional area subgroups within the 5 major categories (e.g., cardiac, pulmonary, etc.). The measures identified as important are highlighted in these tables. We also have prepared histograms of the distributions of ratings for each of the important measures, which are available by request. Appendix E contains the participants' comments from the importance rating process.

RESULTS FROM THE VALIDITY RATINGS

In the validity round of rating, 30 participants rated some or all of the 106 candidate patient safety measures. Presented below are the number of measures in each of the five major categories and number of raters (modal number of raters and the range of the minimum to maximum number of raters for each item).

Highlights of Rating Results

The validity rating results reflect the generally lean situation regarding lack of validation information for the vast majority of the measures. The validity ratings for the inpatient measures were generally low, with several participants stating that this step was difficult to perform.

[6] The mean absolute deviation is the mean of the absolute values of the differences between each rating and the mean of the ratings for a measure.

Table 10
Numbers of Measures and Ratings for the Validity Rating Step

Patient Safety Measure Categories	Number of Measures	Modal Number of Raters (Range)
Surgical Inpatient	35	28 (27-30)
Surgical and Nonsurgical Inpatient	5	27 (25-29)
Nonsurgical Inpatient	31	27 (25-28)
Ambulatory	12	25 (23-25)
Long Term Care	23	23 (22-24)
Total	106	27 (22-30)

Substantial percentages of participants said it was too early to rate all but a few of the ambulatory care and long-term care measures. This reflected the nature of most of these measures as concept measures, which were drawn from measures developed for the inpatient setting. Despite these challenges, the validity ratings provide some useful information that can help guide priorities for future development and validation work for the measures identified as important in the Delphi process.

- For the important inpatient measures identified, the highest median validity rating was an 8, and that was given to only one measure. One other measure had a median validity rating of 7.5, and 11 other measures received median ratings of 7.
- Overall, the surgical inpatient measures received the highest median validity ratings. The validity ratings for the non-surgical inpatient measures are widely distributed, whereas the ratings for the surgical and non-surgical measures are clustered at median ratings of 5 and 6. Except for hospital acquired bloodstream infections, all of the non-surgical inpatient measures with higher validity ratings are security events.
- The important ambulatory care and long term care measures had lower median validity ratings than the three sets of inpatient measures. None of these measures had median validity ratings above 6. The percentage of raters who felt these measures were not ready to be rated varied from 8 percent to 58 percent, depending on the particular measure.

Summary of Validity Ratings for Each Setting

We show here the distributions of the median[7] ratings for all the measures in each of the five sets. The first 3 sets for inpatient measures are shown in Figure 1, and the final two sets for Ambulatory Care and Long Term Care are shown in Figure 2. These figures highlight the higher ratings for the three sets of inpatient measures as well as differences in the distributions of validity ratings across the care settings.

[7] The median is the middle data value when the data are listed in order from smallest to largest (if there are an even number of values and the two middle values differ, the median is defined to be the mean of these two values). Thus half of the participants rated each measure at the median or higher.

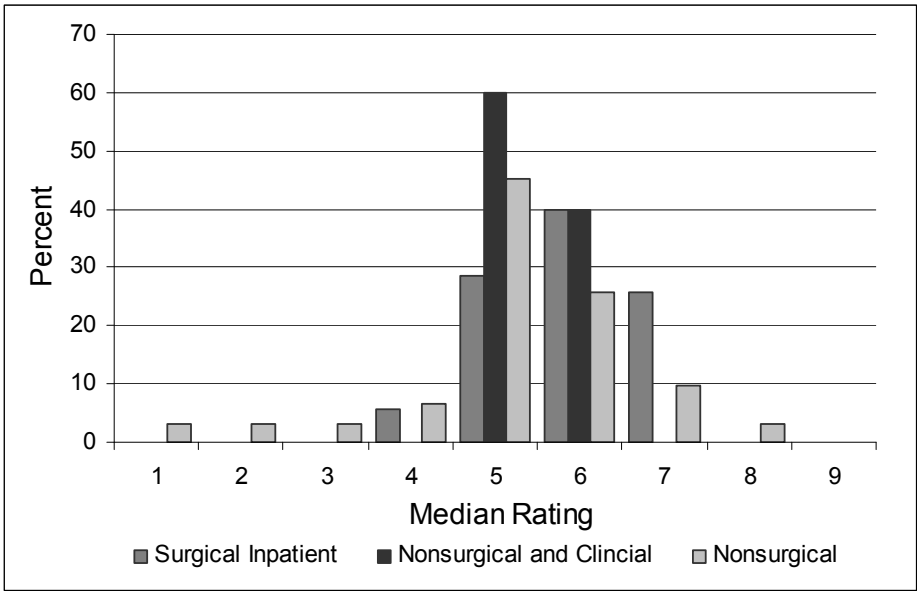

Figure 6 Distributions of Median Validity Ratings for Inpatient Measures

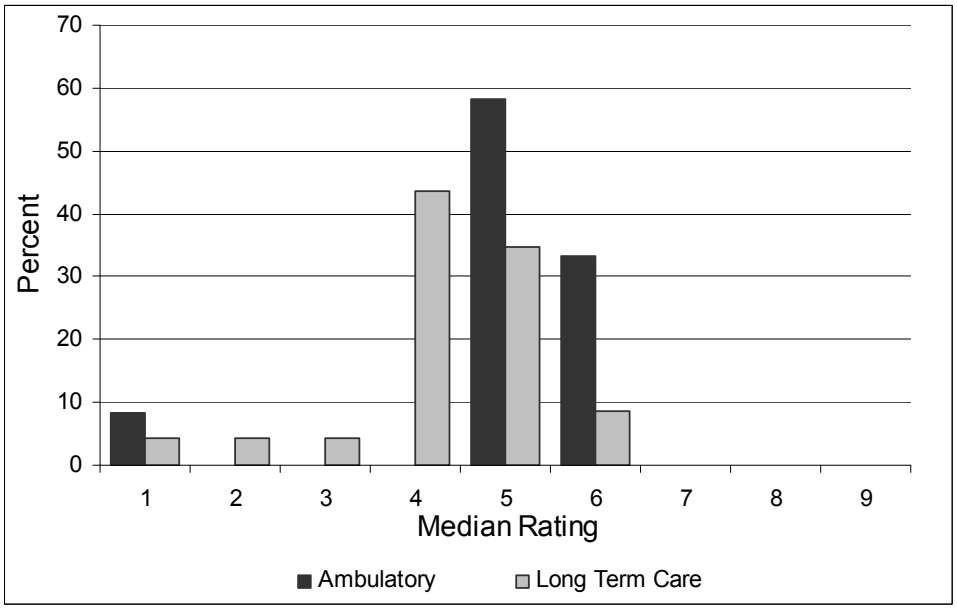

Figure 7 Distributions of Median Validity Ratings for Ambulatory Care and Long Term Care Measures

Full Documentation of Validity Rating Results

Presented in Appendix F are summary statistics for the validity ratings for each candidate measure that was included in the Delphi validity rating task. The full documentation includes the

number of raters, median rating, 20th percentile rating, mean absolute deviation[8], mean, and standard deviation of the ratings for each measure identified as important by the raters, and where applicable, the percentage of raters who judged measures not ready to be rated. The measures in this list are sorted by median validity ratings, and measures with median validity ratings of 7 or higher have been highlighted. Appendix E contains the participants' comments from the validity rating process.

Presented in Appendix G is a listing of the important measures that shows both their importance ratings and validity ratings (medians and 20th percentiles), with the measures sorted by clinical topic area as organized in the rating instrument. Both highly important ratings and validity ratings of 7 or higher are highlighted.

[8] The mean absolute deviation is the mean of the absolute values of the differences between each rating and the mean of the ratings for a measure.

SECTION 4. OVERALL FINDINGS AND ISSUES

The ratings of importance and validity provide the basis for identifying groups of measures that are close to ready for use in national monitoring, as well as other important measures that require further validation before they can be used with confidence. These grouping were developed only for the inpatient outcome measures because most of them were reasonably well developed measures. By contrast, few of the measures identified for ambulatory care and long-term care are well enough validated yet to be ready for use in national trend monitoring. Using the two sets of ratings, RAND identified the following three groups of inpatient measures:

Group 1 – Important measures that are close to being ready to use

Group 2 – Highly important measures that are not well validated

Group 3 – Other important measures that are not well validated

Measures that received high importance ratings (as described in Section 2) and also received high validity ratings (median validity ratings of at least 7) were designated as close to being ready for use (Group 1). These measures are listed in Table 11 at the end of this section. The remaining measures were divided into the other two sets based on their importance ratings – highly important measures that are not well validated and other important measures that are not well validated (in both cases having median validity ratings of less than 7). These two sets of measures are listed in Tables 12 and 13.

COMMENTS AND ISSUES RAISED BY DELPHI PARTICIPANTS

The conference call with the Delphi participants began with a review of the steps taken in the Delphi Process, to provide context for discussion. The strong difference in the results for the importance and validity ratings was highlighted, with the validity ratings being much lower than the importance ratings, and the wide variation in how the participants judged validity.

As participants discussed the logic they used for rating importance, it was learned that they used a variety of approaches. Some focused more on an event's severity of harm, and others focused more on frequency of events. One participant noted that he viewed rare events as less important, so ranked them low. Several participants also stated they had trouble separating importance from validity and other measurement issues. Despite these differences, the rating process yielded strong agreement among participants for a clearly identifiable set of measures.

For the validity ratings, participants agreed that this was a very difficult task because it was hard to separate reliability and other measurement issues from validity. Some noted that the limited information available on validity made the validity ratings nearly impossible. Others reported that they used their professional judgment to rate validity when empirical evidence was not available. It was suggested by one participant that further validation work should address whether the event actually occurred (e.g., did a patient really have DVT), as well as whether the event was the result of an error or potentially poor quality.

Comments on the Measures Identified As Important

There was general agreement that reported events should be tracked, but some participants were not comfortable rating their importance because they did not view them as measures. There was some discussion regarding how best to use reported events for surveillance, with no clear consensus on this issue.

The issue of reported events led to considerable debate about what constitutes a measure, with some participants stating that counts of reported events are not measures. Implicit in this perspective is the difference between counts and rates, in which rates have well-defined denominators. It was stated that many of the measures represent events that have not been tested as measures (i.e., not rates). In addition, for many reported events, the numerators are not well-defined either. Many reporting systems do not provide definitions of the events to be reported, and definitions that do exist vary across reporting systems. Therefore, there is not a common understanding of what is meant by most of the reported events, and reporting organizations may view the same event differently regarding whether to report it or not.

Several participants expressed their concern that there were many duplicative measures in the candidate measure set, and this duplication remained for many of the measures rated as important. Some also expressed surprise that some measures that were not well defined had made it to the final round.

There was general agreement that most of the measures rated as important represent a good starting point for further measure refinement. There also were strong concerns about how the measures would be used, specifically that their use would go beyond the goals of patient safety surveillance defined for this measure selection process. In particular, several participants felt there was a risk the measures would be misused to rank hospitals, for pay-for-performance systems, or regulatory functions. It was agreed that these concerns were well-founded.

There was substantial discussion, and some disagreement, among participants about whether and how to use rare events for national surveillance activities. There was general agreement that rare events should be included for surveillance, but less agreement on how to measure them. Some participants noted that national-level data could detect trends in low frequency events that could not be detected at the hospital or regional levels. There was some agreement that many of the rare events do not have well-defined denominators, and it would not be useful to try to find denominators for them, in particular, for events reported by hospitals to other organizations. It was noted that the Joint Commission requirement of reporting rare events is a constructive approach. It was also noted that rare events can be transformed into useful measures by bundling them or by tracking near misses. However, near misses were unlikely to be captured, except at the hospital level.

There was general agreement that another step is needed – to determine which of the identified measures are best used for which purposes. For example, reported events might be useful for timely corrective actions, whereas other measures might be better used for epidemiological assessment. Specifically, some participants noted that the NQF "never events" would not be useful for rating hospitals or for pay-for-performance, but they do have value for tracking patient safety. Others disagreed, saying that if the goal is to improve performance, "never events" are not useful.

One participant noted that the measures being considered represent a mix of events that includes those clearly due to errors, as well as adverse events that are the result of potentially poor quality. The distinction is important in terms of what kinds of corrective actions they imply. For obvious errors, the emphasis would be on determining the underlying or latent errors in the system. For adverse events due to poor quality, the first step would be a root cause analysis to determine if the event was due to an error or some other reason.

It was suggested that "less is more" – that it would be important to further reduce the number of measures, given the limited resources available for surveillance on a national level. In addition, different measures could be selected for different purposes. It was noted by one participant that the final selection of measures might be shaped by the setting of national priorities by NQF.

Several issues were raised by participants for a few specific measures:

- The sepsis measures were identified as problematic, because it is difficult to define and measure sepsis.
- Measurement of VAP, DVT, and pneumonia all present challenges, as these problems are difficult to define at the bedside and yet more difficult to record.
- Patient harm from restraints and hours of overwhelming pain are two measures that represent important issues but are difficult to operationalize.
- Patient falls in the hospital inpatient setting did not survive the importance ratings, although this measure was rated as important for long-term care.

Comments on Determination of Measure Validity

Concerns were expressed that there is no clear standard for validity, including determination of what is the threshold rating that indicates a measure is valid. Again, there was a sense that assessment of validity will depend to some extent on the purpose for which a measure will be used. One participant expressed dissatisfaction with the measures that have been identified because those that are important are often poorly rated on validity. Several participants felt that if imprecise measures were used for pay-for-performance, hospitals would be hurt badly.

One participant suggested that judgments about potential gains in safety would guide whether a measure was useful. It might be appropriate under certain conditions to use measures that were not well validated, where there clearly were problems with a safety issue and improvement would contribute to safer health care. In response to this point, others noted a related issue of "signal to noise" in measuring outcomes, in which the precision of measurement would be less of an issue when the current outcome performance was bad and large improvements were made.

Reactions to Priority Groups

The participants reviewed the three groups of measures developed by RAND based on the importance and validity ratings. There was general agreement that these three groups were as far as the participants could go in the Delphi process, and that further work to refine patient safety outcome measures should be done by working teams. It also was agreed that if the measures in groups 1 and 2 could be further developed into valid measures, it would be good start. It was noted specifically that the group 1 surgical measures are appropriate, and also that the group 2 measures would be good additions if their validity could be verified. This work should include eliminating duplicative measures to establish one measure that most effectively represents each safety concept. Some participants were concerned, however, that some highly important measures may have been lost due to inconsistency in ratings.

After refining the first two groups of measures, it was suggested that a good next step would be to look for additional measures in group 3 that address safety areas not represented by the measures in the first two groups. This step might also address the concern about losing important measures in the rating process. It was noted by one participant that it would take a lot of work to get Group 3 measures in shape. It also was suggested that it will be important to do additional work on NQF events, but that this task differs from what would need to be done to improve the other measures. It might also be useful to separate the reporting events from the measures that are rates, so that the two types of measures can be looked at as separate tools.

Comments on Ambulatory and Long-Term Care Measures

When asked about what further work is needed for the measures identified for ambulatory care and long-term care settings, participants had mixed responses. Some felt that the surgical care measures could transfer readily from the inpatient to outpatient setting. Others felt that more development and validation work was needed before applying inpatient measures to other settings. On the other hand, as noted by one participant, the measures for these settings are not much behind the inpatient measures, because the inpatient measures are not well validated either.

One participant noted that all of the measures considered are specific to one type of health care provider or setting, which ignores the passage of patients through many providers and settings. This person argued that we need a way to measure across providers because many errors are happening across settings or in transitions between them—hospital, nursing home, home care (ulcers, pain, etc).

FINAL OBSERVATIONS AND NEXT STEPS

Several participants in the Delphi process observed that this rating experience had been an eye-opener for them. It was difficult to get consensus, in part because of the diversity of perspectives and interests involved, and in part because of the complexity and intertwined nature of the importance of measures and relevant measurement issues.

An issue that permeated the discussion was the concern that once outcome measures are developed and made available for use, there is a serious risk that they will be applied in ways that are inconsistent with their intended use. Participants highlighted examples where publicly available measures now are being used to compare hospitals and judge their performance. In addition, pay-for-performance systems are picking up measures without adequate consideration of potentially perverse incentives that might be created. It was strongly urged that RAND point out this issue in the report. It will be important to express the concern of participants about these issues and to make clear statements of limitations of use for the measures.

The results of this rating process serve as a starting point for further work on refining measures, reviewing and seeking solutions to issues, and perhaps most importantly, further validation research to strengthen the evidence base for the validity of the measures identified as important. To help inform this future work, Appendices B and E present participants' comments on the original set of candidate measures and each of the importance and validity rating steps, and Appendix H lists detailed comments made by the participants for each of the candidate measures. In this process, it is advisable to further examine the Group 3 measures, as suggested by the Delphi participants, to identify any additional measures that address important safety issues that are not addressed by measures in the first two groups.

Table 11
Group 1 Measures – Important Inpatient Patient Safety Outcome Measures That Are Close to Ready for Use

Outcome Measure	Importance Ratings		Validity Ratings	
	Median	20th Percentile	Median	20th Percentile
Surgical Inpatient				
NQF 1 - Surgery performed on the wrong body part	9	8	7	5
NQF 2 - Surgery performed on the wrong patient	9	8	7	5
NQF 3 - Wrong surgical procedure performed on a patient	9	8	7	5
STS 6 - Risk-adjusted operative mortality for CABG	8	7	7	5
NQF 13 - Administration of ABO-incompatible blood	9	6	7.5	5
STS 2 - Deep sternal wound infection rate (for CABG surgical patients)	8	6.5	7	5
MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	8	6	7	5
NQF 5 - Intraoperative or immediately postoperative death in an ASA Class 1 patient	8	6	7	5
MP 1 - Postoperative venous thromboembolic events	8	5.5	7	5
Non-surgical inpatient				
NQF 9 - Infant discharged to the wrong person	9	7	8	4
NY 30 - Infant abduction	9	6.5	7	3
NQF 25 - Abduction of a patient of any age	8.5	6	7	4
MP 8 - Hospital-acquired bloodstream infections (BSIs)	8	6	7	5

Acronyms used for measure sources:

MP	Medicare Patient Safety Monitoring System
NQF	National Quality Forum
NS	National Surgical Quality Improvement Program
NY	New York Patient Occurrence Reporting and Tracking System
PSI	AHRQ Patient Safety Initiatives
UT	UT/MO Adverse Events
NSPM	Nursing Sensitive Performance Measures
STS	Society of Thoracic Surgeons Measures
NQFM	National Quality Forum Endorsed Mortality Measures
WI	Wisconsin Medical Injury Prevention Program
HHC	National Quality Forum Endorsed Home Health Care Measures
NH	National Quality Forum Endorsed Nursing Home Measures
SC	Surgical Care Improvement Project
RD	RAND-defined measures
CP	Concepts suggested by Delphi participants

Table 12
Group 2 Measures – Highly Important Inpatient Patient Safety Outcome Measures That Are Not Yet Well Validated

Outcome Measure	Importance Ratings		Validity Ratings	
	Median	20th Percentile	Median	20th Percentile
Surgical Inpatient				
NY 1 – Wrong patient, wrong site surgical procedure	9	7	6.5	5
PSI 5 – Foreign body left in during procedure, Secondary diagnosis field	8	7	6	4
NY 3 – Unintentionally retained foreign body	8	7	5.5	4
NS 26 – Systemic Sepsis (Sepsis) post-surgery	8	7	5	3
NQF 4 – Foreign object retention after surgery	8	7	5	4
NS 27 – Systemic Sepsis (Septic Shock) post-surgery	8	7	4	3
Non-surgical Inpatient				
NY 9 – Medication error occurred that resulted in permanent patient harm	9	8	5	4
NQF 12 – Death or serious disability due to medical error	9	7	6	4
NY 11 – Medication error occurred that resulted in patient death	9	7	5.5	4
NQF 26 – Sexual assault on a patient within or on the grounds of the health care facility	8.5	7	6	3
NY 32 – Rape of a patient	8.5	7	6	3
NQF 17 – Severe pressure ulcers acquired after admission to a healthcare facility	8	7	6	4
NQF 27 – Death or significant injury of a patient or staff member resulting from a physical assault	8	7	6	3
NY 10 – Medication error occurred that resulted in near-death event	8	7	5	3
NQF 24 – Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	8	7	5	3

Table 13
Group 3 Measures – Other Important Inpatient Patient Safety Outcome Measures That Are Not Well Validated

Outcome Measure	Importance Ratings		Validity Ratings	
	Median	20th Percentile	Median	20th Percentile
Surgical Inpatient				
NY 2 - Incorrect procedure or treatment - invasive	9	6	6	3
NSPM 7 - Central line-associated bloodstream infection (CLABSI) rate for intensive care unit locations	8	6	6.5	5
NS 20 - Myocardial infarction intra/post-surgery	8	6	6	4
NS 19 - Cardiac arrest requiring CPR post surgery	8	6	6	4
STS 7 - Risk-adjusted operative mortality for aortic valve replacement (AVR)	8	6	6	5
STS 8 - Risk-adjusted operative-mortality for mitral valve replacement/repair (MVR)	8	6	6	5
STS 9 - Risk-adjusted operative-mortality for MVR + CABG	8	6	6	5
STS 10 - Risk-adjusted operative mortality for AVR + CABG	8	6	6	5
PSI 12 - Postoperative pulmonary embolism (PE) or deep vein thrombosis (DVT)	8	6	6	4
NSPM 8 - Central line-associated bloodstream infections (CLABSI) rate for birthweight categories in neonatal intensive care unit locations	8	6	6	4
NS 3 - Organ/Space surgical site infection (SSI)	8	6	6	4
NQF 14 - Maternal harm in labor and delivery of low-risk pregnancy	8	6	5.5	4
SC 5 - Patients diagnosed with post-operative ventilator-associated pneumonia (VAP) during index hospitalization	8	6	5	4
PSI 17 - Birth trauma - Injury to neonate	8	6	5	4
PSI 20 - Obstetric trauma - cesarean delivery	8	6	5	3
PSI 13 - Postoperative sepsis	8	6	5	3
NS 25 - Systemic Sepsis (SIRS) post-surgery	8	6	5	3
MP 4 - Insertion-site infections associated with c	8	5	6	5
PSI 18 - Obstetric trauma - vaginal delivery with instrumentation	8	5	5	4
SC - 3 Intra/post-operative pulmonary embolism (PE) diagnosed during index hospitalization and within 30 days of surgery	8	5	4	3
Surgical and Non-surgical Inpatient				
NSPM 9 - Ventilator-associated pneumonia (VAP) rate for intensive care unit (ICU) locations	8	6	6	5

NSPM 10 - Ventilator-associated pneumonia (VAP) rate for birthweight categories in the neonatal intensive care unit (NICU) locations	8	6	5	4
NY 12 - New acute pulmonary embolism	8	5.5	5	4
NQF 16 - Failure to treat hyperbilirubinimia	8	5	6	4
NQF 8 - Air embolism	8	5	5	4
Non-surgical Inpatient				
NY 31 - Infant discharged to the wrong family	9	6.5	6.5	4
NQF 19 - Patient harm from electric shock	8	6	6	3
NY 20 - Misadministration of radiation or radioactive material	8	6	5	4
PSI 2 - Death in low mortality DRGs	8	6	5	2
NQF 21 - Patient harm from burns	8	6	5	3
NY 15 - Patient burns	8	6	5	3
NY 8 - Malfunction of equipment with harm	8	6	4.5	3
NY 4 - Unexpected deaths	8	6	4	3
CP 15 - Administration of enteral medications/solutions intravenously	8	6	3	1
CP 7 - Hours of overwhelming pain in terminal patients	8	6	1	1
NQF 23 - Patient harm from use of restraints	8	5	6	4
PSI 4 - Failure to rescue	8	5	5	2
NQF 20 - Delivery of contaminated gas to a patient	8	5	5	3.5
NQF 11 - Patient suicide, or attempted suicide	8	5	5	3
NY 21 - Crime resulting in death or serious injury	8	5	5	3
CP 5 - Burns/scalding during bathing	8	5	2.5	1
NQF 7 - Patient harm from improperly used device	8	4	5	3
NY 6 – Loss of limb or organ	8	2	5	3

APPENDIX A.
LIST OF PARTICIPANTS IN THE IMPORTANCE RATING PROCESS

Name		Degree	Professional Affiliation
Last	**First**		
Annas	Catherine	JD	Massachusetts Department of Public Health
Barnard	Cindy	MS, MBA	Northwestern Memorial Hospital
Bottei	Edward	MD	Iowa Health System, University of Iowa
Burdick	Twila	MBA	Banner Health
Carayon	Pascale	PhD	University of Wisconsin School of Medicine
Combes	John	MD	Center for Healthcare Governance
Crittenden	Michael	MD	Harvard Medical Schoool, VA Boston Healthcare System
Dunagan	W. Claiborne	MD, MS	BJC HealthCare, Washington University School of Medicine
Duthie	Elizabeth	RN, PhD	NYU Hospitals Center
Elson	Robert	MD, MS	MetroHealth
Fairbanks	Rollin	MD, MS	University of Rochester
Feldman	Penny	PhD	Visiting Nurse Service of New York
Ferris	Timothy G.	MD, MPH	Partners HealthCare
Floersheim	Elizabeth A.	RN	Rural Healthcare Quality Network
Foster	Nancy	MS	American Hospital Association
Fried	Marvin	MD	Montefiore Medical Center
Friedman	Amy L.	MD	Yale University
Gandhi	Tejal K.	MD, MPH	Partners HealthCare
Graumlich	James F.	MD	University of Illinois at Chicago
Guise	Jeanne-Marie	MD, MPH	Oregon Health and Science University
Hilborne	Lee	MD, MPH	UCLA Medical Center, RAND Corporation
Kaplan	Harold	MD	Columbia University Medical School
Kovner	Christine	RN, PhD	New York University
Landrigan	Christopher	MD, MPH	Harvard Medical School
Lapane	Kate	PhD	Brown University
Levett	James	MD	Physicians' Clinic of Iowa

Name		Degree	Professional Affiliation
Last	**First**		
Liebovitz	David	MD	Northwestern Memorial Hospital
Lynn	Joanne	MD, MS, MA	RAND Corporation
Metersky	Mark	MD	University of Connecticut School of Medicine
Meurer	Linda	MD, MPH	Medical College of Wisconsin
Mitchell	Pamela	RN, PhD	University of Washington School of Nursing
Mueller	Keith	PhD	University of Nebraska Medical Center
Needleman	Jack	PhD	UCLA School of Public Health
Noskin	Gary	MD	Northwestern Memorial Hospital
Patterson	Mary	MD, MEd	Cincinnati Children's Hospital Medical Center
Poon	Eric	MD, MPH	Partners HealthCare
Pronovost	Peter	MD, PhD	Johns Hopkins University School of Medicine
Ralston	James D.	MD, MPH	Group Health Cooperative of Puget Sound
Rosen	Amy	PhD	Boston University
Schiff	Gordon	MD	Rush Medical College, Cook County Hospital
Shenep	Jerry	MD	St. Jude Children's Research Hospital
Small	Stephen	MD	University of Chicago
Stone	Patricia	RN, MPH, PhD	Columbia University School of Nursing
Teigland	Christie	PhD	New York Association of Homes and Services for Aging (NYAHSA)
Thomas	Eric	MD, MPH	University of Texas-Houston Medical School
Trinkoff	Alison	RN, ScD	University of Maryland, Baltimore
Weissman	Joel	PhD	Harvard Medical School, Massachusetts General Hospital

APPENDIX B.
PARTICIPANTS' COMMENTS ON THE INITIAL SET OF CANDIDATE MEASURES

A NATIONAL FOCUS FOR DETERMINING MEASURE IMPORTANCE

Several questions were raised about the focus and goals of the process to select the most important patient safety outcome measures. The focus is national, reflecting AHRQ's scope of responsibility, although it is anticipated that individual providers or other organizations also might choose to use the measures. As stated in the introductory email, the goal of the Delphi process is to identify a set of patient safety outcome measures, from among the larger set of candidate measures, for the following four purposes:

1. To evaluate the impact of the national patient safety initiative on patient outcomes
2. To be included in the national patient safety database as part of the PSO program
3. For continued use by AHRQ as part of an ongoing, national-level monitoring system on patient safety performance and issues
4. For use by providers for their own safety monitoring and performance benchmarking

APPROACH TO COORDINATING PARTICIPANTS' COMMENTS

As stated in the invitational email for the Delphi process, RAND is the facilitator for the Delphi consensus process, through which participants identify the most important patient safety outcomes for monitoring and evaluation purposes. Therefore, RAND has striven to be inclusive in communicating both the additional suggested candidate measures and issues raised, for consideration by participants during the process. With this information, participants make their own judgments regarding which measures are most important to use for the stated purposes.

RAND consolidated into one document all of the participants' comments from the first Delphi step (anonymous, but retaining the exact words used in the comments), for consideration by all participants during the rating process. In addition, we included almost all the measures suggested by participants, excluding only those that were process measures (see discussion below) or were not defined sufficiently to allow them to be included.

Focus on Outcome Concepts in Performing Ratings of Importance

In rating the importance of the patient safety outcome measures, please focus on the ***basic outcome concepts*** represented by the candidate measures. Additional measurement issues, of which there are many (as highlighted in participants' comments), can then be addressed for those measures that represent the outcome concepts identified as the most important. The rating form prepared for your use organizes the candidate measures by content groups (e.g., inpatient surgical, non-surgical) and within them by functional areas. Similar measures are clustered together within the functional areas so you can compare them during the rating process.

Please rate the importance of each measure based on your own judgment regarding (1) how important the patient outcome concept is, and (2) the relative strength of each of the similar measures in representing that concept. If similar measures receive equal importance ratings, they

will be retained if rated highly, and selection among similar measures will be done in later rating or discussion stages of the process.

ADDITIONAL MEASURE SETS AND INDIVIDUAL MEASURES SUGGESTED

A number of additional measure sets were suggested by the Delphi participants, and outcome measures from all of these suggested sets were added to the candidate measure set. The document entitled ***Summary New Measure Sets*** presents brief descriptions of these sets and identifies which measures from them were added to the candidate set. Again, many of these new measures are very similar to some of those already in the candidate set, but we added them to provide the participants the fullest possible set of choices to consider in the rating process.

No measures of patient safety processes of care were added to the candidate set because the focus in this particular process is on identifying important patient safety outcomes. Cogent arguments were made for the importance of process measures, specifically noting that they are essential for documenting and monitoring the extent to which providers are adopting safe practices. Further work is merited for identification of important process measures, including exploration of measurement methods that can be standardized across multiple providers.

MEASUREMENT ISSUES IDENTIFIED

A substantial number of measurement issues were raised by the Delphi participants, with many comments offered on specific measures and suggestions made for approaches to measuring the outcomes of interest. Please consider these comments as you rate the importance of the candidate measures. Measures that you feel represent important safety outcomes should be given higher ratings, even if they have some measurement problems. Highly rated measures will be subjected to further validation, risk adjustment, and other measurement considerations, as necessary, before they could be used.

In addition, several participants pushed back at RAND's attempt to identify denominators for measures that were designed to be counts of events, rather than rates. We note that both counts and rates are measures, with rates serving to standardize counts of events to a well-defined population that is "at risk" of experiencing a specified event. In some settings (e.g., an individual hospital or health care system), it might be possible to identify the population at risk and calculate rates. However, this is not likely to be the case at the state or national levels. Therefore, in acknowledgement of this feedback, we have deleted all the RAND-developed denominators from the relevant measures, replacing them with a statement of "not determined".

MEASURES FOR AMBULATORY CARE AND LONG TERM CARE SETTINGS

The Delphi participants have identified a large number of measures that they believe are applicable to either the ambulatory care or long term care settings. Many of these measures are from the hospital inpatient measure sets, which they felt are also relevant for these other settings. They also identified other sets of measure designed for one or the other of these settings, which have been added to the candidate set.

For each existing measure identified as also applying to other settings, a line has been added to the measure's definition that identifies the setting(s) to which it applies. However, when applied in other settings, these measures should be considered to be closer to concepts than to established measures, because changes might be needed in the definitions of either or both of

the numerators or denominators to tailor them to the new setting. In addition, it is unlikely that any validation work has been done for the measures in the new settings.

SPECIFIC COMMENTS BY DELPHI PARTICIPANTS

Comments on Candidate Measures

I've taken a look at the list. As you point out in the document, some of the issues overlap by different sources of the measures -- something that will be subsequently reconciled.

There are a few issues that I might define differently (e.g., any anaphylactic type transfusion reaction) although I realize that you have incorporated the definitions as the major taxonomies have defined them. I tried to think of others that might add more depth -- the only one I looked for was the Australia Patient Safety Foundation -- although I didn't find the specific site when I quickly searched.

Seems like the toughest thing about this "task" is that it's a process of exclusion -- what's not on the list is a lot harder to assess (in a vacuum) than what is. I pushed what you sent for whether it would capture the common things we see, and it does -- in one of those taxonomies or another. So I think for a starter set for us to start to look at, it's fine -- and of course needs to be shortened and lumped when the time comes as appropriate.

****

There don't seem to be measures that result from unnecessary care

****

Secondly, you have included the NSQIP measures, which are interesting, but burdensome to collect and not broadly adopted. The SCIP measures are more broadly adopted now and becoming increasingly prevalent. Not sure why NSQIP was chosen over the SCIP measures, but would be interested in hearing that logic.

****

I would like more information on what work has been done to validate some of the measures. In previous work on coming up with quality-based measures-- we ran into difficulties with rare events, instability of estimates because of limited denominators, etc, Issues of case-mix adjustment. Should we be thinking about these issues- or just the clinical relevance of these outcomes.

A few years back I was part of a group charged with developing and validating "quality-indicators" for nursing homes (and special post-acute stay residents). The vast majority are more process, than outcome based- but some of the measures may be of interest. The PI was John Morris at the Hebrew Rehab and it was funded by CMS.

****

In general the measures listed tend to ignore what happens with fetal, neonatal and pediatric patients. My concern is that the general areas listed may not be able to pick out the specific safety issues associated with fetal/neonatal/pediatric resuscitation and care. For example, intracranial hemorrhage in near-term neonates (37 wks EGA) is excluded - yet subgaleal hemorrhage in this population (especially if coupled with instrumented delivery) may be a

marker for a significant breech of safety. I would advise more input from maternal-fetal medicine specialists, pediatricians and pediatric subspecialists.

****

I reviewed the list of measures, and in general think it looks reasonable as a starting point. I have a few thoughts. First, like many other projects that have sought high level means to track error rates and adverse event rates across institutions, this list uses some administrative and coding indicators that are highly specific but insensitive for detecting errors (e.g., foreign body left during procedure on the AHRQ PSI list), and others that are sensitive but nonspecific for errors/adverse events having in fact occurred (e.g. sepsis, pneumonia, failure to rescue). This is a ubiquitous problem when using ICD-9 codes and other indicators for errors; within certain limitations, the indicators are very useful, but they have their limits.

In this case, RAND seems to have proposed substantially supplementing the administrative coding methods for extracting safety data with other methods involving chart review or more intensive clinical-level scrutiny. This could potentially address some limitations of the coding approach to detecting safety hazards, but also further complicates matters. I am unclear on how the summary list of 170ish indicators would be used and weighted to track a hospital's performance. How will clinical, chart-review based indicators line up with administrative coding-based indicators?

This gets to a third issue, which is complexity and purpose. What is the goal of this set of indicators? Is the goal to track an institution's performance, benchmarked against other institutions, or to encourage a specific set of safe practices? Is it to define the incidence of certain key patient safety events? Currently, the list is long and complex, and could potentially be tailored differently depending on the purpose, which remains a bit unclear to me. To track an institution's overall performance, or even performance in a set of defined domains, I suspect a shorter list would be adequate and easier to manage.

Fourthly, although complex and diverse in one sense, the list of measures is narrow in another sense: the long list is principally looking at patient care outcomes / events (e.g. medication errors causing death; death within 30 days of a procedure; sepsis; DVT; etc.), which are detectable through chart review or administrative codes, but it is not looking at the many hospital processes that lead up to these events. It would seem to me to be essential to capture information on hospital processes and safety interventions while trying to look at these outcomes of care. Such indicators might include LeapFrog and other measures of safe hospital processes:

Use and penetrance of CPOE

Use of bar coding

Use of automated

Presence of full time intensivists

Volume of key surgical procedures

Nurse staffing – what is the average staffing ratio in a hospital

Nurse work hours – what percentage of nursing shifts exceed 12 hours (actual vs. scheduled hours)?

Physician work hours – what percent of residents work >16 hours shifts? What percentage work >24 hour shifts? What percentage work >80 hours per week?

And perhaps even cultural metrics:

AHRQ patient safety culture survey for institutions

Teamwork rating in units / hospitals of interest

If we do not measure these process indicators, we will not know the level of adoption of safe practices, which may be far more important than knowing the frequency with which patients suffer incidents of inconsistent relation to real safety hazards. Ideally, such process markers could be collected in parallel with a refined list of outcome markers of safety, to allow both a more robust understanding of an institution's safety, and to facilitate studies of process-outcome relationships.

On more last, specific comment re RAND's new proposed measure, readmission rate. Readmission rate as an indicator of quality may be particularly challenging to interpret. While a high readmission rate suggests frequent, erroneously early discharges (or misdiagnoses), a very low readmission rate could suggest that patients are being kept in the hospital longer than needed. You may get a hint of this on your countervailing measure looking at LOS >2SD above the mean, but such a measure would miss subtle, but important consistent over-stays in the hospital.

****

In reviewing the measures I am concerned about 2 major opportunities for errors in addressing the surgical issues.

First, more and more surgical procedures are performed on an outpatient basis and are essentially excluded from data collection. I do think that a procedure as invasive as a cholecystectomy must be included, yet many patients are now sent home on the same day of surgery and are excluded from all of the measures you have listed. I would propose to include all surgical procedures in these measures, not just inpatients.

Second, the length of stay for inpatients is steadily being shortened, which means that many complications of surgery are only being seen, tracked and treated in the outpatient setting. I do not think that any of these measures collect data on complications that occur after hospital discharge but do not require readmission. This gap can lead to important sources of confounding data. For example, wound infection rates among inpatients might appear to be dropping, since patients are sent home before the infections are recognized. In fact, the infection rate might be disturbingly high, but not require rehospitalization, etc.

Third, of course, these points beg the question about surveillance of the entire outpatient medical scene. There are no measures that do so. Thus, it may be necessary to redefine the collective direction as tracking "inpatient" measures of safety. If we can't develop a reliable method of collecting and reporting outpatient care, this will be required.

****

A stated aim is to have a set of measures that can be used to "measure safety trends." This sounds a little vague to me. However, if the main purpose is to measure trends and not to use

safety measures for regulatory, accreditation, accountability or institution-specific public reporting purposes, then I believe a good argument could be made for using readily available claims-based measures that can be analyzed by the government and do not require labor-intensive record review, investigation, etc. The problem is that such measures will not necessarily capture compelling information such as "wrong site" or "wrong patient" procedures. That would require, I believe, a new national reporting system, and I do not think that "trending" objectives alone would justify the time and effort involved in developing reliable national safety reporting. Since many of the measure sets we have been asked to look at require independent reporting and/or record reviews and safety investigations, I think this will be an important consideration when we get to our ratings.

****

In the next round, in which we will be rating measures, it would be most helpful if you could group like measures together. You indicate that you did eliminate duplicate measures where you could. However, the sets you sent out have many closely related/overlapping measures, and it will be difficult to rate them unless we can look at them "side by side" so to speak. For example, there are multiple measures related to death and/or disability, infection, falls, pneumonia, foreign bodies, ulcers, etc. Some have much clearer language and specifications than others.

****

I assume the lists will be merged after the initial evaluation to prevent duplication.

****

These measure sets do not all seem to be predicated on the same definition of AE. For example, the RAND chronic disease measure would not fit the definition of AE as defined by the Utah group. This project should consider specifying a single definition of AE.

****

The intended use of the measures needs to be more clearly defined. Public reporting versus internal reporting or confidential sharing of data would suggest different levels of reportable events and different levels of reliability. The former would tend to be limited to the most serious ones.

-- I think there needs to be more discussion of the elimination of events like abductions or fires. These are safety events that happen within medical care settings, and may be due to the breakdown of systems.

-- Nearly half of the U.S. states already have reporting systems in place, with definitions. Are there any systems or events or definitions that can gleaned from state systems already in practice?

****

There is obviously a lot of overlap between the different sources for items. I assume the winnowing choices will be made at a later date?

Most of the measures in the pdf concern acute events, with easily measurable consequences, primarily as a result of MD or system errors of commission. There is much less about the harder to measure but, to me, equally important errors of omission.

****

There are limited pediatric safety measures in the listing and in general. With the exception of neonatal complications few are included in the listing. It may actually be that additional measures for pediatric acute care need to be generated in order to be comprehensive in scope for the entire population.

Issues and Considerations for Rating Measures

It seems to me that any outcome that requires a root cause assumption should be eliminated. Complexity is very real and assuming that a patient had an MI "due to" and operative procedure is assuming root cause is identified. This may be an attribution of cause and nothing more.

****

My main concern involves relates to some of the infection related measures. First, the NQF has convened an expert panel to address national reporting of healthcare-associated infections. Therefore, I suggest that you work in parallel with NQF although their guideline is still in the early stages of development. With regards to the infection related measures, the biggest issue is that institutions are "rewarded" by not looking for infections. This is especially problematic for small hospitals that may have only one or a percent of one infection control practitioner. In addition, academic medical centers are penalized by high complexity patients or patients transferred from outside hospitals with infections. I really like the RAND defined measures because they are objective and identifiable from administrative data.

****

There are still several duplicate measures. For our next pass, we should keep them clustered according to the topics covered rather than the source for easy comparison.

In general, I would advocate for a category of measures that use unambiguous items that are not dependent upon after the fact billing codes. This will then facilitate real-time electronic reporting as more hospitals and health care settings adopt CPOE. For the proposed measures, I would suggest eliminating the general "other" lumped categories of adverse events such as "Other CNS occurrence" or "Other respiratory occurrence" and focusing on specific items. I would propose removing all the "poisoning" items and consider specific items with toxicity.

****

Very good job at casting the net broadly in this first grouping. The UT / MO use of ICD-9-CM coding is too broad. For example, on UT29 Shock, the ICD9CM is 785, but that also includes tachycardia (785.0) and gangreen (785.4), etc. Also, the poisoning codes do not exclude patients admitted for poisonings.

****

Based on the assertion that “our purpose in this Delphi process is to facilitate national benchmarking and future epidemiological studies, by developing a core set of valid outcomes measures that capture the most important patient safety events,” we have the follow observations and comments:

ICD-9 based measures have intrinsic problems that make them unreliable for assessing patient safety events. The systems that are included in this list (AHRQ PSI, and the Utah/Missouri measures) have not been clinically validated.

Even if some of these measures are reasonable indicators, that is the event being measured is reliably reflected by the administrative codes, it would be impossible to say whether the measure as identified in this list is useful without reviewing the measure specifications in detail.

There are many redundancies remaining in the measure sets (e.g., similar post-op VTE measures exist in all measure sets). Given the inherent inferiority of ICD-9-based measures, there is little reason to review the VTE measures based on billing codes.

The denominators for most measures are too inclusive to be useful. Most differences in performance will be the result of the mix of patients an individual provider sees, rather than a function of complication rates.

For many event types, the number patients at risk is not the most useful denominator. For example, number of central catheter days is the accepted denominator for CRBSI as promulgated by CDC, APIC, and others.

Voluntary reporting systems will result in non-comparable information within and across systems. This is a problem with NYPORTS and NQF.

Although they all reflect valid safety concerns the MPSMS measure set is based on chart abstraction, the MPSMS measures are inadequately specified to render an opinion about their utility as safety measures. Therefore, one cannot determine whether they are appropriately constituted for the stated purpose of national benchmarking. For example, postoperative UTI is not likely to be accurately captured in medical records as a specific diagnosis. The NSQIP specification is much more likely to yield interpretable information.

NSQIP represents the best collective set of measures for patient safety purposes, although it too has significant limitations:

Many measures represent "miscellaneous" categories (NS 5, 10, 14, 18, 21, and especially NS 28) to allow recording of events not captured by explicit definitions. Such measures are likely to be interpreted inconsistently, unless there are additional specifications that we are not provided.

Many measures may represent appropriate quality metrics, but are not clearly patient safety measures (e.g., return to OR in 30 days, death greater than 30 days after procedure).

Of the Rand measures, 2 (RD 1 and 2) are ICD-9 based and suffer from the above mentioned limitations. LOS (RD 3) must be risk adjusted to be interpretable. RD 4 and 5 are unproven as quality measures in a general measurement context.

****

One issue it might be worth "discussing" among the Delphi panel members prior to rating is the issue of aggregate versus disaggregated or discrete measures – for example all of the different poisoning measures used by UT versus a single poisoning measure that including poisoning from multiple sources. What are the advantages and disadvantages of aggregation vs. disaggregation? Are there any decision rules here or is it all measure-dependent?

Another issue is whether the measures rely on record review versus claims. The former is a richer source of data but will be more burdensome and may be subject to greater unreliability

than claims data tied to payment. (Of course claims data have their own reliability issues.) (See my opening note.)

****

Of course, none of us will have time to check all of the ICD-9 codes included for many of these measures, for example the Utah ones, so for me at least, there has to be a caveat that acceptance of many of the measures is only acceptance of the general topic, and not the complete definition.

Similarly, I don't have time at this point to compare all of the similar measures and decide which is best. In addition, there are varying levels of detail provided for the duplicate measures, so impossible to pick one or the other.

Measurement Issues

One important thing is to differentiate what can be measured validly as a rate and what cannot. to be a valid rate, you need an explicit definition for the numerator and denominator, an

independent surveillance mechanism to capture the num and den, and a method for risk adjustment. Outside of healthcare-acquired infections, we do not have many, if any, outcome measures that are valid as rates. For example, we found a 10 fold increase in the rate of dvvt (defined by ahrq psi) when we implemented a standardized surveillance system.

Most of the measures are not valid as rates and we need to explore how to evaluate progress toward them. I believe it is whether we learned from the mistake. This can be done by the presence of a policy, staffs' knowledge of a policy or appropriate use of a policy.

We need to get the science right. When things are presented as rates we need to adhere to epi 101. Most of safety is not measured as rates and we need to find methods to evaluate progress on them.

NS 20 and NS 21 as well as NS 30 may not reflect errors or breaches the in patient safety for centers that tend to operate on high-risk patients who would otherwise not receive surgery. I would propose that some type of exclusions be included for these measures to accommodate patients who are at high risk for a cardiac event or death prior to undergoing surgery.

****

The list includes the NQF serious reportable events, categorizing these things as if they were measures. They are not measures. The NQF has composed a list of events. The NQF has gone far enough to outline criteria for finding cases that may categorized as an event, which would facilitate case finding that is somewhat standard. Other than that, there is no similarity to a measure. There are no numerators, no denominators, and no testing to see if the specifications actually do lead to reliable case finding. In fact, I am amazed --- and somewhat chagrinned to see that instead of simply including what the NQF list says, Rand seems to have made up denominators for these events. As approved by the NQF (and as proposed in the update currently under review) they do not have denominators. Why? Because they are not measures. These events should be stricken from the list.

I am less familiar with the UT/MO list, but suspect that they did not have denominators because they, too, are not measures. I don't think Rand should be making things up as they proceed in

this task. If we need to discuss the uses of events lists versus measures, then let's have that discussion, but let's not create measures where they don't exist.

****

Why is LOS greater than 2SD of the average LOS a measure of **patient safety**? May be a measure of the severity of illness, etc...

****

The described rationale for this collection of measures is to make progress toward being able to estimate national rates for these AEs. However, it would be naïve to think that once accepted, these measures won't be used for assessing quality of care at the individual hospital level. Many, of these measures will not be useful for that purpose. For example, NQF 1,2,3 -They are all important AEs. However, they are rare events for any given hospital. While the occurrence of any of these should be considered a sentinel event worthy of investigation, the large amount of work to determine a hospital "rate", entailing determining denominators, etc., would be a waste of time, when for most hospitals, the rate would probably be 0 or 1. There will be no statistically significant differences, no matter what the denominator is.

Many also will not be useful for hospital-hospital comparison as they won't be valid without some sort of risk adjustment for patient characteristics, for example-most of the surgical measures will be highly dependent on the type of surgery, so the "number of surgical inpatients" as a denominator is inappropriate.

Comments on Measures for Non-Inpatient Settings

Although the focus appears to be entirely on inpatient measures, in our area there is increasing focus on the Emergency Department, which is an important part of the hospital; in part, because a high percentage of our inpatients come in thru the ED. I would like to see some ED safety measures.

****

I have reviewed the measures and find them to be appropriate. One comment that I have is that all measures seem to reflect the hospital phase of care. None of the measures seem to reflect events occurring in the prehospital (EMS) phase of care, though they could.

****

Another stated aim is to include measures applicable to both hospitals and “long-term care.” I think, however, it would be more accurate to say that you have given us “facility-based” measures to comment on. Much long-term care is provided in home settings, where very, very few of the measures in this group would be applicable. NQF has grappled with safety and quality issues in LTC and so far it has got 2 sets of measures – one for nursing facilities and one for home care. (There are also 13 CMS-defined “potential” adverse events for home care.) If you truly want LTC measures too, I would suggest that you look at the two sets of NQF measures (and the CMS adverse events) to see what, if anything, might be relevant to include in the set, rather than mention long-term care as an afterthought. Also, didn’t RAND support the development of “ACOVE?” Is there anything in that measure set? Obviously the “setting” issue is a difficult one to address in trying to develop priority measures. (Note that the CMS adverse

events for home care include measures such as discharge without referral to other services, in instances where significant disability is present.)

****

Pages 6-7 and beyond. A number of the post-operative events from the Medicare system could occur with outpatient surgical procedures (e.g., pneumonia, urinary tract infection, bloodstream infection), depending on the nature of the operation, but the denominator is surgical inpatients. Is there a way to expand the denominator for these, or to create a separate outpatient post-operative event set of questions, with outpatients as the denominator? NQF denominators seem to do this better. I have the same question for some of the NSQIP measures later on (labeled NP)--how shall we best capture post-surgical complications for outpatient surgeries?

Comments on Specific Measures or Denominators

Following are some of my more specific comments:

1) NQF 1 and 3. Why does exigency mitigate concerns about wrong body part
2) NQF 7. How is serious disability defined (this comes up in other measures also – maybe I missed something re definition specification)
3) NQF 12. Same issue re "reasonable differences"
4) NQF 17 AND 20.Relevant to home care.
5) NS MEASURES. Can we be sure that everyone has a post-surgical follow-up at 30 days? What is the source of this info? Might it be biased to picking up problems only for patients who conscientiously keep a follow-up appt? Or suppose they go to a different hospital or MD to address a serious infection, rather than to the hospital responsible for the problem? How will that be captured and attributed?
6) NS 30. We might want to investigate every death within 30 days of procedure, but we could not immediately attribute to poor medical care without that investigation.
7) NY1. This is a good example of aggregating something that was disaggregated in the NQF Measures – i.e., wrong site and wrong patient are together here
8) NY4 and others – how do we define "unexpected"
9) NY11— see NQF 12 – aggregation/discrete issue
10) PS measures. Note that they are all claims-based. See note 4 above.
11) UT Measures. They raised the question in my mind of who and how it is determined than an event is caused by a medical intervention – e.g. UT 6, 7, 10, 11, etc. This will require thorough record review. But perhaps that is what we are after.
12) UT 33 ff. Here are all the poisoning measures!

****

I do not have a suggestion for any additional measures. However, I do have a comment -- it seems to me that the suggested denominator for some of these (e.g., NQF 7 and NQF 12) would lead to a rather serious under-representation of events. I think that using "number of patients (in a

relevant clinical setting)" instead of "number of patients on whom the device was used" for NQF7 or "number of administered the particular medication" for NQF12 would give a much more accurate picture of the problem.

****

Some of the denominators (too inclusive e.g., number of hospital in-patients) in the NQF measures may cause underestimation of the size of the measure and may therefore not be as accurate as need be.

****

-- Medicare Patient Safety Monitoring System (MPSMS)

Some of the categories mention AEs associated with a procedure but without defining what constitutes an AE. For example, Adverse events associated with knee replacements. It is difficult to say what would be counted here.

NQF

--Why give a reference for NASHP (ref 3) instead of the original source?

-- Many of the NQF measures specify death or serious disability as qualifying for reporting, but other less serious outcomes that may also represent serious injury should be measured as well. Compare with NYPORTS 16 (Falls resulting in x-ray proven fractures, subdural or epidural hematoma, cerebral contusion, traumatic subarachnoid hemorrhage, and/or internal trauma)

-- Some of the categories seem too narrow, e.g., death occurring during or after surgery for a patient with surgical risk defined as ASA Class I. Other ASA classes may be important to consider. On the other hand, NYPORTS 14 ("AMI unrelated to cardiac procedure") may go too far, since it does not restrict the outcome to any ASA class, and so may be counting patients with serious cardiac histories.

--PSIs

Have these ever been validated? For example, some of the birth trauma PSIs may be highly confounded by high risk populations.

--RAND RDs

Use of readmissions is not a reliable patient safety issue, especially for non-surgical admissions; They may be related to other quality issues in certain situations, but because they are events which can be controlled by the provider, they should not be included.

****

MP-2 definition should differentiate aspiration pneumonia

MP-9 10, 11 -How will these be defined, there would seem to be crossover with PE, VAP, etc

NG 16- There is a misprint in the measure for hyperbilirubinemia. The numerator is defined by the number of patients with hypoglycemia, rather than hyperbilirubinemia. I think this is carried over from the measure listed above.

There still appears to be a great deal of redundancy in this list: PSI 12 and MP 1 appear to be the same. UT1 and PSI 9 appear to be the same. Multiple redundant codes for PE, pneumonia,

sepsis, post-operative hemorrhage, perforation, decubitus ulcer. Would suggest these could be pared down.

****

Specific Comments:

MP4 Consider deleleting. Not as clinically significant as catheter related BSI

MP8 Consider deleting. The most likely to be AEs are covered in MP5.

NQF 1,2,3,8,9,10,11,18,19,20,21 Consider deleleting. The described rationale for this collection of measures is to make progress toward being able to estimate national rates for these AEs. However, it would be naïve to think that once accepted, these measures won't be used for assessing quality of care at the individual hospital level. Many, if not most, of these measures will not be useful for that purpose. They are all important AEs. However, they are rare events for any given hospital. While the occurrence of any of these should be considered a sentinel event worthy of investigation, the large amount of work to determine a hospital "rate", entailing determining denominators, etc., would be a waste of time, when for most hospitals, the rate would probably be 0 or 1. There will be no statistically significant differences, no matter what the denominator is.

NQF 7- Consider deleting. Too imprecise of a definition to be useful.

NQF 12- Data collection issues. Many more events will be detected by hospitals using automated pharmacy systems, and those simply looking harder. Hospital-to-hospital comparisons therefore meaningless, as will be time trends as more hospitals adopt these systems.

NS5 Consider deleting. Too imprecise of a definition to be useful.

NS 6 overlaps with MS 2

NS 10, 14,18,21,28 Consider deleting. Too imprecise of a definition to be useful.

NS 11 NS12 is better

NS 25, 26, 27 Not useful measures if looking so as to improve care, more useful to look at the care issues that led to the patient to be at risk, ie. surgical site infection.

Most of the NS measures will need to be adjusted for type of surgery.

NY1 Too rare, see above comments.

NY 7 Consider deleting. Too imprecise of a definition to be useful

NY 9, 10,11 See data collection issues per NQF 12

NY 15 too rare

Many other duplicates of others from other measure sets.

UT measures:

I don't know how the data will be collected for most of these. Coding is notoriously inaccurate and in case, the coding will not distinguish between a condition present prior to hospitalization and one occurring during the hospital stay. If coding is not used, every chart will need to be reviewed.

[RAND: RD5 will need some sort of risk adjustment.]

****

NQF 9, Infant discharged to the wrong person: Consider defining infant. I believe that this measure is intended to be applied to infant discharges after birth, but this is not clear. It might cause confusion because infants have variable age definitions, (up to 3-6 months old) and are commonly seen in emergency departments and outpatient facilities after their initial birth hospitalization. Further it is not uncommon for infants to be brought to EDs and outpatient facilities by grandparents or other family members (or in the case of the ED, babysitters). In these cases discharging the infant with the (non-parent) person who brought them in is not an adverse event.

NQF 10, patient elopement: The exclusion references "involving competent adults," but this should instead say: "involving adults with capacity." Competency is a legal term and can only be determined by the courts, not medical providers. Capacity, on the other hand, is a medical term and can be determined by physicians. This is an important distinction because patients who are legally competent may temporarily lack capacity (for example due to delirium from an underlying medical cause). Patients who lack capacity should be protected from eloping.

NQF 12, Medication Error: You might consider including "omitted medication" in the list of examples. Many researchers have included omissions. For example, this might include death from infection after a failure to provide timely antibiotics.

NQF 14, Maternal harm in labor and delivery of low-risk pregnancy: Why do we specify "while being cared for in a health care facility?" Licensed nurse midwifes who perform home deliveries should be held to the same standard. Also, why do we exclude "death from pulmonary embolism" when it is otherwise considered a quality measure to provide appropriate prophylaxis from VTE to hospitalized patients? (see NS 24, PSI 12, NY 12, UT 12, NS 8).

NS 6, Pneumonia: "Number of patients with non-preoperative pneumonia." If this is to be listed as a general measure, it should specify "Number of post-operative patients with non-preoperative pneumonia" or exclude patients admitted for pneumonia.

NS 9. On Ventilator > 48 hours. Consider excluding non-elective (emergency) surgeries performed on high-risk patients. (For example, the 85 year old patient with COPD who presents to the ED with a surgical trauma or a rupturing AAA). Otherwise this is a measure that might discourage appropriate emergency surgeries, or artificially increase quality measure violation rates at tertiary care centers. This applies to NS 20, 29, 30 and others as well.

NS 29. Return to the OR within 30 days. Consider excluding planned sequential events, such as are not uncommon in trauma patients (to the OR day 1 for neurosurgery, and once stabilized to the OR later for less urgent arm fracture).

NY 4, NY 5. Why is Transfusion related death excluded?

NY 5. Cardiac and/or respiratory arrest requiring ACLS. Need to exclude emergency department patients who present in cardiac arrest or near-arrest. (or change denominator to "number of hospital inpatients" as in NY6) Or is this covered by the "relevant settings" statement?

NY 9,10,11. I would consider combining these with NQF 12—I'm not sure what they add independently.

NY 13. Newly documented deep vein thrombosis (DVT). Is this just for new DVTs during inpatient stays? If so this should be specified.

****

Change Decubitus Ulcer to Pressure Ulcer (the recommended term by the National Pressure Ulcer Advisory Panel)

No additional safety measures to answer

****

MP 3 is unclear to what is meant by mechanical events. If it is pneumothorax that measure is covered by PSI 6. MP 3 could be eliminated

Many of the NQF measures use death and serious disability in the definition section and serious harm and death in the numerator section. This could lead to confusion since there is a difference between harm and disability

NY 18 can be eliminated since it doesn't apply or make sense to anyone outside of NYS

There may be definitional problems with PSI 15 since sometimes planned actions such as decompression enterostomies are coded this way. We need to check.

All the UTs are overly broad and depend solely on coding convention which may capture planned and unplanned events

RD 3 and RD 4 may not be safety issues but social issues and should be eliminated.

****

Great set of measures so far – A few thoughts: For a few I wonder about the ability to capture later events, such as post-op thrombosis or wound infections that get picked up in the outpatient setting, or during a subsequent admission.

For MP-9, unclear to me how "adverse event" is defined.

How would one count a VTE following a hip or knee replacement? Under MP-1 or MP-10/11?

NQF-7: what about using a device for the correct function, but using improper technique?

NQF-13: Administration of ABO-incompatible blood, even without harm, seems like an important quality breech.

Check NQF 16 – numerator states hypoglycemia, but measure is re: hyperbilirubinemia.

****

MP 1 - thromboembolic events not well defined

MP 2 - pneumonia is not well defined

NQF 1 - what if laterality on consent is incorrect?

NQF 5 - anesthesiologists often disagree as to whether a patient is ASA 1 or 2. ASA 2 patients are usually very healthy - why not include? Why not include outpatients?

NQF 7 - hard to imagine this being an important class of events since restricted to use for "functions other than intended." Not sure what is meant here...

NQF 12 - serious disability is vague

NS 6 - "a." in both sets of definitions is by far the weakest of the qualifiers; surprised it was included. Infectious disease is not my primary area of expertise but this is a subjective, vague, nonspecific data point

NS 11 - "progressive" is vague, not defined

NS 12 - seems internally inconsistent: "requiring dialysis...whether or not received dialysis."

NS 17 - are positioning injuries to nerves included? Ie not primarily related to the surgical intervention - or are they to be excluded

NS 20 - not sure of rationale as to why followed out to 30 days post op and not inclusive of other validating data (echocardiogram or stress test etc).

NS 21 - non q wave infarctions - definitions not supplied

NS 23 - no post operative time period is provided

NS 25 - SIRS is not defined

NS 26 - why aren't blood cultures an important qualifier?

NS 27 - why aren't blood cultures an important qualifier?

NY 7 - impairment of organ function not defined

NY10 - some exclusions could be "near death event" - which itself is vague. For example, a patient with blood sugar of 25 due to medication error involving insulin could have just had a seizure, or be in near shock with altered mental status etc. The other exclusion example, narcotic overdose, could be a patient with very low respiratory rate and high carbon dioxide - would require ACLS in another 10 minutes but was rescued.

NY 16 - why is a serious injury required for a report? The numerator is incorrect as listed - should be "number of patients who fall during care AND HAVE INJURY" - although I would think "number of patients who fall during care" is the more useful number. Why exclude falls with only soft tissue injuries? Luck separates minor or no injury from major injury given these types of patients prone to falls (elderly, multiple comorbidities, etc)

PSI 1 - Definition seems quite limited in scope.

PSI 3 - why exclude patients from a long term care facility? It seems that the more important distinction - especially in these patients at high risk of this complication - is What condition were they in on admission, and what has occurred since admission? A decubitus ulcer on admission could be documented as to size, stage etc and followed

PSI 4 - why patients from long term care or other acute care facility excluded? Again, seems distinction should be simply to exclude complications they arrived with. They seem at high risk of developing them after admission.

The UT taxonomy seems least useful - many of the measures are broad and vague, and I do not know what value they would have in assisting evaluation of safety dimensions.

RD 1 - I would suggest that the definition read instead "Any myocardial infarction ASSOCIATED with an operative procedure" (within x Days). "As a result of" indicates causation, probably an inappropriate assumption.

RD3 and RD 5 - very good...

****

Page 4 I suggest reconsideration of the poisoning measure (NYPORTS 27). If we properly consider the quality/safety of care as being the function of the integrated procedures of the whole hospital, poisonings due to dietary functions should be considered part of the hospital care delivery system. This is not usually an example of 'criminal activity' or 'acts of god'. Why include poisonings by antibiotics or other drugs (UT) but not poisonings through the dietary dept?

Comments Regarding Potential New Measure Sets to Include

The Robert Wood Johnson Foundation has been engaged in extensive work on nursing performance measurement, that included funding the NQF work on hospital nursing performance measures, but has now gone far beyond that. They commissioned a series of papers on the state of the art, which are available on the following web site: http://www.inqri.org/

****

See NQF measures for home care and nursing homes. See CMS publicly reported outcomes for nursing homes and home care. See also CMS adverse events for home care. See ACOVE. CURRENT LIST OF MEASURES IS AT BEST A HALF-HEARTED EFFORT TO CAPTURE AMBULATORY, HOME CARE OR LTC OUTCOMES.

****

The Wisconsin Medical Injury Prevention Program developed a set of criteria for identifying injuries that occur to patients during the course of medical care using ICD-9 codes, similar to those developed by the Utah/Missouri team. Five of these were targeted “priority indicators” for guiding injury prevention efforts – see attachment for description of these 5 priority indicators as well as a list of all WMIPP criteria and an article summarizing the process. I hope these are helpful.

****

There are a few other lists that are relevant to nursing sensitive outcomes. Some of these lists have system or employee outcomes that are not relevant to measuring patient safety. And, they do not use administrative discharge data, however, I don't think we can always accurately identify things using administrative discharge data. The lists include the National Database of Nursing Quality Indicators (NDNQI), California Nursing Outcomes Coalition (CalNOC) project and the Healthcare Infection Control Practices Advisory Committee (HICPAC).

****

I assume that you have reviewed the national database of the Society of Thoracic Surgeons. Although the database is specific to cardiac surgery, there may be some valid outcomes measures within the over 200 fields used in the database, so it may be worth reviewing. The database form can be obtained from STS headquarters in Chicago.

****

Consider looking at the VHA measures of primary care effectiveness, from their External Peer Review Program. As you know, these consist primarily of measures of treatment effectiveness (e.g., Hba1c control, BP control, etc.) and measures of the % of time a recommended

procedure/test/referral is actually performed by the facility--i.e., compliance with the electronic physician reminder system. These are not the acute events in the pdf, but they are powerful indicators of the facility's ability to deliver quality and safe care. Or am I out of scope here? These are particularly relevant to ambulatory care, but I think a parallel set of measures for long-term care may (or should) exist.

Also, have you considered measures of patient satisfaction? It is well known, of course, that satisfaction is often not a good measure of technical quality of care, but it is a good and important measure of the quality of emotional connection and care (itself certainly a predictor of lawsuits and very probably a predictor, to a degree, of healing/readjustment, etc)

****

Is there a similar list of "process" measures that hospitals have to report on? If not, could this patient safety "outcomes" measures list include items that are more process-oriented (but which have been shown to contribute to bad outcomes)? Some candidate process measures that bear on hospital (and, for the 2nd and 3rd items, ambulatory care) safety performance could include:

- lag time from medication order entry (written or CPOE) to medication administration

- lag time from abnormal (or critically abnormal) test result availability and acknowledgement by responsible provider

- failure to respond appropriately (or in a timely manner) to an abnormal test result (e.g., discontinue or adjust dose of nephrotoxic drug in patient w/ deteriorating renal function - there's some interesting data, by the way, relating this item to the one above); this category could also include failure to notify a patient of abnormal results that require follow-up (e.g., abnormal mammogram or Pap smear)

- measures of adequacy of prescription history reconciliation upon admission (this is an area I have particular expertise in, but would require lengthier discussion)

measures of adequacy of communication handoffs related to setting and/or care provider transitions

If including process measures directly is not appropriate for consideration here, perhaps creating outcome (e.g., harm) categories linked to these processes would be an option.

I can send attachments / further documentation if there is interest in exploring these areas, but don't want to bother now in case they're off base for current review effort.

****

It is worth at least reviewing the NCC-MERP measures, although they are controversial.

****

Here are some suggestions from some metrics we have been using here

High risk medication errors

1. Use of narcan
2. Use of flumazenil
3. Elevated INRs
4. D50 use

Codes

1. NRCPR data
2. Time to transfer
3. Death in pts admitted to floor excluding DNR pts
4. Change in location within 24 hrs after ER admission

Safety goals/JCAHO goals

1. % meds reconciled
2. Timeliness of receipt of/communication of critical test results
3. Safety pause compliance
4. Handoff compliance
5. Pt identification compliance

Technology

1. eMAR- # errors intercepted, alerts overridden, admins without scanning
2. Smart pumps- # errors intercepted, alerts overridden
3. CPOE- # errors intercepted, alerts overridden
4. Pharmacy barcoding - # errors intercepted
5. ADE monitor- # interventions accepted/month

Candidate Measures Relevant to Long-Term or Ambulatory Care

The list below encompasses candidate Measures to consider as particularly relevant to LTC (note that some already define denominator as "Number of patients [in relevant clinical setting] BUT some define denominator as "Number of hospital inpatients" and will need to be expanded to include other appropriate populations):

NQF7-Harm from improperly used device;

NQF10-Elopement (Note: defined for hospital patients only at this point, but critical issue for nursing home residents as well; definition states "patient death or serious disability associated with patient elopement for more than 4 hours" and we would want to change the definition as applicable to LTC residents-patient safety is compromised whether serious injury/disability results from elopement. Exclusion may also not be appropriate for LTC population.)

NQF11-Suicide and attempted suicides with serious injury (defined only for hospital patients at this point, but relevant to LTC)

NQF12-Medication error (Note: defined as "serious disability/harm or death" and again may want to expand that to encompass other adverse impacts medication errors may have on nursing home residents that do not quite constitute 'serious harm' or death).

NQF15-Patient harm from hypoglycemia - (again defined only for hospital patients, but this happens quite often in LTC)

NQF17-Stage 3 or 4 pressure ulcer acquired after admission (Note: exclusion includes patients with Stage 2 "recognized" on admission that progresses to Stage 3-may want to expand to recognize that some facilities are now doing ultra sound scans or using other means to identify underlying tissue damage that may not be recognizable on surface on admission but can emerge soon thereafter as Stage 3 or 4 pressure ulcer).

NQF19-Harm from electric shock (again, should be considered patient safety issue in absence of "death or serious disability").

NQF20-Delivery of contaminated gas (NOTE: really only oxygen would be relevant for LTC in most instances)

NQF21-Harm from burns (again, should be considered patient safety issue in absence of "death or serious disability").

NQF22-Falls (again, should be considered patient safety issue in absence of "death or serious harm" unless "serious harm" is expanded to include lesser adverse impacts such as skin tears).

NQF23-Harm from use of restraints

NY7-Impairment of limb, organ or body function (Note: not clear if exclusion condition is "Any unexpected adverse occurrence directly related to the natural course of the patient's illness or underlying condition." If so, what is "unexpected" vs. "expected"?)

NY8-Malfunction of equipment w/harm (Note: define "serious harm")

NY9-Medication error occurred that resulted in permanent patient harm (Note: similar to NQF 12 above-here "permanent patient harm" vs. above "serious disability/harm or death"-may want to expand to include lesser adverse impacts related to medication error, such as temporary cognitive or functional decline?)

NY10-Medication error that resulted in near death event (Note: excludes "medication error that resulted in need for treatment, intervention, initial or prolonged hospitalization and caused temporary patient harm." Would want to INCLUDE those to track patient safety events for LTC residents, unless encompassed in a different measure of lesser severity such as NY9.)

NY11-Medication error that resulted in patient death

NY15-Burns (Note: NQF21 above probably too broad for population; similarly, this measure counts only "2nd or 3rd degree burns" and excludes 1st degree burns which may want to include for LTC residents).

NY16-Falls (Note: NQF22 "serious harm" is not defined, better defined here but again may want to broaden even further as relates to LTC residents)

NY19-Specific patient transfers from diagnostic and treatment center to hospital (Note: would need to expand definition to define and include "unnecessary" hospitalizations or "frequent" hospitalizations as related to LTC setting).

UT3-Septicemia, bacteremia

UT4-Pneumonia (Note: may need some exclusion for nursing home residents such as condition present on admission or vaccination given?)

UT5-Other infections (Note: review ICD-9 codes for completeness)

UT9-Drug psychoses (Note: consider conditions beyond single ICD-9 code indicated, such as severe delirium conditions).

UT10-Disorders of the nervous system (Note: review for any appropriate exclusions for LTC population)

UT18-Nausea, vomiting, diarrhea (Note: review for any appropriate exclusions for LTC population)

UT19-Disorders of urinary system (Note: review ICD-9 codes included and review possible exclusions)

UT22-Dermatitis

UT23-Decubitus ulcer (Note: Any vs. NQF17 Stage 3 or 4-which is appropriate for LTC)

UT26-Alterations in mental state (Note: review ICD-9 codes included and review for appropriate exclusions for LTC)

UT27-Rash, spontaneous ecchymoses (Note: appropriate exclusions?)

UT29-Shock

UT31-Sudden death

UT32-Respiratory arrest

UT52-Accidental Falls (NOTE: far more broad than NY16 and NQF22)

NOTE: List below could potentially apply to LTC (possibly with appropriate exclusions applied) but may not warrant individual measures for each rather than a single measure to capture any/all of these conditions

UT33-Poisoning by antibiotics and other anti-infectives

UT34-Poisoning by hormones and synthetic substitutes

UT35-Poisoning by primarily systemic agents

UT36-Poisoning by agents primarily affecting blood constituents

UT37-Poisoning by analgesics, antipyretics, antirheumatics

UT38-Poisoning by anticonvulsant and anti-Parkinsonian drugs

UT39-Poisoning by sedatives and hypnotics

UT40-Poisoning by other CNS depressants, stimulants, anesthetics, nervous system agents

UT41-Poisoning by psychotropic agents

UT42-Poisoning by other agents

UT53-UT63-Similar to UT33-42 above, with a few additional agents, except measure "adverse effects" vs. "poisoning" - same comment regarding need for individual vs. composite? Also, question whether poisoning can be considered an "adverse effect" again for simplification purposes, depending on level of detail needed in reporting/tracking/monitoring.

Question regarding inclusion as candidate measure:

UT66-Poisoning (definition of "undetermined whether accidentally or purposely inflicted" IS NOT CONSISTENT with stated guideline to EXCLUDE measures that indicate criminal intent)

NOTE: A much longer list of candidate measures applies to LTC POST-ACUTE, including: all of MP measures; NQF8; NQF15; NS1,NS2, NS4, NS5, NS6, NS8, NS13, NS15, NS19, NS20, NS23-27, NS29-31; NY12-14; NY17; PSI3, PSI8, PSI9, PSI12-15; UT12

****

The medication error items would be useful in an ambulatory setting (NQF 12, NY 9-11).

****

Some of the measures should be definitely applicable to LTC or ambulatory care. Among the MPSMS measures, MP2 could be associated with ambulatory care, while MP3-5 could be associated with LTC. Among the NQF measures, NQF 1-4 could be associated with ambulatory care, while NQF 6-8,12,13,15, 21-23 could be associated with LTC. Among the NSQIP measures, several of them could be associated with ambulatory care: NS 2-6, NS8, NS10. NS13-15, NS17, NS19, 22,24, 26, and 29 could be associated with ambulatory care. Among the NYPORTS measures, NY 1 and NY17 could be associated with ambulatory care, NY4,5,7-9, 12,15-16, and 19 with LTC. Among the AHRQ PSIs, PSI3 could be associated with LTC, and PSI5-7, PSI11-15 with ambulatory care. Among the Utah/Missouri events, UT1,-4, 16,1719,22, 32, could be associated with ambulatory care, and UT23 with LTC. The drug poisoning events could be associated with either LTC or ambulatory care that are listed on p.49-50. UT 52 could be associated with LTC, and the adverse effects drug events could be associated with either ambulatory care or LTC on p. 53-55. The Rand measures, RD 1 and 2, could be associated with ambulatory care.

Because I am not a clinician, it would be important to obtain clinical input on these events to determine if they do pertain to ambulatory care and/or LTC.

****

"Falls" would be applicable in LTC and ambulatory care settings.

****

Several of the measures could also apply to outpatient setting, such as NQF 3 “wrong procedure”, NS 1 “superficial incisional surgical site infection measures – as many procedures are done in the outpatient setting too.

APPENDIX C.
DOCUMENTATION OF EVIDENCE ON VALIDITY

Approach to the Validity Rating Step

The purpose of this Delphi step is to develop consensus information on the strengths and limitations of the measures that participants have identified as addressing important patient safety issues. This information will be provided to the Agency for Healthcare Research and Quality (AHRQ) and the patient safety community. In particular, AHRQ can use this information in decisions for funding additional measure development work, as well as decisions on use of measures for various patient safety activities.

In the previous Delphi step, we asked you to view each measure as representing the concept it was intended to address, without yet considering validity or measurability. It now is time to focus on the integrity of each specific measure that has surfaced from the previous rating step as important for patient safety.

The primary consideration is assessment of measure validity. However, we also provide text boxes on the rating form in which you can provide feedback on measurability issues. (*Note*: In preparing the report on results of this process, RAND also will draw upon comments you have submitted at each step of the process.) As we all know, numerous types of validity are used in research and other measurement processes. We are not going to attempt to work with the terminologies involved in those different validities because their definitions vary widely.

When assessing the validity of each measure being considered, we ask you to address the following question:

> ***For the type of event the measure is intended to represent, how well does the measure accurately capture the frequency of actual events that are occurring during clinical care processes?***

To draw a conclusion on measure validity, the following logical steps should be taken:

1. Determine the extent to which evidence regarding the validity of a measure is available in the published scientific literature.
2. If such published evidence is available, assess what the evidence says about how accurately the measure captures the underlying actual events.
3. In the absence of such published evidence (or limited evidence availability), determine the extent to which you feel it is appropriate to use clinical judgment and a consensus rating process to judge the validity of the measure.

RAND has searched for published papers providing validity evidence for the measures that were identified as important in the previous Delphi process step. We have summarized the evidence we have been able to find, which is presented for your use in this document. However, we might not have found everything that has been published, and validity assessment work also has been done that has not (yet) been published.

Therefore, as you assess measure validity, please consider not only the materials RAND is providing, but also other information you may have from your own work or knowledge of other

work. As you will note in the information we provide, many of the measure sets have not yet been validated with respect to their accuracy in capturing the actual events they are intended to represent.

Validity Consideration for Each Type of Measure

The list of "surviving" candidate measures include three distinct types of measures – event reporting measures, measures based on administrative data, and measures based on patient health record data. Different considerations are involved in assessing the validity of each type of measures (referred to below as validity criteria), which are discussed here.

- **Event reporting measures**. These are events that may be reported by health care providers, either into their own internal reporting systems or to external entities. Typically, sets of these measures have been assessed through the processes used to select the measures, usually through a consensus process. Measure sets included in this category include:
  - NQF Serious Reportable Events
  - New York Patient Occurrence Reporting and Tracking System (NYPORTS).

  Validity criterion: How completely the relevant events are being reported.

- **Administrative measures.** These measures are calculated using administrative data with testing that ranges from validation through chart review to tests for precision, bias, or coding issues. Measure sets in this category are:
  - AHRQ Patient Safety Indicators
  - UT/MO Adverse Events
  - NQF Endorsed Mortality Measures
  - Wisconsin Medical Injury Prevention Program

  Validity criterion: How accurately the counts or rates estimated with these measures match those that would be obtained using data from patient health records.

- **Chart-Based Measures.** Data for these measures are collected directly from patient health records. The chart-based measures in our candidate list include:
  - Medicare Patient Safety Monitoring System
  - National Surgical Quality Improvement Program (NSQIP) *
  - Nursing Sensitive Performance Measures
  - Society of Thoracic Surgeons Measures *
  - NQF Endorsed Home Health Measures *
  - NQF Endorsed Nursing Home Measures *
  - Surgical Care Improvement Project

  Validity criterion: How consistently data is abstracted out of the health records, as measured by inter-rater reliability.

  * Note: Data for the four measure sets followed by "*" are available in currently existing data sets, which contain data abstracted from health records (e.g., MDS data for nursing homes, OASIS data for home health)

The remainder of this document contains summary information on the validation information that RAND was able to find in the published literature for the important measures identified by the Delphi participants. The measures are organized by health care setting

(inpatient, ambulatory care, and long-term care) and within setting, by measure set. Only the measures selected for importance are included in each measure set listing. A summary of the validity evidence that we found for each of the measure sets is provided in the following table.

Summary of Validity Information for Measure Sets and Concepts

Note: When information directly applicable to the validity criterion corresponding to the Measure Set is available, it is in **bold face** in the table below.

Measure Set	Validity Information	Pages
Inpatient Care:		
MPSMS	Inter-rater reliability tests for 3 years	4-5
NQF Serious Reportable Events	Expert consensus process	6-7
NSQIP	Inter-rater reliability tests performed	8-9
NYPORTS	Expert consensus **Validation of reporting against administrative data**	10-11
AHRQ PSIs	Validity concerns documented Chart review of one measure	12-17
UT/MO Adverse Events	Chart review of a subset of measures	18
Nursing Sensitive Performance Measures	Expert consensus process	19
Society of Thoracic Surgeons Measures	Expert consensus Validation of STS data against MEDPAR data Regional tests of inter-rater reliability	20
SCIP	No validity information	21
Concepts		21
Ambulatory Care:		
NQF Serious Reportable Events	Expert consensus process	22
Concepts	No validity information	23
Long Term Care:		
NQF Home Health Measures	Expert Consensus Inter-rater reliability on the OASIS data elements	24
NQF Nursing Home Measures	Expense Consensus Inter-rater reliability of MDS data elements Validity assessment for each measure	25-26
NQF Serious Reportable Events	Expert Consensus	26
Concepts	No validity information	27

INPATIENT MEASURES

MEDICARE PATIENT SAFETY MONITORING SYSTEM

The Medicare Patient Safety Monitoring System (MPSMS) is the Center for Medicare & Medicaid Services' (CMS) national outcomes measure system. The system utilizes trained abstractors to extract data related to the index hospital stay from medical charts. Each year, more than 40,000 Medicare inpatient charts are sent to two Clinical Data Abstraction Centers (CDAC) where the abstractors process each chart.[9] Medical charts are randomly selected from national discharge records of inpatient Medicare recipients. Additionally, the system uses Medicare administrative data to track 30-day post-discharge mortality and re-admission within 30 days. An algorithm is then applied to the extracted data to determine the rate of adverse events.

Validation of the data derived from the chart abstraction is carried out through regular checks of inter-rater reliability. Each month, 40 randomly selected charts are exchanged between the two CDACs and are re-abstracted. The MPSMS contractor, Qualidigm, and the Connecticut Quality Improvement Organization adjudicate any discrepancies between abstracts. Results from the first year of data collection (2002) are summarized below [1].

Completeness measure (average completeness of all cases)	99.5%
Number of cases produced	40,620
Number of cases sampled	480
Inter-rater reliability	96.7%
Overall accuracy (aggregate agreement rate across all data elements in all cases in the sample)	98.0%

Additional validation work for the MSPMS measures is currently in process, focusing on the accuracy of the algorithm applied to the abstracted data. This work compares event identification using the algorithm to event identification by physicians reviewing the source medical charts. However, no published findings on this work are currently available.

---

[9] Hunt, Verzier et al. Fundamentals of Medicare Patient Safety Surveillance: Intent, Relevance, and Transparency. AHRQ Compendium of Research Advances in Patient Safety: From Research to Implementation, 2005.

Measure Number	Measure Name/Definition	Evidence
MP 1	Postoperative venous thromboembolic events	**Inter-rater reliability** (all data elements): 96.66% (2002), 98.18% (2003), 97.10% (2004) none available for individual measure **Algorithm validation per chart review**: no published information available.
MP 4	Insertion-site infections associated with central vascular catheters	**Inter-rater reliability** (all data elements): 96.66% (2002), 98.18% (2003), 97.10% (2004) none available for individual measure **Algorithm validation per chart review**: no published information available.
MP 5	Blood stream infections associated with central venous catheters	**Inter-rater reliability** (all data elements): 96.66% (2002), 98.18% (2003), 97.10% (2004); none available for individual measure **Algorithm validation per chart review**: no published information available.
MP 8	Hospital-acquired blood stream infections	**Inter-rater reliability** (all data elements): 96.66% (2002), 98.18% (2003), 97.10% (2004); none available for individual measure **Algorithm validation per chart review:** no published information available.

NATIONAL QUALITY FORUM (NQF) SERIOUS REPORTABLE EVENTS

The NQF events were generated through a formal expert consensus process to enable standardized data collection and reporting for **all licensed healthcare facilities**. The Steering Committee determined that the primary purpose of the list is for public accountability and that the events must be serious, unambiguous and "usually preventable". The Steering Committee also determined that there must be evidence or formal expert opinion on strategies to reduce the risk of any of the events occurring.[10]

Validation regarding completeness of reporting: No published studies were found.

[10] Additional information available in The National Quality Forum Serious Reportable Events in Healthcare Consensus Report, June 2002. Portions of the report available online at www.qualityforum.org.

Measure Number	Measure Name/Definition	Evidence
NQF 1	Surgery performed on the wrong body part	Expert consensus: NQF process
NQF 2	Surgery performed on the wrong patient	Expert consensus: NQF process
NQF 3	Wrong surgical procedure performed on a patient	Expert consensus: NQF process
NQF 4	Foreign object retention after surgery	Expert consensus: NQF process
NQF 5	Intraoperative or immediately post-operative death in an ASA Class 1 patient	Expert consensus: NQF process
NQF 7	Patient harm from improperly used device	Expert consensus: NQF process
NQF 8	Air embolism	Expert consensus: NQF process
NQF 9	Infant discharged to the wrong person	Expert consensus: NQF process
NQF 11	Patient suicide or attempted suicide	Expert consensus: NQF process
NQF 12	Death or serious disability due to medication error	Expert consensus: NQF process
NQF 13	Administration of AB)-incompatible blood	Expert consensus: NQF process
NQF 14	Maternal harm in labor and delivery of low-risk pregnancy	Expert consensus: NQF process
NQF 16	Failure to treat hyperbilirubinimia	Expert consensus: NQF process
NQF 17	Severe pressure ulcer acquired in a health care facility	Expert consensus: NQF process
NQF 19	Patient harm from electric shock	Expert consensus: NQF process
NQF 20	Delivery of contaminated gas to a patient	Expert consensus: NQF process
NQF 21	Patient harm from burns	Expert consensus: NQF process
NQF 23	Patient harm from use of restraints	Expert consensus: NQF process
NQF 24	Any instance of care ordered by or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	Expert consensus: NQF process
NQF 25	Abduction of a patient of any age	Expert consensus: NQF process
NQF 26	Sexual assault on a patient within or on the grounds of a health care facility	Expert consensus: NQF process
NQF 27	Death or significant injury of a patient or staff member resulting from a physical assault	Expert consensus: NQF process

NATIONAL SURGICAL QUALITY IMPROVEMENT PROGRAM (NSQIP)

The National Surgical Quality Improvement Program (NSQIP) is a nationwide voluntary outcomes measurement program developed in 1994 by the Veterans Administration (VA). Data are collected by surgical clinical nurse reviewers (SCNR) and submitted via the internet to the NSQIP database. Data collected include pre-operative risk factors, intra-operative variables, and 30-day mortality and morbidity outcomes. These risk-adjusted measures are all based on surgical outcomes only and do not include other patient care related outcomes.

Validation of these NSQIP measures requires tests of inter-rater reliability for data extracted by the nurse reviewers. The National VA Surgical Risk Study was conducted in 44 VA Medical Centers (VAMCs) between 1991 and 1993 with the purpose of developing the risk-adjustment models for surgical outcomes which are now used in NSQIP. As part of this study, inter-rater reliability was tested for the preoperative intra-operative and post-operative variables extracted by the nurses at each VAMC.[11] Two nurse coordinators visited each center and abstracted a random sample of 16 cases at each. The results of this study are as follows:

- 41% of the variables collected had perfect to near perfect agreement (kappa values of .81-1.0)
- 29% showed substantial agreement (0.61-0.80)
- 11% showed moderate agreement (0.41-0.60)
- 9% showed fair agreement (0.21-0.40)
- 10% showed poor agreement (<0.20)

Subsequently, in an effort to verify a decrease in post-surgical morbidity in the VA in the 1994-1997 time period, several of the VAMCs conducted inter-rater reliability studies. They did not find any systematic bias toward underreporting.[12] Limited resources prevented similar studies in all 123 VAMCs. Since the American College of Surgeons began offering NSQIP nationally in 2004, a third-party contractor, QCMetrix, has conducted ongoing training and consulting services for nurse reviewers, including annual inter-rater reliability site visits. These annual tests of inter-rater reliability are a requirement of participation in the NSQIP. We do not have access to the QCMetrix inter-rater reliability data.

---

[11] Khuri S et al. Risk Adjustment of the Postoperative Mortality Rate for the Comparative Assessment of the Quality of Surgical Care: Results of the National Veterans Affairs Surgical Risk Study. Journal of the American College of Surgeons 1997; 185:325-338.

[12] Khuri S et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer controlled program for the measurement and enhancement of the Quality of Surgical Care. *Annals of Surgery*; 228:491-507.

In addition, extensive work has been carried out to determine the extent to which the NSQIP morbitity and mortality measures are sensitive to case-mix adjustment based on prospectively collected case-mix adjustors.[313] This work concludes that the morbitity measures, when combined to form a single measure, are mildly to moderately sensitive to case-mix adjustment. Case-mix adjustment sensitivity information is not available separately for the individual morbidity measures.

Measure Number	Measure Name/Definition	Evidence
NS 3	Organ/Space surgical site infection	**Inter-rater reliability**: Tests performed but no data available
NS 19	Cardiac arrest requiring CPR post-surgery	**Inter-rater reliability**: Tests performed but no data available
NS 20	Myocardial infarction intra/post-surgery	**Inter-rater reliability**: Tests performed but no data available
NS 25	Systemic sepsis (SIRS) post-surgery	**Inter-rater reliability**: Tests performed but no data available
NS 26	Systemic sepsis (Sepsis) post-surgery	**Inter-rater reliability**: Tests performed but no data available
NS 27	Systemic sepsis (Septic shock) post-surgery	**Inter-rater reliability**: Tests performed but no data available

[13] Daley et al. Risk Adjustment of the Postoperative Morbidity Rate for the Comparative Assessment of the Quality of Surgical Care: Results of the National Veterans Affairs Surgical Risk Study. Journal of the American College of Surgeons 1997; 185: 339-352.

NEW YORK PATIENT OCCURRENCE REPORTING AND TRACKING SYSTEM (NYPORTS)

The New York Patient Occurrence Reporting and Tracking System (NYPORTS) is a mandatory adverse event reporting system created under New York State law that began operations in 1998. The reporting system is primarily focused on patient outcomes in a hospital setting. Submissions are made electronically via the internet. Event codes are based on New York State-specific medical codes, and serious events leading to death or bodily impairment require the hospital or center to conduct a root cause analysis. The NYPORTS Statewide Council, made up of facility representatives, provider association representatives, and NY Department of Health Staff (NYSDOH), advises the NYSDOH on issues related to NYPORTS, while the refinement subcommittee determines whether codes need to be added, deleted or modified.

The initial validation of these reporting events was the consensus process used to develop the list. Several incidents were specified in the 1998 statute establishing NYPORTS, while most have been added by the NYSDOH upon recommendation of the subcommittees. Additionally, in response to analysis of the first full year of reporting data, the results of which showed large regional and hospital variation in reporting, the New York Department of Health announced that it would use its acute care hospital administrative database to validate hospital reporting.[14] Individual hospitals also undertook internal reviews, and significant under-reporting of events was revealed. For example, there were 1030 cases of "death within 48 hours of a procedure" identified by the administrative data for 1999. Of these, only 167 (16.2%) were reported. As a result, code definitions were refined and additional compliance surveillance and educational sessions for hospital staff have taken place. In 2000, the percentage of cases reported for this code increased to 80%.[15] The University at Albany School of Public Health is also comparing the PSIs to the similar NYPORTS occurrence codes to determine whether reportable codes should be altered in any way.

---

[14] Tuttle D, Panzer R, Baird T. Using administrative data to improve compliance with mandatory state event reporting. *Journal on Quality Improvement*; 28: 349-358.

[15] Information available in NYPORTS: The New York Patient Occurrence Reporting and Tracking System Annual Report 2000/2001. July 2002. Available online at http://www.health.state.ny.us/nysdoh/hospital/nyports/annual_report/2000-2001/annual_report.htm.

Measure Number	Measure Name/Definition	Evidence
NY 1	Wrong patient/wrong site surgical procedure	Expert consensus
NY 2	Incorrect procedure or treatment-invasive	Expert consensus
NY 3	Unintentionally retained foreign body	Expert consensus
NY 4	Unexpected deaths	Expert consensus Reporting completeness: Utilizing 1999 administrative data for validation of reporting, 16.2% of the cases identified in the administrative data were found to be reported to NYPORTS
NY 6	Loss of limb or organ	Expert consensus
NY 8	Malfunction of equipment with harm	Expert consensus
NY 9	Medication error occurred that resulted in permanent patient harm	Expert consensus
NY 10	Medication error occurred that resulted in a near-death event	Expert consensus
NY 11	Medication error occurred that resulted in patient death	Expert consensus
NY 12	New acute pulmonary embolism	Expert consensus
NY 15	Patient burns	Expert consensus
NY 20	Misadministration of radiation or radioactive material	Expert consensus
NY 21	Crime resulting in death or serious injury	Expert consensus
NY 30	Infant abduction	Expert consensus
NY 31	Infant discharged to wrong family	Expert consensus
NY 32	Rape of a patient	Expert consensus

AHRQ PATIENT SAFETY INDICATORS (PSI)

The PSIs are constructed from administrative data and are broadly defined through a series of ICD-9 coding exclusions and inclusions. Validation of these measures would require chart review. **While we identified one study using medical charts to validate PSI 15, Accidental Puncture or Laceration[16], we are not aware of any published studies validating the PSIs identified as "important" by the Delphi group through chart review**. However, an expert panel initially evaluated all the PSIs on six areas of evidence as follows:

- **Precision.** Is there a substantial amount of provider- or community-level variation that is not attributable to random variation?
- **Minimum bias.** Is there either little effect on the indicator of variations in patient disease severity and comorbidities, or is it possible to apply risk adjustment and statistical methods to remove most or all bias?
- **Construct validity.** Does the indicator perform well in identifying true (or actual) quality of care problems?
- **Fosters real quality improvement.** Is the indicator insulated from perverse incentives for providers to improve their reported performance by avoiding difficult or complex cases, or by other responses that do not improve quality of care?
- **Application.** Has the measure been used effectively in practice? Does it have potential for working well with other indicators?

The panel evaluated face validity through a consensus process and evaluated construct validity through a literature review. After a subset of indicators was selected, experts in ICD-9-CM coding reviewed codes to determine their accuracy in capturing the complication (numerator) and the population at risk (denominator).

For each measure, a set of validity concerns is identified, its empirical performance is described and the strength of the evidence from the literature is assessed. [17]

- **Validity Concerns.** The following concerns, raised during the panel review, are listed if they were judged to affect the validity of the particular indicator:

---

[16] Gallagher B, Cen L, Hannan E. Validation of AHRQ's Patient Safety Indicator for Accidental Puncture or Laceration. *Advances in Patient Safety: From Research to Implementation*. Volumes 1-4, AHRQ Publication Nos. 050021 (1-4). February 2005. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/advances/

[17] Additional information available in Guide to Patient Safety Indicators Version 3.0a, May 26. Available online at http://www.qualityindicators.ahrq.gov/psi_download.htm.

- **Rare —** This indicator is relatively rare and may not have adequate statistical power for some providers.
- **Condition definition varies—**This indicator includes conditions for which diagnosis may be subjective, depending on the threshold of the physician, and patients with the same clinical state may not have the same diagnosis.
- **Underreporting or screening—** Conditions included in this indicator may not be systematically reported (leading to an artificially low rate) or may be routinely screened for (leading to a higher rate in facilities that screen).
- **Adverse consequences—** Use of this indicator may have undesirable effects, such as increasing inappropriate antibiotic use.
- **Stratification suggested—** This indicator includes some high risk patient groups and stratification is recommended when examining rates,
- **Unclear preventability—** As compared to other PSIs, the conditions included in this indicator may be less preventable by the health system.
- **Heterogeneous severity—**This indicator includes codes that encompass several levels of severity of a condition that cannot be ascertained by the codes.
- **Case mix bias—** This indicator was felt to be particularly subject to systematic bias, and DRG and co-morbidity risk adjustment may not adequately address the concern.
- **Denominator unspecific—** The denominator for this indicator is less than ideal, because the true population at risk could not be identified using ICD-9-CM codes. Some patients are likely included who are not truly at risk, or some patients who are at risk are not included.

- **Empirical Performance.** The performance of each indicator is measured for the following:
  - **Rate—** The rate measures the number of adverse events per 1,000 population at risk. Rates represent the average rate of the indicator for a nationwide sample of hospitals.
  - **Deviation—** Standard deviation is an estimate of systematic variation. For the PSIs, standard deviation is reported between providers.
  - **Bias—** Bias represents the degree to which the results may be influenced by outside factors. Bias ratings are based on a series of tests of bias using DRG and co-morbidity risk adjustment. Those indicators flagged with X+ demonstrated substantial bias and should be risk adjusted. Those indicators flagged with X also demonstrated some bias. Those without a flag did not demonstrate substantial bias in empirical tests, but may nonetheless be substantially biased in a manner not detectable by the bias tests. Those marked with N/A did not undergo empirical testing of bias due to lack of systematic variation.

- **Strength of Evidence.** The following key findings represent a review of the limited literature assessing the validity of the indicators:
  - **Coding—** Sensitivity is defined as the proportion of patients who suffered an adverse event, based on detailed chart review or prospective data collection, for whom that event was coded on a discharge abstract or Medicare claim. Predictive value is defined as the proportion of patients with a coded adverse event who were confirmed as having suffered that event, based on detailed chart review or prospective data collection.
  - **Construct, explicit process—** Adherence to specific, evidence-based or expert-endorsed processes of care, such as appropriate use of diagnostic modalities and effective therapies. The construct is that hospitals that provide better processes of care should experience fewer adverse events.
  - **Construct, implicit process—** Adherence to the "standard of care" for similar patients, based on global assessment of quality by physician chart reviewers. The construct is that hospitals that provide better overall care should experience fewer adverse events.
  - **Construct, staffing—** The construct is that hospitals that offer more nursing hours per patient day, better nursing skill mix, better physician skill mix, or more experienced physicians should have fewer adverse events.

The following distinctions were used to summarize the strength of the published evidence for each indicator:

-- Published evidence suggests that the indicator lacks validity in this domain (i.e., less than 50% sensitivity or predictive value; explicit or implicit process failure rates no more frequent than among control patients).

0 No published evidence regarding this domain of validity.

± Published evidence suggests that the indicator may be valid in this domain, but different studies offer conflicting results (although study quality may account for these conflicts).

\+ Published evidence suggests that the indicator is valid, or is likely to be valid, in this domain (i.e., one favorable study).

++ There is strong evidence supporting the validity of this indicator in this domain (i.e., multiple studies with consistent results, or studies showing both high sensitivity and high predictive value). When content validity is exceptionally high, as for transfusion reaction or iatrogenic pneumothorax, construct validity becomes less important.

In addition to the work of the panel evaluating the PSIs, one study looked at the sensitivity of PSIs to differences in administrative databases utilizing the Department of Veteran's Affairs (VA) discharge database.[18] Modifications were made to the VA database to make it comparable to the HCUP data. They found that the PSI rates were sensitive to differences in data file structure, definitions and sources of data elements. The fact that PSI rates are inherently low, means that data differences that even slightly change the numerator or denominator can affect the PSI rate and impact comparisons of rates across health systems.

Validation through chart review comparisons:

No published information available for any PSI measures.

[18] Rivard E et al. Applying Patient Safety Indicators (PSIs) Across Health Systems: Achieving Data Comparability. *Advances in Patient Safety: From Research to Implementation*. Volumes 1-4, AHRQ Publication Nos. 050021 (1-4). February 2005. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/advances/

Measure Number	Measure Name/Definition	Evidence
PSI 2	Death in low morality DRGs	Validity Concerns: Heterogeneous severity Empirical Evidence: Provider rate = 2.599 Provider SD = 31.803 Pop. Rate = .620 Bias = X+ Strength of Evidence: + Coding 0 Explicit Process + Implicit Process 0 Staffing
PSI 4	Failure to rescue	Validity Concerns: Underreporting or screening, Heterogeneous severity, Case Mix bias Empirical Evidence: Provider rate = 105.390 Provider SD = 88.150 Pop. Rate = 127.687 Bias = X+ Strength of Evidence: + Coding 0 Explicit Process + Implicit Process ++ Staffing
PSI 5	Foreign body left during procedure	Validity Concerns: Rare, Stratification suggested, Denominator unspecific Empirical Evidence: Provider rate = .071 Provider SD = .340 Pop. Rate = .084 Bias = N/A Strength of Evidence: 0 Coding 0 Explicit Process 0 Implicit Process 0 Staffing
PSI 12	Postoperative PE or DVT	Validity Concerns: Underreporting or screening, Stratification suggested Empirical Evidence: Provider rate = 10.944 Provider SD = 39.208 Pop. Rate = 9.830 Bias = X+ Strength of Evidence: + Coding + Explicit Process

		+ Implicit Process ± Staffing
PSI 13	Postoperative sepsis	Validity Concerns: Condition definition varies, Adverse consequences Empirical Evidence: Provider rate = 17.023 Provider SD = 58.224 Pop. Rate = 10.872 Bias = X+ Strength of Evidence: ± Coding 0 Explicit Process 0 Implicit Process - Staffing
PSI 17	Birth trauma--injury to neonate	Validity Concerns: Condition definition varies, Unclear preventability, Heterogeneous severity Empirical Evidence: Provider rate = 5.425 Provider SD = 17.182 Pop. Rate = 5.531 Bias = N/A Strength of Evidence: - Coding 0 Explicit Process 0 Implicit Process 0 Staffing
PSI 18	Obstetric trauma--vaginal delivery with instrument	Validity Concerns: Unclear preventability, Case mix bias Empirical Evidence: Provider rate = 191.203 Provider SD = 140.435 Pop. Rate = 191.006 Bias = N/A Strength of Evidence: + Coding 0 Explicit Process 0 Implicit Process 0 Staffing
PSI 20	Obstetric trauma—cesarean delivery	Validity Concerns: Unclear preventability, Case mix bias Empirical Evidence: Provider rate = 4.460 Provider SD = 20.871 Pop. Rate = 4.315 Bias = N/A Strength of Evidence: + Coding

		0 Explicit Process 0 Implicit Process 0 Staffing

UTAH/MISSOURI PATIENT SAFETY PROJECT ADVERSE EVENTS (UT/MO)

The Utah/Missouri Patient Safety Project formed an expert panel of raters to classify adverse events by ICD-9-CM and E codes. The twenty-three panelists were asked to rate each code on three dimensions-medical care/causality, harm and preventability. Codes were rated on a scale of 1 to 5 by the panel.

Additionally, two separate studies were performed to validate the use of the ICD-9-CM codes in detecting adverse drug events[19] and surgery-related adverse events[20]. The study of surgery-related events estimated the validity of 23 measures through medical record review. As a group, the 23 measures identified adverse events associated with care management during the hospital stay with an average predictive value positive (PVP) of 45%. PVP is defined as "the proportion of persons identified as having cases who actually do have the condition under surveillance".[21] They identified adverse events specifically related to surgery with a PVP of 37%.

Note: None of the UT/MO measures were rated "important" for inpatient settings.

JCAHO NURSING-SENSITIVE PERFORMANCE MEASURES

The Nursing-Sensitive Performance (NSPM) measures were reviewed and endorsed by NQF in 2003 and 2004.[22] Candidate measures were drawn from existing measurement activities and nursing outcomes initiatives and evaluated by experts on their importance, scientific acceptability, usability and feasibility. In order to be considered as a "nursing-sensitive" indicator, measures

---

[19] Hougland P et al. Performance of International Classification of Diseases, 9th Revision, Clinical Modification codes as an adverse drug event surveillance system. Medical Care 2006; 44(7): 629-36.

[20] Tuinen M et al. Surveillance of Surgery-Related Adverse Events in Missouri Using ICD-9-CM-Codes. *Advances in Patient Safety: From Research to Implementation*. Volumes 1-4, AHRQ Publication Nos. 050021 (1-4). February 2005. Agency for Healthcare Research and Quality, Rockville, MD. www.ahrq.gov/quality/advances.

[21] Center for Disease Control. Guidelines for evaluating surveillance systems. MMWR-37 (S5) 1-18. Publication date 5/6/88. Available online at http://wonder.cdc.gov/wonder/prevguid/p0000112/p0000112.asp.

[22] Background information on the Nursing Sensitive Measures is available from: National Quality Forum. National Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set: A Consensus Report. 2004.

had to have a direct relationship to some element of nurse staffing that has been associated with better quality care or could be quantifiably influenced by nursing personnel.[23]

These NQF-endorsed measures require primary data collection or chart abstraction. We have not identified any published studies of inter-rater reliability for these measures.

Measure Number	Measure Name/Definition	Evidence
NSPM 7	Central line-associated bloodstream infection (CLABSI) for intensive care unit locations	**Expert consensus:** NQF process **Inter-rater reliability**: None
NSPM 8	Central line-associated bloodstream infection (CLABSI) for birthweight categories in neonatal intensive care unit locations	**Expert consensus**: NQF process **Inter-rater reliability:** None
NSPM 9	Ventilator-associated pneumonia (VAP) rate for intensive care unit locations	**Expert consensus:** NQF process **Inter-rater reliability**: None
NSPM 10	Ventilator-associated pneumonia (VAP) rate for birthweight categories in the neonatal intensive care unit locations	**Expert consensus**: NQF process **Inter-rater reliability**: None

[23] Needleman J, Kurtzman E, Kizer K. Performance Measurement of Nursing Care: State of the Science and Current Consensus. Available on the Interdisciplinary Nursing Quality Research Initiative website at http://www.inqri.org/#papers.

SOCIETY OF THORACIC SURGEONS (STS) MEASURES (NQF APPROVED)

The Society of Thoracic Surgeons established a national voluntary cardiac surgery database in 1989, the National Adult Cardiac Surgery Database (NCD), to support national quality improvement efforts. It allows providers to benchmark their risk-adjusted results regionally and nationally and provides data for research.

These STS measures went through the formal NQF consensus process and were approved in 2004 based on expert judgment of their importance, scientific acceptability, usability, and feasibility[24]. Additionally, there are ongoing efforts to validate the STS data in the NCD.[25] The Iowa Quality Improvement Organization has performed several quality audits of data supplied to the database by Iowa hospitals. The result was an overall agreement rate for the audited data variables of 95.8% and 96.2% for the two audits. Additionally, STS worked with the University of Iowa to compare the NCD data to MEDPAR data in an effort to compare the number of CABG cases submitted to NCD with claims submitted to Medicare. They also compared unadjusted mortality rates. The comparison took place over a five year period, 1994-1999. The volume of patients and the in-hospital mortality was consistently slightly higher in the NCD, although the NCD data did not include patients enrolled in health maintenance organizations. The findings suggest that the majority of STS participants were providing complete case records and mortality results. This is consistent with results of a comparison of STS mortality rates for valve surgery with those from the New York State Cardiac Surgery Reporting system database.

Measure Number	Measure Name/Definition	Evidence
STS 2	Deep sternal wound infection rate for CABG surgical patients	**Expert consensus**: NQF process
STS 6	Risk-adjusted operative morality rate for CABG patients	**Expert consensus**: NQF process
STS 7	Risk-adjusted operative mortality for aortic valve replacement (AVR)	**Expert consensus**: NQF process
STS 8	Risk-adjusted operative mortality for mitral valve replacement/repair	**Expert consensus**: NQF process
STS 9	Risk-adjusted operative mortality for MVR + CABG surgery	**Expert consensus:** NQF process
STS 10	Risk-adjusted operative mortality for AVR+CABG	**Expert consensus:** NQF process

[24] Additional information available in The National Quality Forum National Voluntary Consensus Standards for Cardiac Surgery. National Quality Forum 2004.

[25] Welke K et al. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. *Ann Thorac Surg* 2004;77:1137-9.

SURGICAL CARE IMPROVEMENT PROJECT (SCIP)

These measures are currently being aligned with JCAHO/CMS measures and have not been explicitly operationalized nor has there been any validation work performed at this point.

Measure Number	Measure Name/Definition	Evidence
SC 3	Intra- or post-operative acute myocardial infarction diagnosed during index hospitalization and within 30 days of surgery	None
SC 5	Patients diagnosed with post-operative ventilator-assisted pneumonia during index hospitalization	None

CONCEPTS

These measures were suggested by the Delphi participants and have not been explicitly defined or validated.

Measure Number	Measure Name/Definition	Evidence
CP 5	Burns/scalding during bathing	None
CP 7	Hours of overwhelming pain in terminal patients	None
CP 15	Administration of enteral medications/solutions intravenously	None

AMBULATORY CARE MEASURES

The ambulatory measures are largely measures that have been borrowed from other settings and were not specifically designed for ambulatory care. The exceptions are the NQF measures that are applicable to outpatient surgery performed in a "licensed healthcare facility" and the NYPORTS events, which must be reported by hospitals as well as "diagnostic and treatment centers". As such, the majority of these measures are more accurately defined as "concepts" that have not been operationalized in an ambulatory care setting. We divide the ambulatory care measures into measures or concepts applicable to outpatient surgery and those that are non-surgical.

OUTPATIENT SURGERY

National Quality Forum (NQF) Serious Reportable Events

The NQF events were generated through a formal expert consensus process to enable standardized data collection and reporting for **all licensed healthcare facilities**. The Steering Committee determined that the primary purpose of the list is for public accountability and that the events must be serious, unambiguous and "usually preventable". The Steering Committee also determined that there must be evidence or formal expert opinion on strategies to reduce the risk of any of the events occurring.

Measure Number	Measure Name/Definition	Evidence
NQF 1	Surgery performed on the wrong body part	**Expert consensus**: NQF process
NQF 2	Surgery performed on the wrong patient	**Expert consensus:** NQF process
NQF 3	Wrong surgical procedure performed on a patient	**Expert consensus**: NQF process
NQF 4	Foreign object retention after surgery	**Expert consensus**: NQF process

Surgical measures not designed or validated for ambulatory care

Measure Number	Measure Name/Definition	Evidence
NS 3	Organ/space surgical site infection	No validation for this setting
NY 1	Wrong patient, wrong site, surgical procedure	No validation for this setting
PSI 5	Foreign body left in during procedure	No validation for this setting

Non-Surgical

None of the non-surgical measures has been validated for the ambulatory care setting.

Measure Number	Measure Name/Definition	Evidence
NQF 12	Death or serious disability due to medication error	No validation for this setting
NY 9	Medication error occurred that resulted in permanent patient harm	No validation for this setting
NY 10	Medication error occurred that resulted in near-death event	No validation for this setting
NY 11	Medication error occurred that resulted in patient death	No validation for this setting
CP 14	Failure to follow-up on test results	No validation

LONG TERM CARE MEASURES

The Home Health Care Measures and the Nursing Home Measures were designed specifically for this setting, while the NQF Serious Reportable Events are applicable to any licensed healthcare facility. The remaining measures have not been validated for this setting. As such, these measures are more accurately defined as "concepts" that have not been operationalized or validated in long term care settings.

NQF ENDORSED HOME HEALTH CARE MEASURES

The NQF-endorsed home health care measures are all drawn from the Outcome and Assessment Information Set (OASIS) which contains data on all Medicare and Medicaid patients receiving skilled services. The measures were endorsed by NQF in 2005 based on expert judgment about their importance, scientific acceptability, usability and feasibility[26]. One of NQF's stated priorities for research is testing the validity and reliability of the15 measure set.

CMS implemented OASIS in 1999, and a number of studies have tested inter-rater reliability for OASIS data items. A summary of the findings is located in the following table.

[26] Information available in National Voluntary Consensus Standards for Nursing Home Care. National Quality Forum 2004.

Author, Date	Methodology	Findings
Kinatukara, Rosati, Huang. 2005	Used delayed (n=259) and simultaneous (n=105) assessments to test inter-rater reliability as well as comparison of OASIS entries with the CMS 485 or Plan of Care form (n=141) in a large urban agency.	For the delayed assessments, 39 items has poor inter-rater reliability estimates (<.40), 17 had moderate reliability estimates (>.40-.60), 2 had substantial reliability estimates (>.60-.80), and 2 items had excellent estimates (>.80). The simultaneous assessments resulted in 19 items with poor reliability estimates, 24 items with moderate reliability estimates, 19 items with substantial reliability estimates, and 4 items with excellent reliability estimates. The comparison of OASIS data and the CMS 485 showed the greatest inconsistencies between the records for functional status (47.5%), medications (25.5%) and prognosis (17.7%)
Madigan, Fortinsky 2004	Evaluated inter-rater reliability among 88 patients from 21 agencies	Of the 25 OASIS items studied, all but 2 had weighted kappa values of $\geq$ .60
Hittle, Shaughnessey, Crisler et al. 2003	Inter-rater reliability was estimated using independent assessments by two clinicians for a sample of 66 patients from 5 agencies	Of the 38 OASIS items studies, all but 2 had co-efficients >.60, and 25 items had co-efficients > .70.

Measure Number	Measure Name/Definition	Evidence
HHC 11	Emergent Care for improper medication administration, medication side effects	**Expert Consensus**: NQF process **Inter-rater reliability:** not available in the studies we identified

NQF ENDORSED NURSING HOME MEASURES

In 1997, CMS funded a five-year study (the Mega QI Project) in order to develop and validate measures for sub-acute, post-acute and long-term care settings. Relevant measures already in use were identified, and new measures were developed all using data available from the nursing home Minimum Data Set (MDS). The MDS is a core set of screening and assessment elements, including standardized definitions and coding categories, that make up the comprehensive assessment for all residents of nursing facilities certified to participate in Medicare or Medicaid. The Mega QI Project was completed in 2002, and NQF endorsed the measures that came out of that study in 2003.[27] NQF utilized validity information on the indicators that was produced in a study by Abt

[27] Information available in National Voluntary Consensus Standards for Nursing Home Care. National Quality Forum 2004.

Associates.[28] Tests of inter-rater reliability were performed and overall validity was assessed for each measure. Indicators were considered valid when they 1) are accurately measured; and 2) reflect a positive relationship between the care reflected by the measure and the care processes and structures in place to achieve those care processes. Measures were classified as Level I (top validity) Level II (mid validity), and Level III (not validated).

Measure Number	Measure Name/Definition	Evidence
NH 10	Average risk residents with pressure ulcers	**Expert Consensus:** NQF process **Inter-rater reliability**: 98.73% agreement, .83 weighted Kappa **Overall validity:** Ratings were assigned to indicators including 1) high and low risk residents-I, high risk residents-I, low risk residents-II, but were not assigned for "average" risk residents)

NATIONAL QUALITY FORUM (NQF) SERIOUS REPORTABLE EVENTS (LICENSED HEALTHCARE FACILITY)

The NQF events were generated through a formal expert consensus process to enable standardized data collection and reporting for **all licensed healthcare facilities** including long term care. The Steering Committee determined that the primary purpose of the list is for public accountability and that the events must be serious, unambiguous and "usually preventable". The Steering Committee also determined that there must be evidence or formal expert opinion on strategies to reduce the risk of any of the events occurring.

[28] Morris J et al. Validation of long term and post-acute care quality indicators. June 10, 2003. Available at http://www.cms.hhs.gov/NursingHomeQualityInits/35_NHQIArchives.asp

Measure Number	Measure Name/Definition	Evidence
NQF 7	Patient harm from improperly used device	**Expert Consensus**: NQF process
NQF 11	Patient suicide or attempted suicide	**Expert Consensus**: NQF process
NQF 12	Death or serious disability due to medication error	**Expert Consensus:** NQF process
NQF 13	Administration of ABO-incompatible blood	**Expert Consensus**: NQF process
NQF 17	Severe pressure ulcers acquired in a health care facility	**Expert Consensus**: NQF process
NQF 21	Patient harm from burns	**Expert Consensus**: NQF process
NQF 24	Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	**Expert Consensus**: NQF process
NQF 26	Sexual assault of a patient within or on the grounds of the health care facility	**Expert Consensus**: NQF process
NQF 27	Death or significant injury of a patient or staff member resulting from a physical assault	**Expert Consensus**: NQF process

Measures not defined or validated for long term care:

Measure Number	**Measure Name/Definition**	**Evidence**
MP 4	Insertion site infections associated with central vascular catheters	No validation for this setting
MP 5	Blood stream infections associated with central vascular catheters	No validation for this setting
NY 7	Impairment of limb, organ or body function	No validation for this setting
NY 8	Malfunction of equipment with harm	No validation for this setting
NY 9	Medication error occurred that resulted in permanent patient harm	No validation for this setting
NY 10	Medication error occurred that resulted in near-death event	No validation for this setting
NY 11	Medication error that resulted in patient death	No validation for this setting
NY 21	Crime resulting in death or serious injury	No validation for this setting
NY 32	Rape of a patient	No validation for this setting
UT 23	Decubitis ulcer	No validation for this setting
CP 5	Burns/scalding during bathing	No validation for this setting
CP 7	Hours of overwhelming pain in terminal patients	No validation for this setting

REFERENCES

Medicare Patient Safety Monitoring System:

1. Background information describing the MPSMS measures is available at http://www.qualitytools.ahrq.gov/disparitiesreport/2004/browse/browse.aspx
2. Qualidigm. (2006). Report for 2006 NHQR Medicare Patient Safety Monitoring System (MPSMS, Measure Specifications and Algorithms, Year 3, NHRQ. May 30, 2006.
3. Hunt, Verzier et al. Fundamentals of Medicare Patient Safety Surveillance: Intent, Relevance, and Transparency. AHRQ Compendium of Research Advances in Patient Safety: From Research to Implementation, 2005.

NSQIP:

1. Background information describing the NSQIP measures is available at https://acsnsqip.org/main/resources_downloads.asp.
2. Khuri S et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer controlled program for the measurement and enhancement of the quality of surgical care. *Annals of Surgery*; 228:491-507.
3. Khuri S et al. Risk Adjustment of the Postoperative Mortality Rate for the Comparative Assessment of the Quality of Surgical Care: Results of the National Veterans Affairs Surgical Risk Study. *Journal of the American College of Surgeons* 1997; 185:325-338.
4. Daley et al. Risk Adjustment of the Postoperative Morbidity Rate for the Comparative Assessment of the Quality of Surgical Care: Results of the National Veterans Affairs Surgical Risk Study. *Journal of the American College of Surgeons* 1997; 185: 339-352.

Patient Safety Indicators:

1. AHRQ. Guide to Patient Safety Indicators, AHRQ Pub No. 03-R203, Version 3.0a, May 1, 2006.
2. Rivard E et al. Applying Patient Safety Indicators (PSIs) Across Health Systems: Achieving Data Comparability. *Advances in Patient Safety: From Research to Implementation*. Volumes 1-4, AHRQ Publication Nos. 050021 (1-4). February 2005. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/advances/
3. Gallagher B, Cen L, Hannan E. Validation of AHRQ's Patient Safety Indicator for Accidental Puncture or Laceration. *Advances in Patient Safety: From Research to Implementation*. Volumes 1-4, AHRQ Publication Nos. 050021 (1-4). February 2005. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/advances/

Utah/Missouri Measures:

1. Background information describing the UT/MO measures is available through the Utah Department of Health website, at http://health.utah.gov/psi/icd9.htm.
2. Hougland P, Mashetor C, and Xu W. 2002. "The 2002 Report on the Findings of Rating the Utah/Missouri ICD-9-CM Adverse Event Codes" AHRQ HS11885 Grant Report. Utah

Department of Health Office of Health Care Statistics. http://health.utah.gov/psi/pubs/Expertpanel.pdf.

3. Tuinen M et al. Surveillance of Surgery-Related Adverse Events in Missouri Using ICD-9-CM-Codes. *Advances in Patient Safety: From Research to Implementation*. Volumes 1-4, AHRQ Publication Nos. 050021 (1-4). February 2005. Agency for Healthcare Research and Quality, Rockville, MD. www.ahrq.gov/quality/advances.

Nursing Sensitive Performance Measures:

1. National Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set. National Quality Forum 2004.

2. Needleman J, Kurtzman E, Kizer K. Performance Measurement of Nursing Care: State of the Science and Current Consensus. Available on the Interdisciplinary Nursing Quality Research Initiative website at http://www.inqri.org/#papers.

3. Needleman J et al. Nurse staffing and patient outcomes in hospitals. Health Resources and Services (HRSA) Report No. 230-99-0021; February 28, 2001.

NQF Serious Reportable Events:

1. Serious Reportable Events in Healthcare: A Consensus Report. National Quality Forum. (2002).

NYPORTS:

1. Tuttle D, Panzer R, Baird T. Using administrative data to improve compliance with mandatory state event reporting. *Journal on Quality Improvement*; 28: 349-358.

2. Information available in NYPORTS: The New York Patient Occurrence Reporting and Tracking System Annual Report 2000/2001. July 2002. Available online at http://www.health.state.ny.us/nysdoh/hospital/nyports/annual_report/2000-2001/annual_report.htm.

Society of Thoracic Surgeon (STS) Measures:

1. National Voluntary Consensus Standards for Cardiac Surgery. National Quality Forum 2004.

2. Welke K et al. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. *Ann Thorac Surg* 2004;77:1137-9.

CMS Mortality Measures:

1. Krumholz HM et al., An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction, Circulation 2006;113: 1683-1692.

Home Health Care Measures:

1. National Voluntary Consensus Standards for Nursing Home Care. National Quality Forum 2004.

2. Kinatukara S, Rosati R, Huang L. Assessment of OASIS reliability and validity using several methodological approaches. *Home Health Care Services Quarterly* 2005; 24(3); 23-38.

3. Hittle D, Shaughnessy P, Crisler K et al. A study of reliability and burden of home health assessment using OASIS. *Home Health Care Services Quarterly 2003*; 22(4); 43-63.

4. Madigan E, Fortinsky R. Inter-rater reliability of the Outcomes and Assessment Information Set: results from the field. The Gerontologist 2004; 44(5): 689-692.

Nursing Home Measures:

1. National Voluntary Consensus Standards for Nursing Home Care. National Quality Forum 2004.

2. Morris J et al. Validation of long term and post-acute care quality indicators. June 10, 2003. Available at http://www.cms.hhs.gov/NursingHomeQualityInits/35_NHQIArchives.asp

APPENDIX D.
SUMMARY STATISTICS FOR IMPORTANCE RATINGS FOR EACH MEASURE

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
Surgical Inpatient							
NQF 1 - Surgery performed on the wrong body part	9	8	1.16	8.16	1.78	32	high
NQF 2 - Surgery performed on the wrong patient	9	8	1.12	8.22	1.77	32	high
NQF 3 - Wrong surgical procedure performed on a patient	9	8	1.12	8.19	1.73	32	high
NY 1 - Wrong patient, wrong site surgical procedure	9	7	1.21	8.00	1.71	33	high
NY 2 - Incorrect procedure or treatment - invasive	9	6	1.64	7.69	2.10	32	moderate
NQF 13 - Administration of ABO-incompatible blood	9	6	1.66	7.56	2.15	32	moderate
STS 6 - Risk-adjusted operative mortality for CABG	8	7	0.87	7.94	1.12	33	high
NS 26 - Systemic Sepsis (Sepsis) post-surgery	8	7	1.08	7.48	1.48	33	high
NS 27 - Systemic Sepsis (Septic Shock) post-surgery	8	7	1.16	7.66	1.58	32	high
PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	8	7	0.89	7.68	1.14	31	high
NQF 4 - Foreign object retention after surgery	8	7	1.24	7.52	1.77	31	high
NY 3 - Unintentionally retained foreign body	8	7	0.93	7.78	1.21	32	high
STS 2 - Deep sternal wound infection rate (for CABG surgical patients)	8	6.5	1.25	7.40	1.65	30	moderate
NS 20 - Myocardial infarction intra/post-surgery	8	6	1.22	7.15	1.56	33	moderate
NS 19 - Cardiac arrest requiring CPR post surgery	8	6	1.40	7.32	1.95	34	moderate
STS 7 - Risk-adjusted operative mortality for aortic valve replacement (AVR)	8	6	1.35	7.47	1.68	32	moderate
STS 8 - Risk-adjusted operative mortality for mitral valve replacement/repair (MVR)	8	6	1.32	7.41	1.66	32	moderate
STS 9 - Risk-adjusted operative mortality for MVR and CABG surgery	8	6	1.38	7.39	1.80	31	moderate
STS 10 - Risk-adjusted operative mortality for AVR and CABG surgery	8	6	1.36	7.44	1.78	32	moderate
SC 5 - Patients diagnosed with post-operative ventilator-associated pneumonia (VAP) during index hospitalization	8	6	1.38	7.34	1.73	35	moderate

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
PSI 12 - Postoperative pulmonary embolism or deep vein thrombosis (DVT)	8	6	1.31	7.11	1.57	35	moderate
NQF 14 - Maternal harm in labor and delivery of low-risk pregnancy	8	6	1.45	7.18	2.00	28	moderate
PSI 17 - Birth trauma - Injury to neonate	8	6	1.42	7.66	1.82	29	moderate
PSI 20 - Obstetric trauma - cesarean delivery	8	6	1.75	6.96	2.21	27	moderate
PSI 13 - Postoperative sepsis	8	6	1.05	7.63	1.29	32	moderate
NS 25 - Systemic Sepsis (SIRS) post-surgery	8	6	1.32	7.27	1.68	33	moderate
MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	8	6	1.39	7.35	1.67	34	moderate
NSPM 7 - Central line-associated bloodstream infection (CLABSI) rate for intensive care unit locations	8	6	1.17	7.45	1.42	33	moderate
NSPM 8 - Central line-associated bloodstream infection (CLABSI) rate for birthweight categories in neonatal intensive care unit locations	8	6	1.18	7.29	1.47	31	moderate
NS 3 - Organ/Space surgical site infection (SSI)	8	6	1.06	7.24	1.27	29	moderate
NQF 5 - Intraoperative or immediately postoperative death in an ASA Class 1 patient	8	6	1.54	7.58	1.96	31	moderate
MP 1 - Postoperative venous thromboembolic events	8	5.5	1.54	7.00	1.89	35	moderate
SC - 3 Intra/post-operative pulmonary embolism diagnosed during index hospitalization and within 30 days of surgery	8	5	1.59	6.91	2.01	33	moderate
PSI 18 - Obstetric trauma - vaginal delivery with instrumentation	8	5	1.61	6.96	2.14	27	moderate
MP 4 - Insertion-site infections associated with central vascular catheters (CVCs)	8	5	1.31	7.15	1.67	34	moderate
NS 15 - CVA/Stroke post-surgery	7.5	6	1.30	7.10	1.65	30	low
PSI 19 - Obstetric trauma - vaginal delivery without instrumentation	7.5	5	1.59	6.96	2.01	26	low
NSPM 1 - Death among surgical inpatients with treatable serious complications	7.5	4.5	1.86	6.73	2.39	30	low
NS 8 - Pulmonary embolism post-surgery	7	7	1.11	7.32	1.47	34	low
PSI 11 - Postoperative respiratory failure	7	6	1.14	7.15	1.37	33	low
NS 6 - Pneumonia Post-Surgery	7	6	1.03	7.09	1.26	34	low
UT 12 - Pulmonary embolism and infarction (surgical)	7	6	1.24	7.06	1.65	34	low
SC 4 - Intra/post-operative deep vein thrombosis diagnosed during index hospitalization and within 30 days of surgery	7	6	1.43	7.06	1.81	35	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
NS 12 - Acute renal failure post-surgery	7	6	1.13	7.06	1.46	32	low
UT 3 - Septicemia, bacteremia (surgical)	7	6	1.36	6.94	1.85	33	low
NS 2 - Deep incisional surgical site infection (SSI)	7	6	1.10	7.00	1.46	31	low
NY 17 - Post-operative surgical wound infection	7	6	1.13	6.91	1.31	33	low
WI 2 - Surgical site infections	7	6	1.16	6.88	1.34	33	low
NS 17 - Peripheral nerve injury post-surgery	7	6	0.90	7.00	1.26	31	low
MP 2 - Postoperative pneumonia events	7	5.5	1.30	6.80	1.51	35	low
PSI 1 - Complications of anesthesia	7	5.5	1.27	6.50	1.59	30	low
SC 2 - Intra/post-operative acute myocardial infarction diagnosed during index hospitalization and within 30 days of surgery	7	5	1.50	6.52	1.86	33	low
NS 7 - Unplanned Intubation post-surgery	7	5	1.29	6.71	1.53	34	low
NS 24 - DVT/Thrombophlebitis post-surgery	7	5	1.63	6.71	2.14	35	low
NS 11 - Progressive renal insufficiency post-surgery	7	5	1.34	6.09	1.86	32	low
STS 4 - Post-operative renal insufficiency	7	5	1.19	6.22	1.58	32	low
WI 3 - Central venous catheter injuries (with procedure code 38.93)	7	5	1.46	6.24	1.81	34	low
SC 1 - Post-operative wound infection diagnosed during index hospitalization	7	5	1.34	6.70	1.69	33	low
PSI 14 - Postoperative wound dehiscence, Secondary diagnosis field	7	5	1.38	6.25	1.73	28	low
PSI 9 - Postoperative hemorrhage or hematoma	7	5	1.30	6.52	1.50	31	low
UT 1 - Reopening of a surgical site, control of post-procedure hemorrhage	7	5	1.47	6.77	1.69	31	low
STS 3 - Stroke/cerebrovascular accident (for surgical CABG inpatients)	7	5	1.38	6.72	1.69	29	low
NS 16 - Coma >24 hours post-surgery	7	5	1.74	6.48	2.05	29	low
NS 30 - Death within 30 days of surgery	7	5	1.44	6.72	1.73	32	low
SC 6 - Mortality within 30 days of surgery	7	5	1.44	6.67	1.73	30	low
RD 2 - Postoperative shock	7	5	1.51	6.67	1.84	30	low
PSI 15 - Accidental puncture or laceration, Secondary diagnosis field (surgical)	7	5	1.30	6.38	1.66	32	low
UT 48 - Accidental cut, puncture, perforation, or hemorrhage (surgical)	7	5	1.61	6.19	2.02	31	low
UT 11 - Acute myocardial infarction (surgical)	7	4	1.58	6.39	1.84	33	low
NS 9 - On ventilator >48 hours post-surgery	6.5	4	1.54	5.96	1.84	28	low
WI 1 - Perioperative events during non-cardiac surgery	6	5	1.50	6.24	1.89	34	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
STS 1 - Prolonged intubation (ventilation, for CABG surgical patients	6	5	1.50	6.30	1.79	33	low
NS 4 - Surgical wound disruption	6	5	1.38	5.77	1.76	30	low
SC 7 - Readmission within 30 days of surgery	6	5	1.52	5.90	1.92	30	low
PSI 8 - Postoperative hip fracture	6	5	1.44	6.03	1.87	30	low
STS 5 - Surgical re-exploration (for CABG surgical inpatients)	6	5	1.37	6.30	1.73	27	low
NY 14 - Acute mycocardial infarction (AMI) unrelated to a cardiac procedure	6	4	1.85	6.03	2.22	34	low
UT 20 - Complications occurring mainly in the course of labor and delivery	6	4	1.66	6.04	2.20	26	low
NS 29 - Return to the OR within 30 days of surgery	6	4	1.56	5.74	1.83	31	low
PSI 10 - Postoperative physiologic and metabolic derangements	6	4	1.67	5.69	2.11	29	low
NS 22 - Bleeding requiring > 4 units PRBC/Whole blood transfusions within 1st 72 hours postop	6	4	1.59	5.83	2.02	29	low
NS 23 - Surgical Graft/Prothesis/Flap failure	6	4	1.61	5.58	2.00	26	low
UT 2 - Perforation or laceration (surgical)	6	4	1.58	5.87	2.00	30	low
MP 6 - Postoperative urinary tract infections (UTIs)	6	3.5	1.59	5.66	1.94	35	low
NS 13 - Urinary tract infection post-surgery	6	3.5	1.60	5.63	1.99	35	low
CP 18 - Effusions related to improperly positioned central venous catheter	6	3.5	1.81	5.37	2.19	30	low
MP 9 - Adverse events associated with hip joint replacement due to degenerative conditions	6	3	1.73	5.25	2.05	28	low
MP 10 - Adverse events associated with hip joint replacement due to fracture	6	3	1.82	5.29	2.11	28	low
MP 11 - Adverse events associated with knee replacements	6	3	1.82	5.30	2.13	27	low
MP 3 - Mechanical adverse events associated with central vascular catheters	6	3	1.74	5.66	2.06	29	low
PSI 16 - Transfusion reaction, Secondary diagnosis field	6	3	1.78	5.54	2.15	28	low
NS 1 - Superficial incisional surgical site infection (SSI)	5.5	4	1.44	5.75	1.83	32	low
NS 21 - Other cardiac occurrence post surgery	5	4	1.60	5.73	1.93	30	low
UT 5 - Other infections (surgical)	5	4	1.74	5.34	2.07	32	low
NS 10 - Other respiratory occurrences, post operative	5	3	1.65	4.56	1.97	32	low
NS 14 - Other urinary tract occurrence post-surgery	5	3	1.53	4.57	1.81	30	low
NS 5 - Other surgical wound occurrence	5	3	1.70	4.76	1.98	29	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
NS 18 - Other CNS occurrence post-surgery	5	3	1.83	5.10	2.24	29	low
NS 31 - Death greater than 30 days after surgery if in acute care	5	3	1.84	5.41	2.24	29	low
UT 13 - Heart disease (surgical)	5	2	1.65	4.52	2.03	27	low
UT 50 - Surgical operation/procedure as cause of abnormal reaction of later complications	5	2	2.22	4.39	2.47	28	low
UT 17 - Postoperative GI disorders	4	2	1.38	4.07	1.80	28	low
UT 44 - Complication perculiar to specified procedure	4	2	1.82	4.19	2.14	26	low
UT 46 - Other complications of procedures	4	2	1.82	3.81	2.08	26	low
NS 28 - Other occurrences post-surgery	3	2	1.86	3.62	2.19	26	low
UT 51 - Other procedures without mention of misadventure	3	1	1.80	3.38	2.12	24	low
Surgical and Nonsurgical Inpatient							
NSPM 9 - Ventilator-associated pneumonia (VAP) rate for intensive care unit (ICU) locations	8	6	1.29	7.30	1.76	33	moderate
NSPM 10 - Ventilator-associated pneumonia (VAP) rate for birthweight categories in the neonatal intensive care unit (NICU) locations	8	6	1.49	7.16	1.82	25	moderate
NY 12 - New acute pulmonary embolism	8	5.5	1.44	7.14	1.73	35	moderate
NQF 8 - Air embolism	8	5	1.56	7.16	2.07	32	moderate
NQF 16 - Failure to treat hyperbilirubinimia	8	5	2.10	6.93	2.68	28	moderate
NY 13 - Newly documented deep vein thrombosis (DVT)	7.5	5	1.49	6.94	1.81	34	low
PSI 6 - Iatrogenic pneumothorax, secondary diagnosis field	7	6	1.13	6.97	1.51	33	low
CP 11 - Unplanned extubations	7	5	1.67	6.23	2.09	31	low
CP 29 - Re-hospitalization secondary to CHF within 30 days of CHF hospitalization	7	5	1.48	6.45	1.80	31	low
CP 33 - Re-hospitalization secondary to PE within 4 weeks of any hospitalization	7	5	1.58	6.39	1.93	31	low
CP 12 - Unplanned re-intubation	7	4	1.75	6.16	2.11	32	low
NSPM 6 - Urinary catheter-associated urinary tract infection (CAUTI) rate for intensive care unit locations)	7	4	1.56	5.86	1.88	35	low
CP 30 - Re-hospitalization secondary to sepsis within 2 weeks of any hospitalization	6.5	4	1.63	6.31	1.97	32	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
CP 36 - Re-hospitalization secondary to pneumonia within 2 weeks of any hospitalization	6	5	1.48	5.71	1.83	31	low
CP 35 - Re-hospitalization secondary to renal failure within 2 weeks of any hospitalization	6	4	1.45	6.03	1.91	31	low
CP 32 - Re-hospitalization secondary to DVT within 4 weeks of any hospitalization	6	4	1.76	6.19	2.07	32	low
WI 5 - Catheter-related urinary tract infections (UTIs)	6	3	1.70	5.85	2.05	34	low
PSI 7 - Selected infections due to medical care, secondary diagnosis field	6	3	1.79	5.39	2.14	31	low
CP 28 - Re-hospitalization secondary to COPD within 30 days of COPD hospitalization	5.5	4	1.71	5.43	2.13	28	low
UT 4 - Pneumonia	5	3	1.86	4.91	2.25	34	low
UT 49 - Other misadventures of surgical and medical care	5	2	2.02	4.23	2.33	30	low
CP 34 - Re-hospitalization secondary to cellulitis within 2 weeks of any hospitalization	5	2	1.79	4.83	2.13	30	low
CP 31 - Re-hospitalization secondary to a UTI within 2 weeks of any hospitalization	5	2	1.71	4.41	2.01	32	low
CP 37 - Re-hospitalization secondary to malnutrition within 4 weeks of any hospitalization	4	2	1.92	4.55	2.26	31	low
Nonsurgical Inpatient							
NY 9 - Medication error occurred that resulted in permanent patient harm	9	8	1.06	8.00	1.79	36	high
NQF 12 - Death or serious disability due to medical error	9	7	1.01	8.19	1.52	37	high
NY 11 - Medication error occurred that resulted in patient death	9	7	1.30	8.00	1.94	37	high
NQF 9 - Infant discharged to the wrong person	9	7	1.75	7.65	2.45	34	high
NY 31 - Infant discharged to the wrong family	9	6.5	1.68	7.60	2.34	35	moderate
NY 30 - Infant abduction	9	6.5	1.72	7.63	2.41	35	moderate
NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	8.5	7	1.85	7.39	2.58	36	high
NY 32 - Rape of a patient	8.5	7	1.89	7.31	2.54	36	high
NQF 25 - Abduction of a patient of any age	8.5	6	1.80	7.44	2.40	36	moderate

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
NQF 17 - Severe pressure ulcers acquired after admission to a health care facility	8	7	1.32	7.37	1.83	35	high
NY 10 - Medication error occurred that resulted in near-death event	8	7	1.51	7.54	2.18	37	high
NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	8	7	1.80	7.31	2.55	36	high
NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	8	7	1.80	7.36	2.46	36	high
MP 8 - Hospital-acquired bloodstream infections (BSIs)	8	6	1.13	7.56	1.33	34	moderate
NY 20 - Misadministration of radiation or radioactive material	8	6	1.75	7.12	2.28	34	moderate
NY 8 - Malfunction of equipment with harm	8	6	1.56	7.32	2.13	34	moderate
NY 4 - Unexpected deaths	8	6	1.84	7.17	2.29	35	moderate
PSI 2 - Death in low mortality DRGs	8	6	1.50	7.37	1.85	35	moderate
NQF 19 - Patient harm from electric shock	8	6	1.66	6.88	2.12	32	moderate
NQF 21 - Patient harm from burns	8	6	1.39	7.09	1.96	35	moderate
NY 15 - Patient burns	8	6	1.98	6.52	2.58	33	moderate
CP 7 - Hours of overwhelming pain in terminal patients	8	6	1.38	7.32	1.82	34	moderate
CP 15 - Administration of enteral medications/solutions intravenously	8	6	1.96	7.03	2.53	33	moderate
PSI 4 - Failure to rescue	8	5	1.81	6.97	2.29	35	moderate
NQF 20 - Delivery of contaminated gas to a patient	8	5	1.85	7.03	2.43	34	moderate
CP 5 - Burns/scalding during bathing	8	5	1.85	6.82	2.42	34	moderate
NQF 11 - Patient suicide, or attempted suicide	8	5	1.89	6.85	2.29	33	moderate
NQF 23 - Patient harm from use of restraints	8	5	1.52	6.89	1.83	35	moderate
NY 21 - Crime resulting in death or serious injury	8	5	2.23	6.92	2.74	36	moderate
NQF 7 - Patient harm from improperly used device	8	4	1.79	7.12	2.23	33	moderate
NY 6 - Loss of limb or organ	8	2	2.81	6.07	3.08	29	moderate
MP 7 - Ventilator-associated pneumonia (VAP) events	7.5	6	1.44	7.09	1.94	34	low
NQFM 1 - Acute myocardial infarction 30-day mortality	7.5	5	1.38	7.13	1.64	32	low
NQF 15 - Patient harm from hypoglycemia	7	6	1.49	6.82	1.98	33	low
NQFM 2 - Congestive heart failure 30-day mortality	7	5	1.56	6.75	1.92	32	low
NY 29 - Malfunction of equipment with potential for patient harm	7	5	1.50	6.48	1.87	33	low
NSPM 2- Pressure ulcer prevalence	7	5	1.48	6.55	1.97	33	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
NY 22 - Suicides and attempted suicides related to inpatient hospitalization, with serious injury	7	5	1.92	6.64	2.34	33	low
CP 13 - Readmission to ICU within 48 hours	7	5	1.24	6.73	1.55	33	low
CP 14 - Failure to follow-up test results	7	5	1.54	6.79	1.90	34	low
CP 44 - Over anticoagulation	7	5	1.62	6.49	2.05	35	low
MP 12 - Adverse drug events associated with the anticoagulant warfarin	7	5	1.69	6.59	2.09	37	low
MP 13 - Adverse drug events associated with the anticoagulant IV heparin	7	5	1.73	6.44	2.18	36	low
CP 17 - Failure to maintain a safe environment in an MRI room	7	5	1.91	6.39	2.33	33	low
NSPM 4 - Falls with injury	7	5	1.52	6.66	1.88	35	low
CP 21 - Injuries related to use of restraints	7	5	1.42	6.74	1.65	35	low
CP 20 - Injury occurring due to mechanical bed entrapment	7	5	1.62	6.42	2.05	33	low
PSI 3 - Decubitus ulcer	7	4.5	1.69	6.43	2.10	35	low
NY 5 - Cardiac and/or respiratory arrest requiring ACLS	7	4	1.95	6.59	2.24	32	low
UT 23 - Decubitus ulcer	7	4	1.76	6.21	2.23	34	low
NY 27 - Poisoning occurring within the hospital (water, air, food)	7	4	2.21	6.50	2.69	34	low
WI 4 - Anticoagulation injuries	7	4	1.88	6.11	2.25	36	low
CP 45 - Over anticoagulation with heparin	7	4	1.85	6.23	2.24	35	low
NQF 6 - Contaminated product causing patient harm	7	4	2.30	6.14	2.65	36	low
CP 16 - Fires in the OR/surgical suite with patient harm	7	4	1.98	6.59	2.45	34	low
UT 31 - Sudden death	7	3	2.47	6.18	2.89	33	low
MP 17 - Hospital-acquired antibiotic associated clostridiun difficile (C. Diff)	7	3	2.11	5.78	2.52	32	low
NY 7 - Impairment of limb, organ, or body function	7	2	2.95	5.29	3.15	31	low
NQF 22 - Patient falls in a health care facility	6	5	1.49	6.36	1.90	36	low
NSPM 3 - Patient falls	6	5	1.44	6.34	1.71	35	low
UT 52 - Accidental falls	6	4.5	1.55	5.74	1.99	35	low
CP 46 - Hospital acquired hyperglycemia	6	4	1.73	5.81	2.24	31	low
CP 10 - Emergency department return within 72 hours	6	4	1.44	6.00	1.70	32	low
MP 14 - Adverse drug events associated with the anticoagulant low molecular weight heparin or factor Xa	6	4	1.81	5.97	2.31	36	low
NY 16 - Patient falls	6	4	1.43	6.12	1.75	34	low
NSPM 5 - Restraint prevalence (vest and limb)	6	4	1.57	5.75	1.93	36	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
NY 23 - Elopement from hospital resulting in death or serious injury	6	4	2.18	5.94	2.55	35	low
NQF 10 - Patient elopement	6	3.5	1.99	5.51	2.36	35	low
CP 41 - New hyperkalemia	6	3	1.98	5.52	2.39	31	low
CP 6 - Choking incidents involving consumption of inappropriate food	6	3	2.03	5.61	2.46	33	low
RD 5 - 30-day readmission rates	6	3	1.87	5.38	2.24	34	low
CP 50 - Use of narcan, flumazenil, or dextrose	6	3	2.09	5.74	2.44	31	low
MP 15 - Adverse drug events associated with hypoglycemic agents	6	3	1.90	5.78	2.34	36	low
MP 16 - Adverse drug events associated with digoxin	6	3	2.02	5.56	2.42	36	low
CP 40 - Oversedation	6	3	2.06	5.42	2.37	36	low
NY 28 - Hospital fire or other internal disaster disrupting patient care or causing harm to patients or staff	6	1	2.57	5.12	2.99	34	low
CP 8 - Emergency department admission holding time for inpatient beds	5	4	1.48	4.81	1.99	32	low
CP 9 - Departure from emergency department without treatment	5	4	1.50	5.58	1.88	31	low
UT 32 - Respiratory arrest	5	3	1.79	5.53	2.16	32	low
CP 39 - Hospital acquired delirium	5	3	2.03	5.15	2.35	33	low
UT 33 - Poisoning by antibiotics and other antiinfectives	5	3	2.09	5.08	2.48	36	low
CP 16 - IV rate error	5	3	2.09	5.26	2.37	35	low
CP 42 - Over hydration	5	3	1.98	5.09	2.35	34	low
UT 39 - Poisoning by sedatives and hypnotics	5	2.5	2.07	4.94	2.51	35	low
UT 21 - Complications of the puerperium	5	2	2.26	4.84	2.56	25	low
UT 25 - Maternal causes of perinatal morbidity and mortality, drug reactions and intoxications specific to newborn	5	2	2.48	4.96	2.87	23	low
UT 29 - Shock	5	2	2.22	5.09	2.69	34	low
UT 9 - Drug psychoses	5	2	1.80	4.74	2.18	34	low
UT 53 - Adverse effect of antibiotics and other antiinfectives	5	2	2.02	4.61	2.33	36	low
UT 34 - Poisoning by hormones and synthetic substitutes	5	2	2.22	4.61	2.66	33	low
UT 35 - Poisoning by primarily systemic agents	5	2	2.18	4.60	2.56	35	low
UT 59 - Adverse effects of sedatives and hypnotics	5	2	2.10	4.65	2.44	34	low
NY 18 - Other serious occurrence warranting DOH notification	5	1	2.49	4.12	2.70	26	low
NY 24 - Strike by hospital staff	5	1	2.59	4.50	2.93	32	low
NY 26 - Termination of any services vital to continued safe operation	5	1	2.59	5.17	3.02	29	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
CP 49 - Discharge after resuscitation	4.5	3	2.14	4.79	2.48	28	low
UT 16 - Acute GI ulcer, GI bleed, other GI disorders	4	3	2.05	4.76	2.42	33	low
UT 40 - Poisoning by other CNS depressants, stimulants, anesthetics, nervous system agents	4	2.5	2.11	4.71	2.47	35	low
RD 3 - Length of stay	4	2	1.96	4.24	2.29	33	low
UT 54 - Adverse effect of hormones and synthetic substitutes	4	2	1.89	3.94	2.20	34	low
UT 55 - Adverse effects of primarily systemic agents	4	2	2.04	4.09	2.31	35	low
UT 36 - Poisoning by agents primarily affecting blood constituents	4	2	2.18	4.43	2.55	35	low
UT 37 - Poisoning by analgesics, antipyretics, antirheumatics	4	2	2.27	4.51	2.65	35	low
UT 38 - Poisoning by anticonvulsant and anti-Parkinsonian drugs	4	2	2.19	4.51	2.57	35	low
UT 58 - Adverse effects of anticonvulsant and anti-Parkinsonian drugs	4	2	2.15	4.14	2.40	35	low
UT 60 - Adverse effect of other CNS depressants, stimulants, anesthetics, nervous system agents	4	2	2.08	4.39	2.35	36	low
UT 41 - Poisoning by psychotropic agents	4	2	2.13	4.57	2.54	35	low
UT 61 - Adverse effects of psychotropic agents	4	2	2.16	4.14	2.43	35	low
UT 62 - Adverse effect of agents primarily affecting the cardiovascular system	4	2	2.34	4.40	2.66	35	low
UT 42 - Poisoning by other agents	4	1	2.26	4.06	2.65	33	low
UT 66 - Poisoning (undetermined whether accidental or purposely inflicted)	4	1	2.48	4.31	2.84	32	low
UT 57 - Adverse effects of analgesics, antipyretics, antirheumatics	3.5	2	2.13	4.12	2.43	34	low
UT 26 - Alterations in mental state	3.5	1	2.00	3.88	2.34	32	low
UT 15 - Diseases of the respiratory system	3	2	1.73	3.52	2.08	31	low
UT 56 - Adverse effect of agents primarily affecting blood constituents	3	2	2.09	4.09	2.35	34	low
UT 63 - Adverse effect of other drugs, biological, medicinal substances in therapeutic use	3	2	2.18	4.00	2.44	34	low
UT 10 - Disorders of the nervous system	3	1	1.76	3.21	2.14	29	low
UT 18 - Nausea, vomiting, diarrhea	3	1	1.67	3.03	2.10	32	low
UT 24 - Urticaria	3	1	1.49	2.94	1.75	31	low
UT 28 - Epistaxis, hemorrhage from throat	3	1	1.61	3.10	2.04	31	low
UT 30 - Hemoptysis	3	1	2.13	3.86	2.57	29	low
CP 43 - New constipation	3	1	1.90	3.33	2.19	33	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
UT 47 - Complications of medical care, not elsewhere classified	3	1	1.88	3.23	2.18	26	low
NY 25 - External disaster outside the control of the hospital which affects facility operation	3	1	2.17	3.44	2.54	34	low
UT 43 - Certain adverse effects not elsewhere classified	2.5	1	1.94	3.18	2.29	28	low
UT 45 - Complications affecting specified body systems	2.5	1	1.86	3.10	2.25	30	low
UT 14 - Diseases of veins and lymphatics, other diseases of the circulatory system	2	1	1.30	2.72	1.67	29	low
UT 19 - Disorders of urinary system	2	1	1.45	2.69	1.71	29	low
UT 6 - Endocrine disorders	2	1	1.32	2.63	1.69	30	low
UT 7 - Metabolic and immunity disorders	2	1	1.17	2.36	1.57	28	low
UT 8 - Anemias, coagulation defects and hemorrhagic conditions	2	1	1.43	2.66	1.93	29	low
UT 22 - Dermatitis	2	1	1.26	2.53	1.52	32	low
UT 27 - Rash, spontaneous ecchymoses	2	1	1.38	2.55	1.73	31	low
UT 45 - Complications affecting specified body systems	2	1	1.88	3.15	2.18	27	low
UT 43 - Certain adverse effects not elsewhere classified	2	1	2.04	3.14	2.36	29	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
Ambulatory Care							
AMB NQF 1 - Surgery performed on the wrong body part	9	8	1.20	7.90	1.92	21	high
AMB NQF 3 - Wrong surgical procedure performed on a patient	9	8	0.82	8.29	1.06	21	high
AMB NQF 12 - Death or serious disability due to medical error	9	8	0.48	8.67	0.58	21	high
AMB NY 11 - Medication error occurred that resulted in patient death	9	8	0.88	8.29	1.55	21	high
AMB NY 1 - Wrong patient, wrong site surgical procedure	9	7	1.09	8.05	1.53	21	high
AMB NQF 2 - Surgery performed on the wrong patient	9	7	1.34	7.81	1.97	21	high
AMB NY 9 - Medication error occurred that resulted in permanent patient harm	8	8	0.78	8.10	1.51	21	high
AMB NY 10 - Medication error occurred that resulted in near-death event	8	8	1.22	7.62	2.11	21	high
AMB PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	8	7	0.70	8.00	0.92	20	high

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
AMB NQF 4 - Foreign object retention after surgery	8	7	0.67	8.00	0.95	21	high
AMB NS 3 - Organ/space surgical site infection (SSI)	8	5	1.52	6.82	1.78	17	moderate
AMB CP 14 - Failure to follow-up test results	8	5	1.05	7.00	1.25	19	moderate
AMB PSI 6 - Iatrogenic pneumothorax, secondary diagnosis field	7.5	6	1.31	6.85	1.73	20	low
AMB NS 2 - Deep incisional surgical site infection (SSI)	7.5	6	1.17	7.06	1.39	18	low
AMB NQF 18 - Patient death or serious disability due to spinal manipulative therapy	7.5	3	2.22	6.50	2.73	18	low
AMB NS 8 - Pulmonary embolism post-surgery	7	6	1.27	7.11	1.66	19	low
AMB PSI 12 - Postoperative pulmonary embolism or deep vein thrombosis (DVT)	7	6	1.42	7.05	1.78	19	low
AMB RD 2 - Postoperative shock	7	6	1.09	6.83	1.42	18	low
AMB CP 30 - Re-hospitalization secondary to sepsis within 2 weeks of any hospitalization	7	6	1.06	6.50	1.38	18	low
AMB MP 2 - Postoperative pneumonia events	7	5	1.36	6.79	1.84	19	low
AMB NS 6 - Pneumonia post-surgery	7	5	1.51	6.26	1.82	19	low
AMB NS 24 - DVT/Thrombophlebitis post-surgery	7	5	1.36	6.84	1.71	19	low
AMB PSI 13 - Postoperative sepsis	7	5	1.16	7.00	1.45	19	low
AMB NS 26 - Systemic Sepsis (Sepsis) post-surgery	7	5	1.20	6.88	1.41	17	low
AMB NY 17 - Post-operative surgical wound infection	7	5	1.20	6.71	1.57	17	low
AMB UT 1 - Reopening of a surgical site, control of post-procedure hemorrhage	7	5	1.38	6.44	1.82	16	low
AMB NS 15 - CVA/stroke post-surgery	7	5	1.37	6.56	1.63	16	low
AMB CP 27 - Re-hospitalization secondary to asthma within 30 days of asthma hospitalization	7	5	1.47	6.38	1.86	21	low
AMB CP 33 - Re-hospitalization secondary to PE within 4 weeks of any hospitalization	7	5	1.34	6.55	1.61	20	low
AMB PSI 11 - Postoperative respiratory failure	7	4	1.68	6.47	2.07	17	low
AMB NS 29 - Return to the OR within 30 days of surgery	7	4	1.75	6.24	2.08	17	low
AMB CP 23 - Hospitalization secondary to warfarin induced bleeding	7	4	1.71	6.00	2.05	21	low
AMB CP 35 - Re-hospitalization secondary to renal failure within 2 weeks of any hospitalization	7	4	1.23	6.06	1.52	17	low
AMB NS 17 - Peripheral nerve injury post-surgery	6.5	6	1.11	6.56	1.42	18	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
AMB CP 29 - Re-hospitalization secondary to CHF within 30 days of CHF hospitalization	6.5	4.5	1.10	6.20	1.36	20	low
AMB UT 3 - Septicemia, bacteremia (surgical)	6.5	3	2.06	5.94	2.39	18	low
AMB CP 48 - Composite outpatient diabetes care	6.5	2	2.56	5.50	2.88	16	low
AMB NS 22 - Bleeding requiring > 4 units PRBC/Whole blood transfusions within 1st 72 hours postop	6	6	1.07	6.67	1.24	18	low
AMB PSI 15 - Accidental puncture or laceration, Secondary diagnosis field (surgical)	6	6	0.91	6.37	1.21	19	low
AMB NS 4 - Surgical wound disruption	6	5	1.40	6.29	1.90	17	low
AMB PSI 14 - Postoperative wound dehiscence, Secondary diagnosis field	6	5	1.29	6.29	1.72	17	low
AMB CP 32 - Re-hospitalization secondary to DVT within 4 weeks of any hospitalization	6	4.5	1.41	6.10	1.68	20	low
AMB NS 13 - Urinary tract infection post-surgery	6	4	1.49	5.68	1.73	19	low
AMB CP 22 - Hospitalization secondary to warfarin usage	6	4	1.68	5.75	1.89	20	low
AMB CP 24 - Hospitalization secondary to NSAID induced bleeding	6	4	1.84	5.43	2.18	21	low
AMB CP 2 - Patient use of duplicative medications or medications with potentially harmful drug interactions	6	4	1.62	6.33	1.88	21	low
AMB CP 36 - Re-hospitalization secondary to pneumonia within 2 weeks of any hospitalization	6	4	1.45	5.50	1.79	20	low
AMB CP 28 - Re-hospitalization secondary to COPD within 30 days of COPD hospitalization	6	4	1.21	5.61	1.42	18	low
AMB UT 2 - Perforation or laceration (surgical)	6	3.5	1.81	5.45	2.14	20	low
AMB PSI 7 - Selected infections due to medical care, secondary diagnosis field	6	3	2.17	5.44	2.59	18	low
AMB CP 47 - Polypharmacy in the geriatric populatiom	6	3	1.88	5.20	2.14	20	low
AMB CP 3 - Patient use of improper medication for age group	6	3	1.74	5.45	2.11	22	low
AMB UT 52 - Accidental falls	6	3	1.41	5.30	1.75	20	low
AMB CP 34 - Re-hospitalization secondary to cellulitis within 2 weeks of any hospitalization	6	3	1.31	5.58	1.68	19	low
AMB CP 51 - Elevated international normalized ratios	6	2	2.13	5.00	2.58	16	low
AMB CP 26 - Hospitalization secondary to stroke in patients with atrial fbrillation while not on warfarin	5.5	4.5	1.50	5.80	1.77	20	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
AMB NY 19 - Unplanned specific patient transfers from diagnostic and treatment centers to hospital	5.5	3	1.65	5.15	1.98	20	low
AMB CP 37 - Re-hospitalization secondary to malnutrition within 4 weeks of any hospitalization	5.5	3	1.56	5.00	1.88	18	low
AMB CP 1 - Patient home use of incorrect medication	5	4	1.84	5.67	2.15	21	low
AMB UT 16 - Acute GI ulcer, GI bleed, other GI disorders	5	3	1.96	5.17	2.33	18	low
AMB CP 4 - Patient improperly informed on use of home medical equipment	5	3	1.67	5.15	1.87	20	low
AMB CP 31 - Re-hospitalization secondary to a UTI within 2 weeks of any hospitalization	5	3	1.53	5.05	1.84	19	low
AMB RD 4 - Chronic disease admissions	5	2	1.95	4.50	2.35	20	low
AMB UT 39 - Poisoning by sedatives and hypnotics	4.5	3	1.68	4.68	2.08	22	low
AMB UT 59 - Adverse effects of sedatives and hypnotics	4.5	2	2.18	4.55	2.44	22	low
AMB UT 17 - Postoperative GI disorders	4	3	1.34	4.19	1.80	16	low
AMB UT 40 - Poisoning by other CNS depressants, stimulants, anesthetics, nervous system agents	4	3	1.76	4.41	2.17	22	low
AMB UT 41 - Poisoning by psychotropic agents	4	3	1.61	4.18	2.06	22	low
AMB CP 38 - Re-hospitalization secondary to medication non-compliance within 4 weeks of any hospitalization	4	3	1.63	4.30	2.05	20	low
AMB UT 4 - Pneumonia	4	2	1.90	4.28	2.32	18	low
AMB NS 10 - Other respiratory occurrence post-surgery	4	2	1.72	4.06	2.05	17	low
AMB CP 25 - Hospitalization secondary to seizures in patients with known seizure disorder	4	2	1.56	4.00	1.91	18	low
AMB UT 33 - Poisoning by antibiotics and other antiinfectives	4	2	1.45	4.00	2.05	22	low
AMB UT 53 - Adverse effect of antibiotics and other antiinfectives	4	2	2.21	4.36	2.46	22	low
AMB UT 34 - Poisoning by hormones and synthetic substitutes	4	2	1.50	4.09	2.04	22	low
AMB UT 35 - Poisoning by primarily systemic agents	4	2	1.56	3.82	2.06	22	low
AMB UT 38 - Poisoning by anticonvulsant and anti-Parkinsonian drugs	4	2	1.64	4.00	2.14	22	low
AMB UT 60 - Adverse effect of other CNS depressants, stimulants, anesthetics, nervous system agents	4	2	2.26	4.43	2.52	21	low
AMB UT 55 - Adverse effects of primarily systemic agents	3.5	2	2.19	4.09	2.47	22	low
AMB UT 37 - Poisoning by analgesics, antipyretics, antirheumatics	3.5	2	1.68	3.95	2.15	22	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
AMB UT 62 - Adverse effect of agents primarily affecting the cardiovascular system	3.5	2	2.14	4.27	2.47	22	low
AMB CP 53 - Primary care, urgent care or ER visit following PC visit	3.5	1	2.06	3.72	2.37	18	low
AMB UT 19 - Disorders of urinary system	3	2	1.63	3.76	2.05	17	low
AMB UT 54 - Adverse effect of hormones and synthetic substitutes	3	2	2.11	4.09	2.39	22	low
AMB UT 36 - Poisoning by agents primarily affecting blood constituents	3	2	1.86	3.95	2.30	22	low
AMB UT 56 - Adverse effect of agents primarily affecting blood constituents	3	2	2.31	3.95	2.55	22	low
AMB UT 57 - Adverse effects of analgesics, antipyretics, antirheumatics	3	2	2.31	3.95	2.59	22	low
AMB UT 58 - Adverse effects of anticonvulsant and anti-Parkinsonian drugs	3	2	2.29	3.86	2.53	22	low
AMB UT 61 - Adverse effects of psychotropic agents	3	2	2.30	4.27	2.51	22	low
AMB UT 42 - Poisoning by other agents	3	2	1.71	3.47	2.20	19	low
AMB UT 63 - Adverse effect of other drugs, biological, medicinal substances in therapeutic use	3	2	2.14	3.95	2.42	21	low
AMB NS 5 - Other Surgical wound occurrence	2	1	1.70	3.06	2.05	16	low
AMB UT 22 - Dermatitis	2	1	1.47	2.95	1.77	21	low
AMB CP 52 - Hospitalization following primary care visit	2	1	1.88	3.00	2.21	17	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
Longterm Care							
LTC NQF 12 - Death or serious disability due to medical error	9	8	0.85	8.38	1.39	13	high
LTC NY 9 - Medication error occurred that resulted in permanent patient harm	9	8	0.66	8.38	0.87	13	high
LTC NY 11 - Medication error occurred that resulted in patient death	9	8	0.59	8.62	0.87	13	high
LTC NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	9	8	0.95	8.23	1.42	13	high

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
LTC NQF 17 - Severe pressure ulcers acquired after admission to a health care facility	9	7	1.31	7.85	1.86	13	high
LTC NY 32 - Rape of a patient	9	7	0.95	8.23	1.24	13	high
LTC NY 21 - Crime resulting in death or serious injury	9	6	1.28	8.08	1.61	13	moderate
LTC NY 10 - Medication error occurred that resulted in near-death event	8	8	1.27	7.62	2.14	13	high
LTC NQF 13 - Administration of ABO-incompatible blood	8	7	1.06	7.83	1.34	12	high
LTC CP 5 - Burns/scalding during bathing	8	7	0.89	7.46	1.13	13	high
LTC NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	8	7	1.09	7.69	1.55	13	high
LTC NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	8	7	0.85	8.08	1.12	13	high
LTC NQF 7 - Patient harm from improperly used device	8	6	1.37	7.15	1.72	13	moderate
LTC NY 8 - Malfunction of equipment with harm	8	6	0.97	7.38	1.19	13	moderate
LTC NQF 21 - Patient harm from burns	8	6	0.97	7.38	1.19	13	moderate
LTC CP 7 - Hours of overwhelming pain in terminal patients	8	6	1.28	7.38	1.71	13	moderate
LTC MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	8	5	1.40	6.91	1.81	11	moderate
LTC UT 23 - Decubitus ulcer	8	5	1.98	6.77	2.52	13	moderate
LTC NH 10 - Average risk residents with pressure ulcers	8	5	1.75	6.85	2.30	13	moderate
LTC NY 7 - Impairment of limb, organ, or body function	8	5	2.28	6.64	2.87	11	moderate
LTC NQF 11 - Patient suicide, or attempted suicide	8	5	1.50	7.50	1.83	12	moderate
LTC HHC 11 - Emergent care for improper medication administration, medication side effects	8	5	1.62	6.85	2.27	13	moderate
LTC MP 4 - Insertion-site infections associated with central vascular catheters (CVCs)	8	3	2.23	6.45	2.54	11	moderate
LTC PSI 8 - Postoperative hip fracture	7.5	5	1.83	6.50	2.32	12	low
LTC NS 30 - Death within 30 days of surgery	7.5	3	2.28	6.17	2.72	12	low
LTC NY 28 - Hospital fire or other internal disaster disrupting patient care or causing harm to patients or staff	7.5	2	2.50	6.00	3.05	12	low
LTC NQF 8 - Air embolism	7	7	1.02	6.82	1.78	11	low
LTC NQF 20 - Delivery of contaminated gas to a patient	7	6	1.09	7.15	1.34	13	low
LTC NY 15 - Patient burns	7	6	0.86	7.17	1.11	12	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
LTC NH 3 - Residents with moderate to severe pain	7	6	1.67	6.31	2.39	13	low
LTC CP 21 - Injuries related to use of restraints	7	6	1.33	7.23	1.64	13	low
LTC NY 12 - New acute pulmonary embolism (PE)	7	5	2.08	6.09	2.47	11	low
LTC NS 8 - Pulmonary embolism post-surgery	7	5	1.80	6.18	2.40	11	low
LTC NY 13 - Newly documented deep vein thrombosis (DVT)	7	5	1.28	6.33	1.61	12	low
LTC PSI 13 - Postoperative sepsis	7	5	1.60	6.36	2.20	11	low
LTC NS 25 - Systemic Sepsis (SIRS) post-surgery	7	5	1.64	6.00	2.14	11	low
LTC NS 26 - Systemic Sepsis (Sepsis) post-surgery	7	5	1.80	6.18	2.32	11	low
LTC NS 27 - Systemic Sepsis (Septic Shock) post-surgery	7	5	1.80	6.18	2.32	11	low
LTC UT 3 - Septicemia, bacteremia (surgical)	7	5	1.68	6.42	1.93	12	low
LTC MP 8 - Hospital-acquired bloodstream infections (BSIs)	7	5	1.64	6.30	2.21	10	low
LTC NH 13 - Recently hospitalized residents with pressure ulcers	7	5	1.88	6.25	2.34	12	low
LTC NQF 19 - Patient harm from electric shock	7	5	1.29	6.75	1.54	12	low
LTC NQF 6 - Contaminated product causing patient harm	7	5	1.42	6.50	1.62	12	low
LTC CP 29 - Re-hospitalization secondary to CHF within 30 days of CHF hospitalization	7	5	1.43	6.38	1.80	13	low
LTC NQF 23 - Patient harm from use of restraints	7	5	1.33	7.00	1.71	12	low
LTC NY 26 - Termination of any services vital to continued safe operation	7	5	1.92	6.69	2.50	13	low
LTC NY 5 - Cardiac and/or respiratory arrest requiring ACLS	7	4	2.21	6.23	2.59	13	low
LTC UT 32 - Respiratory arrest	7	4	2.08	6.08	2.39	12	low
LTC HHC 10 - Emergent care for wound infections, deteriorating wound status	7	4	1.89	6.38	2.36	13	low
LTC NY 4 - Unexpected deaths	7	4	1.64	6.54	2.07	13	low
LTC NQF 15 - Patient harm from hypoglycemia	7	4	2.00	6.17	2.48	12	low
LTC HHC 12 - Emergent care for hypo/hyperglycemia	7	4	1.81	6.17	2.37	12	low
LTC CP 6 - Choking incidents involving consumption of inappropriate food	7	4	2.05	6.46	2.63	13	low
LTC CP 3 - Patient use of improper medication for age group	7	4	1.83	6.15	2.23	13	low
LTC UT 9 - Drug psychoses	7	3	2.38	5.69	2.81	13	low
LTC MP 15 - Adverse drug events associated with hypoglycemic agents	7	3	2.00	5.75	2.38	12	low
LTC NS 19 - Cardiac arrest requiring CPR post surgery	7	2	2.40	5.92	2.84	13	low
LTC NH 9 - High-risk residents with pressure ulcers	6.5	4	1.75	6.08	2.27	12	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
LTC NY 19 - Unplanned specific patient transfers from diagnostic and treatment centers to hospital	6.5	4	1.61	5.92	1.93	12	low
LTC NSPM 4 - Falls with injury	6.5	4	1.78	5.92	2.27	12	low
LTC NS 29 - Return to the OR within 30 days of surgery	6.5	3	1.94	5.67	2.42	12	low
LTC UT 31 - Sudden death	6.5	3	2.39	5.83	2.86	12	low
LTC MP 16 - Adverse drug events associated with digoxin	6.5	3	1.69	5.83	2.04	12	low
LTC NQF 10 - Patient elopement	6	6	1.61	6.92	1.89	13	low
LTC PSI 12 - Postoperative pulmonary embolism or deep vein thrombosis (DVT)	6	5	1.72	5.67	2.27	12	low
LTC NY 17 - Post-operative surgical wound infection	6	5	1.92	5.91	2.47	11	low
LTC NS 4 - Surgical wound disruption	6	5	1.76	6.08	2.35	12	low
LTC NS 15 - CVA/Stroke post-surgery	6	5	1.74	5.91	2.34	11	low
LTC PSI 15 - Accidental puncture or laceration, Secondary diagnosis field (surgical)	6	5	1.64	6.00	2.24	11	low
LTC NH 12 - Recently hospitalized residents who experience moderate to severe pain	6	5	1.50	5.69	2.14	13	low
LTC NS 2 - Deep incisional surgical site infection (SSI)	6	4.5	1.64	5.80	2.25	10	low
LTC NS 6 - Pneumonia Post-Surgery	6	4	1.73	5.85	2.19	13	low
LTC UT 4 - Pneumonia	6	4	1.86	5.83	2.41	12	low
LTC UT 12 - Pulmonary embolism and infarction (surgical)	6	4	2.00	5.80	2.44	10	low
LTC MP 1 - Postoperative venous thromboembolic events	6	4	1.92	5.42	2.43	12	low
LTC NS 24 - DVT/Thrombophlebitis post-surgery	6	4	1.92	5.42	2.43	12	low
LTC PSI 14 - Postoperative wound dehiscence, Secondary diagnosis field	6	4	1.60	5.50	2.17	10	low
LTC CP 47 - Polypharmacy in the geriatic population	6	4	1.69	6.17	2.04	12	low
LTC MP 17 - Hospital-acquired antibiotic associated clostridiun difficile (C. Diff)	6	4	1.83	5.82	2.23	11	low
LTC CP 33 - Re-hospitalization secondary to PE within 4 weeks of any hospitalization	6	4	1.50	6.00	1.81	12	low
LTC CP 32 - Re-hospitalization secondary to DVT within 4 weeks of any hospitalization	6	4	1.85	6.00	2.12	13	low
LTC NSPM 3 - Patient falls	6	4	1.67	5.64	2.06	11	low
LTC CP 35 - Re-hospitalization secondary to renal failure within 2 weeks of any hospitalization	6	3.5	1.48	5.60	1.84	10	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
LTC MP 2 - Postoperative pneumonia events	6	3	1.82	5.77	2.28	13	low
LTC NH 1 - Residents who lost too much weight	6	3	1.82	5.58	2.19	12	low
LTC UT 41 - Poisoning by psychotropic agents	6	3	2.38	5.08	2.69	13	low
LTC CP 30 - Re-hospitalization secondary to sepsis within 2 weeks of any hospitalization	6	3	2.18	6.00	2.57	11	low
LTC CP 36 - Re-hospitalization secondary to pneumonia within 2 weeks of any hospitalization	6	3	1.66	5.46	2.03	13	low
LTC UT 52 - Accidental falls	6	3	2.08	4.92	2.33	13	low
LTC NQF 22 - Patient falls in a health care facility	6	3	1.88	5.85	2.27	13	low
LTC NY 16 - Patient falls	6	3	1.75	5.50	2.07	12	low
LTC NSPM 5 - Restraint prevalence (vest and limb)	6	3	2.26	5.62	2.79	13	low
LTC NS 20 - Myocardial infarction intra/post-surgery	6	2	2.18	5.42	2.68	12	low
LTC MP 10 - Adverse events associated with hip joint replacement due to fracture	6	1.5	2.50	5.10	2.96	10	low
LTC NS 23 - Surgical Graft/Prothesis/Flap failure	5.5	3.5	1.80	5.40	2.27	10	low
LTC CP 28 - Re-hospitalization secondary to COPD within 30 days of COPD hospitalization	5.5	3	1.67	5.33	2.02	12	low
LTC PSI 9 - Postoperative hemorrhage or hematoma	5	4	1.59	5.45	2.11	11	low
LTC CP 34 - Re-hospitalization secondary to cellulitis within 2 weeks of any hospitalization	5	4	1.09	5.00	1.67	11	low
LTC CP 37 - Re-hospitalization secondary to malnutrition within 4 weeks of any hospitalization	5	4	1.60	5.50	1.96	10	low
LTC CP 31 - Re-hospitalization secondary to a UTI within 2 weeks of any hospitalization	5	4	1.39	5.17	1.95	12	low
LTC NS 1 - Superficial incisional surgical site infection (SSI)	5	3.5	1.54	5.10	2.28	10	low
LTC MP 6 - Postoperative urinary tract infections (UTIs)	5	3	2.09	5.23	2.52	13	low
LTC NS 13 - Urinary tract infection post-surgery	5	3	2.09	5.23	2.52	13	low
LTC NH 7 - Residents with a urinary tract infection	5	3	1.74	5.62	2.18	13	low
LTC UT 33 - Poisoning by antibiotics and other antiinfectives	5	3	2.01	4.85	2.48	13	low
LTC UT 34 - Poisoning by hormones and synthetic substitutes	5	3	1.92	4.91	2.39	11	low
LTC CP 38 - Re-hospitalization secondary to medication non-compliance within 4 weeks of any hospitalization	5	3	2.02	5.31	2.43	13	low
LTC MP 9 - Adverse events associated with hip joint replacement due to degenerative conditions	5	2	2.12	4.91	2.63	11	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
LTC MP 11 - Adverse events associated with knee replacements	5	2	2.28	4.91	2.77	11	low
LTC UT 26 - Alterations in mental state	5	2	2.07	4.69	2.50	13	low
LTC UT 61 - Adverse effects of psychotropic agents	5	2	2.77	5.00	3.16	13	low
LTC MP 12 - Adverse drug events associated with the anticoagulant warfarin	5	2	2.17	4.73	2.61	11	low
LTC MP 13 - Adverse drug events associated with the anticoagulant IV heparin	5	2	2.36	4.83	2.82	12	low
LTC MP 14 - Adverse drug events associated with the anticoagulant low molecular weight heparin or factor Xa	5	2	2.26	4.92	2.78	12	low
LTC UT 5 - Other infections (surgical)	5	1	2.42	4.22	2.68	9	low
LTC NS 5 - Other surgical wound occurrence	5	1	2.22	4.00	2.50	9	low
LTC NY 25 - External disaster outside the control of the hospital which affects facility operation	5	1	2.65	4.85	3.08	13	low
LTC UT 37 - Poisoning by analgesics, antipyretics, antirheumatics	4.5	2	1.92	4.42	2.43	12	low
LTC UT 38 - Poisoning by anticonvulsant and anti-Parkinsonian drugs	4.5	2	2.00	4.50	2.50	12	low
LTC UT 36 - Poisoning by agents primarily affecting blood constituents	4	3	1.75	4.27	2.24	11	low
LTC UT 39 - Poisoning by sedatives and hypnotics	4	3	2.06	4.77	2.52	13	low
LTC UT 40 - Poisoning by other CNS depressants, stimulants, anesthetics, nervous system agents	4	3	2.37	4.77	2.68	13	low
LTC NY 14 - Acute mycocardial infarction (AMI) unrelated to a cardiac procedure	4	2	2.33	4.50	2.75	12	low
LTC UT 19 - Disorders of urinary system	4	2	2.17	4.17	2.48	12	low
LTC CP 25 - Hospitalization secondary to seizures in patients with known seizure disorder	4	2	1.50	3.50	1.69	8	low
LTC UT 29 - Shock	4	2	2.44	4.67	2.90	12	low
LTC UT 53 - Adverse effect of antibiotics and other antiinfectives	4	2	2.36	4.62	2.66	13	low
LTC UT 35 - Poisoning by primarily systemic agents	4	2	2.40	4.36	2.77	11	low
LTC UT 54 - Adverse effect of hormones and synthetic substitutes	3.5	2	2.26	4.08	2.57	12	low
LTC UT 57 - Adverse effects of analgesics, antipyretics, antirheumatics	3.5	2	2.36	4.17	2.72	12	low
LTC UT 58 - Adverse effects of anticonvulsant and anti-Parkinsonian drugs	3.5	2	2.56	4.33	2.84	12	low

Measure	Median	20th Percentile	Mean Abs Dev	Mean	StdDev	N	Rating
LTC UT 62 - Adverse effect of agents primarily affecting the cardiovascular system	3.5	2	2.58	4.50	2.88	12	low
LTC UT 55 - Adverse effects of primarily systemic agents	3	2	2.76	4.36	3.04	11	low
LTC UT 56 - Adverse effect of agents primarily affecting blood constituents	3	2	2.48	4.09	2.81	11	low
LTC UT 59 - Adverse effects of sedatives and hypnotics	3	2	2.75	4.69	3.01	13	low
LTC UT 60 - Adverse effect of other CNS depressants, stimulants, anesthetics, nervous system agents	3	2	2.89	4.54	3.10	13	low
LTC UT 42 - Poisoning by other agents	3	2	2.48	4.27	2.80	11	low
LTC UT 63 - Adverse effect of other drugs, biological, medicinal substances in therapeutic use	3	2	2.46	4.09	2.77	11	low
LTC UT 27 - Rash, spontaneous ecchymoses	3	1	1.88	3.46	2.18	13	low
LTC UT 22 - Dermatitis	2	2	1.99	3.42	2.31	12	low
LTC UT 18 - Nausea, vomiting, diarrhea	2	1	2.45	3.64	2.91	11	low
LTC UT 10 - Disorders of the nervous system	1	1	1.98	2.78	2.49	9	low

APPENDIX E.
PARTICIPANT COMMENTS FROM THE IMPORTANCE AND VALIDITY RATING PROCESSES

IMPORTANCE

Several Delphi participants provided comments on the rating process, and on some of the candidate measures being considered in the process. Similar to the previous comments offered by the Delphi participants as the candidate measures were initially being compiled, these comments raise issues that should provide context for determining: (1) what additional work is needed before measures identified as important are ready for widespread use, and (2) what guidance should be provided for appropriate use of the measures.

These issues will also be discussed during the conference call planned for the end of the Delphi process. At that time, participants will work as a group to consider the measures that surface from the Delphi process, assessing their limitations and caveats, and providing guidance on how to present the measures in a final product from the process.

General Comments from Participants:

- "These seem to cover most of what is important to capture. I did not have to eliminate many. Some I felt to be important, but as they were so specific, they would not occur in appreciable amounts to track as an indicator (e.g.- assault on a patient rated as more important than rape of patients)."
- "There were lots of duplicates. Some more specific than others. I assume at some point we will have a chance to pick the best of the similar measures."
- "With the exception of neonatal ICU patients, there are very few safety measures that are applicable to pediatric patients."
- "There are several items on these lists for which I have checked a very low rating in terms of measurement. Most often, that is because these types of occurrences should be investigated as individual instances and the lessons learned about risks and safety improvement derived from those analyses, not that there should be a tabulation of the events."

Comments on Difficulty in Rating Importance. *Several participants found the rating process difficult, for a variety of reasons. Participants identified the following reasons for difficulty in rating:*

- "Many of the measures (especially those nominated by a participant and those UT) were inadequately documented. In many cases, it was unclear how to weigh very low rates of error (or of the situation itself) - such a consideration would be important in any real application. Likewise, less egregious errors might be very important because so many are affected. I'm sure mine shows more than the usual incoherence."
- "Very difficult to limit evaluation to importance domain; feasibility and validity likely influenced my responses."

- "The largest problem with this part is the vague definition of some of these measures. In particular, there is a heavy reliance on ICD-9-CM coding. Some of the codes are very non-descript and many patients may get these codes upon discharge, the reason for them getting the code having nothing to do with a patient safety issue."
- "Some issues were difficult to rate because of ambiguity. Hard to be really precise."
- "Some of the measures were so poorly worded or so vague that it was not possible to rate them....for these, I listed the lowest number."

Comments on Specific Measures:

- "Throughout the document, the terms 'emergency department' or 'ED' should be used rather than 'emergency room' or 'ER' since these are the industry standard (the use of the latter terms can make the measures be perceived as outdated by some)."
- "NS 22—needs to have exclusions for emergent surgical issues that involve blood loss, such as trauma or GI bleed."
- "CP 52—doesn't take into account patients who are determined (during a PMD visit) to need hospitalization. Provides a negative incentive for MDs to admit patient after visit."
- "CP 53—Follow-up visits after initial PCP visit, ED visit, or urgent care visits are considered the "safety net" and are a very important part of proper care. In fact, primary care physicians who are not accessible for urgent follow-up visits pose more of a quality problem. This measure would cause a negative incentive for PCPs to provide timely follow-up."
- "CP3—needs better definition. Has the potential to be misinterpreted. For example, benadryl is considered inappropriate for regular or chronic use in the elderly, but is appropriate for one time use when an elderly patient is having an allergic reaction."
- "RD4—I am concerned that this measure can have the unintended consequence of providing a negative incentive to admit patients who are having exacerbations of these conditions and require admission for appropriate care."
- "Question 14 on surgical PSI 19 is only helpful for the category of pregnant women who are multiparous and without macrosomia or assisted delivery."
- "MP14 Xa determinant of anticoagulation level rather than RX?"

Comment on Transferring Measures to Ambulatory Care and Long Term Care Settings:

"Thank you for the opportunity to participate in the RAND Patient Safety Outcomes Measures Delphi project. The project's goal of selecting the most important patient safety outcome measures from existing measurement sets is an important undertaking. However, we are greatly concerned about the arbitrary rollover of measures to the ambulatory and long term care settings. This strategy was illogical at best and nonsensical in many cases.

It was difficult for us (and I presume other raters) to interpret the measures when the context is subject to interpretation. For example, if the ambulatory care setting includes outpatient surgery centers, then the inclusion of surgical outcomes measures may be appropriate. However, the majority of ambulatory care is provided at medical

offices for which surgical outcomes are not relevant. In other cases, its not clear that the measures can even be applied at all. For example, how were we to interpret post surgical outcomes in the long term care setting?

We believe this lack of measurement context will render whatever information you receive on the ambulatory and long term care measures borrowed from the inpatient setting very problematic."

VALIDITY

Summary of Comments Made

All of the comments provided by participants at each step of the Delphi process are being pulled together for inclusion in the report that presents the results of this process. We list below the general comments made during the validity rating step of the process. Additional, more specific comments were provided on the individual measures that had been rated as important, and therefore, were addressed in the validity rating step; they are not included here but will be listed in the report.

The general comments made by participants are organized into topic areas of quality of validity information available to consider in the rating process, other measurement issues, issues specific to source groups of measures, and use of measures for ambulatory care and long-term care settings.

Quality of Information on Validity

I found this quite difficult to do and tended to not have a lot of variation in my answers

Many of the measures are duplicative.

I have concern about the validity of the results of this Delphi survey for several reasons. The enormity of the task was such that even with the information that you provided, I suspect that most of us had to give answers based on "gestalt". Ideally, one would like to go to the actual literature about these issues and see how valid similar types of measures were. Also, most measures had little validation, so we were really predicting, not judging. Also, at least at this stage, we had little information on how cases would be collected, ie. chart review? (and if chart review, by whom? trained abstractors vs the hospital nurse who is on light duty due to a back injury vs other). Or if not chart review, by what reporting mechanism, with what sanctions if reporting is inaccurate. All these will affect validity. I have grave concerns, if this survey is used for policy decisions. This is not a criticism of the methodology of the survey, it is just that I think that there is not enough information available to determine how these measures would fare in the real world.

Second, it appears that most of the measures have no data regarding their validity. In my opinion, I do not think experts (like myself and your panelists) can judge validity - it is an empirical issue to be tested empirically. It is not a concept that should be measured using opinion.

Comparison of validity of these complex groups of measures, collected using various methodologies, is a difficult task. I would in general suggest that the better-validated measures based on collection of administrative data or voluntary reporting are likely to be valid insofar as the data were reported or coded, but even if "valid" in this sense, they are not so useful if they're not consistently coded or reported. Hence, they may be insensitive measures of the conditions of concern. They may be of variable validity across institutions that have variable reporting behaviors. Chart review that uses standardized review criteria tends to be more sensitive and systematically sound (as long as extractors are collecting data in standard ways), and the concept of "validity" is more focused on interrater agreement about the categorization of events - this, however, may vary highly depending on how exactly the reviews were done, and the inter-rater reliability in one study may or may not apply to the next. To compare the use of each of these

measures, across variable populations and variable hospitals is difficult at best. I am not sure how well I can comparatively rate the validity of many of these measures on short notice (with "validity" having different meanings in different contexts), but here are my best impressions based on the data available!

I wish there were more information regarding the validity and reliability of these measures.

We agree with your working definition of validity, which is an assessment of whether or not the candidate measure is an accurate measurement of the intended care process or outcome. As you know, this is separate from the ability to repeated apply the measure and get the same (whether accurate or inaccurate) result. While you have identified (and provided) some data on inter-rater reliability, you did not provide information on validity. This may be because of the sparseness of such data or the difficulty in locating it through standard retrieval methods.

Hard to keep disentangling validity from other important concepts like sensitivity in use.

Other Measurement Issues

I'm having a hard time completing the survey. …For evidence of validity, the documents often report data on the reliability of the measure. Reliability and validity are two distinct concepts. So why tell us about reliability when you want us to assess validity?

I still think it will be confusing to the panel to present reliability data when you are asking them to judge validity. As you know, a measure can be reliable but not valid, or valid but not reliable. If the panel is incorporating information about reliability and validity when making these judgments, then I think it would be more accurate to say these are judgments about the utility or overall potential usefulness, than validity.

I was frustrated by the lack of details regarding the denominators and/or risk adjustment.

The biggest problem with almost all of these measures is that there is no denominator.

Most of the measures have undefined denominators, and it seems that we should better define them in this process. Overall, this makes them less valid as measures for comparisons, since different institutions might interpret differently.

A number of these are important but are not 'measures' - they are 'never' events, absolute numbers with a target/threshold of zero. They are only valid as sentinel events. Others are important (e.g. medication errors) but more work is needed to achieve necessary specificity and reliable/consistent reporting and classification. Overall a slightly discouraging exercise.

Issues Specific to Source Sets of Inpatient Measures

STS. Several of the numerator definitions include a clause that allows exemption of cases if the cause of death is clearly unrelated to the operation. This clause needs specifications so it can be applied in a consistent and standard manner by different users in order to have validity as a measure.

SCIP. None of the SCIP outcome measures have been specified (or at least the specifications have not yet been publicly shared) and so should be listed as not yet ready since it is impossible to evaluate the validity without specifications.

NQF. The concepts represented by the NQF Serious Reportable Adverse Events are all important events. However, as they are currently written they cannot be operationalized for measurement. Based on the limited definitions provided, they cannot be implemented in a standardized manner. They need complete specification before they will be useful as measures.

NSPM. NSPM are well defined but the validity could be severely impaired depending on how they are implemented. Specifications not detailed enough for standardization.

AHRQ. The AHRQ Quality [patient safety] Indicators were designed not as standardized measures but as indicators and meant to serve as surveillance screening tools. They have not been systematically validated for the purpose of measurement. In supporting documentation, AHRQ provides a list of potential concerns and recommends that those who intend such use of the indicators perform their own evaluation of the validity.

NYPORTS. None of these candidates have denominators and so are event counts and not rates. They are intended to serve as adverse event reporting (or surveillance) monitoring. They are voluntary and based on expert consensus. Without specification of denominator, they cannot be converted to rates or adjusted and are therefore not comparable across institutions so can't be used as measures.

Use of Measures in Ambulatory Care and Long-term Care Settings.

While we believe it is important to evaluate surgical care in settings other than hospitals and in ambulatory and long term care, in general, the candidate measures provided have not been tested for use in those settings and should undergo evaluation before being used in those settings.

I still worry about how these indicators will be used, since validity is related to candor and insensitivity of surveillance. and ultimately improvement is only facilitated in a supportive overall environment that, done wrong, these report could degrade validity.

APPENDIX F. SUMMARY STATISTICS FOR THE VALIDITY RATINGS FOR EACH MEASURE

Measure	Median	20th Percentile	Mean Abs Dev	Mean	Std Dev	N	Percentage of respondents who said this measure is not ready to be rated for validity
Surgical Inpatient							
NQF 13 - Administration of ABO-incompatible blood	7.5	5	1.71	7.00	1.96	28	.
STS 6 - Risk-adjusted operative mortality for CABG	7	5	1.18	6.55	1.50	29	.
MP 1 - Postoperative venous thromboembolic events	7	5	1.27	6.17	1.47	29	.
STS 2 - Deep sternal wound infection rate (for CABG surgical patients)	7	5	1.15	6.46	1.45	28	.
MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	7	5	1.19	6.66	1.47	29	.
NQF 1 - Surgery performed on the wrong body part	7	5	1.79	6.78	2.12	27	.
NQF 2 - Surgery performed on the wrong patient	7	5	1.76	6.82	2.04	28	.
NQF 3 - Wrong surgical procedure performed on a patient	7	5	1.60	6.89	1.89	27	.
NQF 5 - Intraoperative or immediately postoperative death in an ASA Class 1 patient	7	5	1.49	6.56	1.87	27	.
NSPM 7 - Central line-associated bloodstream infection (CLABSI) rate for intesive care unit locations	6.5	5	1.39	6.32	1.66	28	.
NY 1 - Wrong patient, wrong site surgical procedure	6.5	5	1.93	6.43	2.23	28	.
NS 20 - Myocardial infarction intra/post-surgery	6	4	1.53	5.59	1.84	29	.
NS 19 - Cardiac arrest requiring CPR post surgery	6	4	1.40	5.93	1.79	29	.
STS 7 - Risk-adjusted operative mortality for aortic valve replacement (AVR)	6	5	1.22	6.24	1.50	29	.
STS 8 - Risk-adjusted operative-mortality for mitral valve replacement/repair (MVR)	6	5	1.19	6.21	1.52	29	.
STS 9 - Risk-adjusted operative-mortality for MVR + CABG	6	5	1.30	6.07	1.65	28	.

Measure	Median	20th Percentile	Mean Abs Dev	Mean	Std Dev	N	Percentage of respondents who said this measure is not ready to be rated for validity
STS 10 - Risk-adjusted operative mortality for AVR + CABG	6	5	1.33	6.11	1.66	28	.
PSI 12 - Postoperative pulmonary embolism (PE) or deep vein thrombosis (DVT)	6	4	1.35	5.87	1.57	30	.
MP 4 - Insertion-site infections associated with c	6	5	1.33	6.23	1.59	30	.
NSPM 8 - Central line-associated bloodstream infections (CLABSI) rate for birthweight categories in neonatal intensive care unit locations	6	4	1.46	6.11	1.74	27	.
NS 3 - Organ/Space surgical site infection (SSI)	6	4	1.40	5.30	1.64	27	.
PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	6	4	1.85	5.90	2.24	29	.
NY 2 - Incorrect procedure or treatment - invasive	6	3	1.91	6.11	2.30	28	.
NQF 14 - Maternal harm in labor and delivery of low-risk pregnancy	5.5	4	1.54	5.46	1.79	28	.
NY 3 - Unintentionally retained foreign body	5.5	4	1.86	5.79	2.15	28	.
SC 5 - Patients diagnosed with post-operative ventilator-associated pneumonia (VAP) during index hospitalization	5	4	1.58	5.32	1.98	28	.
PSI 17 - Birth trauma - Injury to neonate	5	4	1.55	5.25	1.78	28	.
PSI 18 - Obstetric trauma - vaginal delivery with instrumentation	5	4	1.42	5.29	1.67	28	.
PSI 20 - Obstetric trauma - cesarean delivery	5	3	1.55	5.25	1.82	28	.
PSI 13 - Postoperative sepsis	5	3	1.70	5.27	2.08	30	.
NS 25 - Systemic Sepsis (SIRS) post-surgery	5	3	1.46	4.72	1.81	29	.
NS 26 - Systemic Sepsis (Sepsis) post-surgery	5	3	1.43	4.90	1.88	29	.
NQF 4 - Foreign object retention after surgery	5	4	1.84	5.78	2.19	27	.
SC - 3 Intra/post-operative pulmonary embolism (PE) diagnosed during index hospitalization and within 30 days of surgery	4	3	1.58	4.86	1.87	29	.
NS 27 - Systemic Sepsis (Septic Shock) post-surgery	4	3	1.44	4.89	1.72	27	.

Measure	Median	20th Percentile	Mean Abs Dev	Mean	Std Dev	N	Percentage of respondents who said this measure is not ready to be rated for validity
Surgical and Nonsurgical Inpatient							
NSPM 9 - Ventilator-associated pneumonia (VAP) rate for intensive care unit (ICU) locations	6	5	1.36	6.10	1.72	29	.
NQF 16 - Failure to treat hyperbilirubinimia	6	4	1.64	5.96	1.99	25	.
NSPM 10 - Ventilator-associated pneumonia (VAP) rate for birthweight categories in the neonatal intensive care unit (NICU) locations	5	4	1.44	5.72	1.77	29	.
NY 12 - New acute pulmonary embolism	5	4	1.49	5.26	1.79	27	.
NQF 8 - Air embolism	5	4	1.38	5.44	1.60	27	.
Nonsurgical Inpatient							
NQF 9 - Infant discharged to the wrong person	8	4	1.89	6.78	2.17	27	.
MP 8 - Hospital-acquired bloodstream infections (BSIs)	7	5	1.25	6.41	1.42	27	.
NY 30 - Infant abduction	7	3	2.13	6.14	2.46	28	.
NQF 25 - Abduction of a patient of any age	7	4	2.07	6.07	2.40	27	.
NY 31 - Infant discharged to the wrong family	6.5	4	2.11	6.39	2.35	28	.
NQF 17 - Severe pressure ulcers acquired after admission to a healthcare facility	6	4	1.38	6.00	1.72	26	.
NQF 19 - Patient harm from electric shock	6	3	1.79	5.36	2.10	25	.
NQF 12 - Death or serious disability due to medical error	6	4	1.75	5.89	2.14	27	.
NQF 23 - Patient harm from use of restraints	6	4	1.46	5.59	1.67	27	.
NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	6	3	1.80	5.48	2.17	27	.
NY 32 - Rape of a patient	6	3	1.83	5.67	2.18	27	.
NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	6	3	1.90	5.67	2.25	27	.
NY 11 - Medication error occurred that resulted in patient death	5.5	4	1.89	5.89	2.20	28	.

Measure	Median	20th Percentile	Mean Abs Dev	Mean	Std Dev	N	Percentage of respondents who said this measure is not ready to be rated for validity
NY 20 - Misadministration of radiation or radioactive material	5	4	1.40	5.35	1.70	26	.
NQF 7 - Patient harm from improperly used device	5	3	1.54	5.00	1.96	26	.
PSI 2 - Death in low mortality DRGs	5	2	2.09	5.15	2.49	26	.
PSI 4 - Failure to rescue	5	2	2.05	4.72	2.35	25	.
NQF 20 - Delivery of contaminated gas to a patient	5	3.5	1.60	5.20	1.98	25	.
NQF 21 - Patient harm from burns	5	3	1.84	5.56	2.14	27	.
NY 15 - Patient burns	5	3	1.86	5.41	2.14	27	.
NY 6 - Loss of limb or organ	5	3	1.98	5.30	2.38	27	.
NQF 11 - Patient suicide, or attempted suicide	5	3	1.78	5.58	2.06	26	.
NY 9 - Medication error occurred that resulted in permanent patient harm	5	4	1.80	5.67	2.08	27	.
NY 10 - Medication error occurred that resulted in near-death event	5	3	1.82	5.32	2.16	28	.
NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	5	3	2.09	5.15	2.43	26	.
NY 21 - Crime resulting in death or serious injury	5	3	2.15	5.33	2.45	27	.
NY 8 - Malfunction of equipment with harm	4.5	3	1.69	4.92	2.06	26	.
NY 4 - Unexpected deaths	4	3	1.80	4.59	2.15	27	.
CP 15 - Administration of enteral medications/solutions intravenously	3	1	4.15	4.00	3.20	27	44.44
CP 5 - Burns/scalding during bathing	2.5	1	4.42	3.81	3.16	26	46.15
CP 7 - Hours of overwhelming pain in terminal patients	1	1	5.82	2.56	2.62	27	62.96
Ambulatory Care							
****AMB NQF 1 - Surgery performed on the wrong body part	6	3	2.10	5.64	2.60	25	12.00
****AMB NQF 2 - Surgery performed on the wrong patient	6	3.5	2.22	5.80	2.69	25	12.00

Measure	Median	20th Percentile	Mean Abs Dev	Mean	Std Dev	N	Percentage of respondents who said this measure is not ready to be rated for validity
AMB NQF 12 - Death or serious disability due to medical error	6	2.5	2.16	5.28	2.57	25	12.00
AMB NY 11 - Medication error occurred that resulted in patient death	6	3.5	1.93	5.68	2.41	25	8.00
AMB PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	5.5	2	2.21	5.33	2.63	24	12.50
AMB NY 1 - Wrong patient, wrong site, surgical procedure	5.5	3	2.42	5.50	2.77	24	12.50
AMB NS 3 - Organ/space surgical site infection	5	2	2.24	4.65	2.35	23	13.04
****AMB NQF 4 - Foreign object retention after surgery	5	4	1.86	5.43	2.27	23	8.70
****AMB NQF 3 - Wrong surgical procedure performed on the patient	5	3	2.10	5.48	2.55	25	12.00
AMB NY 9 - Medication error occurred that resulted in permanent patient harm	5	3	1.78	5.16	2.30	25	8.00
AMB NY 10 - Medication error occurred that resulted in near-death event	5	2	2.24	4.68	2.50	25	12.00
AMB CP 14 - Failure to follow-up test results	1	1	5.13	3.08	2.75	25	56.00
Long Term Care							
LTC NQF 17 - Severe pressure ulcers acquired after admission to a health care facility	6	3	1.84	5.52	2.25	23	8.70
****LTC NH 10 - Average risk residents with pressure ulcers	6	2	2.02	5.39	2.48	23	13.04
LTC NY 32 - Rape of a patient	5.5	3	2.35	4.83	2.48	24	12.50
LTC MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	5	1	2.30	5.00	2.61	23	21.74
LTC NQF 13 - Administration of ABO-incompatible blood	5	1	2.76	4.70	2.84	23	26.09
LTC UT 23 - Decubitus ulcer	5	1	2.27	4.52	2.21	23	21.74
LTC NQF 12 - Death or serious disability due to	5	2	2.21	4.92	2.47	24	16.67

Measure	Median	20th Percentile	Mean Abs Dev	Mean	Std Dev	N	Percentage of respondents who said this measure is not ready to be rated for validity
medical error							
LTC NY 11 - Medication error occurred that resulted in patient death	5	3	2.42	5.09	2.71	23	17.39
LTC NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	5	3	2.38	4.88	2.52	24	12.50
LTC NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	5	3	2.43	5.00	2.73	23	17.39
LTC NQF 21 - Patient harm from burns	4.5	2	2.41	4.50	2.41	22	18.18
LTC NY 9 - Medication error occurred that resulted in permanent patient harm	4.5	2	2.28	4.67	2.43	24	16.67
LTC MP 4 - Insertion-site infections associated with central vascular catheters (CVCs)	4	1	2.45	4.52	2.45	23	21.74
LTC NQF 7 - Patient harm from improperly used device	4	1	2.05	3.87	1.96	23	17.39
LTC NY 8 - Malfunction of equipment with harm	4	2	2.38	3.91	2.19	23	17.39
LTC NQF 11 - Patient suicide, or attempted suicide	4	2	2.67	4.52	2.59	23	17.39
LTC NY 10 - Medication error occurred that resulted in near-death event	4	2	2.21	4.26	2.36	23	13.04
****LTC HHC 11 - Emergent care for improper medication administration, medication side effects	4	1	3.16	3.96	2.70	23	30.43
LTC NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	4	1	2.53	4.39	2.59	23	17.39
LTC NY 21 - Crime resulting in death or serious injury	4	3	2.52	4.88	2.71	24	12.50
LTC NY 7 - Impairment of limb, organ, or body function	3	1	2.74	3.22	1.88	23	26.09
LTC CP 5 - Burns/scalding during bathing	2	1	4.55	3.54	2.93	24	50.00
LTC CP 7 - Hours of overwhelming pain in terminal patients	1	1	5.44	2.63	2.45	24	58.33

APPENDIX G.
LISTING OF IMPORTANT MEASURES WITH IMPORTANCE AND VALIDITY RATINGS

Measure	Median Importance Rating	20th Percentile Importance Rating	Median Validity Rating	20th Percentile Validity Rating	Percentage of Step 3 respondents who said this measure is not ready to be rated for validity
Surgical Inpatient					
NQF 1 - Surgery performed on the wrong body part	9	8	7	5	.
NQF 2 - Surgery performed on the wrong patient	9	8	7	5	.
NQF 3 - Wrong surgical procedure performed on a patient	9	8	7	5	.
NY 1 - Wrong patient, wrong site surgical procedure	9	7	6.5	5	.
NQF 13 - Administration of ABO-incompatible blood	9	6	7.5	5	.
NY 2 - Incorrect procedure or treatment - invasive	9	6	6	3	.
STS 6 - Risk-adjusted operative mortality for CABG	8	7	7	5	.
PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	8	7	6	4	.
NY 3 - Unintentionally retained foreign body	8	7	5.5	4	.
NS 26 - Systemic Sepsis (Sepsis) post-surgery	8	7	5	3	.
NQF 4 - Foreign object retention after surgery	8	7	5	4	.
NS 27 - Systemic Sepsis (Septic Shock) post-surgery	8	7	4	3	.
STS 2 - Deep sternal wound infection rate (for CABG surgical patients)	8	6.5	7	5	.
MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	8	6	7	5	.
NQF 5 - Intraoperative or immediately postoperative death in an ASA Class 1 patient	8	6	7	5	.
NSPM 7 - Central line-associated bloodstream infection (CLABSI) rate for intesive care unit locations	8	6	6.5	5	.
NS 20 - Myocardial infarction intra/post-surgery	8	6	6	4	.
NS 19 - Cardiac arrest requiring CPR post surgery	8	6	6	4	.

Measure	Median Importance Rating	20th Percentile Importance Rating	Median Validity Rating	20th Percentile Validity Rating	Percentage of Step 3 respondents who said this measure is not ready to be rated for validity
STS 7 - Risk-adjusted operative mortality for aortic valve replacement (AVR)	8	6	6	5	.
STS 8 - Risk-adjusted operative-mortality for mitral valve replacement/repair (MVR)	8	6	6	5	.
STS 9 - Risk-adjusted operative-mortality for MVR + CABG	8	6	6	5	.
STS 10 - Risk-adjusted operative mortality for AVR + CABG	8	6	6	5	.
PSI 12 - Postoperative pulmonary embolism (PE) or deep vein thrombosis (DVT)	8	6	6	4	.
NSPM 8 - Central line-associated bloodstream infections (CLABSI) rate for birthweight categories in neonatal intensive care unit locations	8	6	6	4	.
NS 3 - Organ/Space surgical site infection (SSI)	8	6	6	4	.
NQF 14 - Maternal harm in labor and delivery of low-risk pregnancy	8	6	5.5	4	.
SC 5 - Patients diagnosed with post-operative ventilator-associated pneumonia (VAP) during index hospitalization	8	6	5	4	.
PSI 17 - Birth trauma - Injury to neonate	8	6	5	4	.
PSI 20 - Obstetric trauma - cesarean delivery	8	6	5	3	.
PSI 13 - Postoperative sepsis	8	6	5	3	.
NS 25 - Systemic Sepsis (SIRS) post-surgery	8	6	5	3	.
MP 1 - Postoperative venous thromboembolic events	8	5.5	7	5	.
MP 4 - Insertion-site infections associated with c	8	5	6	5	.
PSI 18 - Obstetric trauma - vaginal delivery with instrumentation	8	5	5	4	.
SC - 3 Intra/post-operative pulmonary embolism (PE) diagnosed during index hospitalization and within 30 days of surgery	8	5	4	3	.

Surgical and Nonsurgical Inpatient

Measure	Median Importance Rating	20th Percentile Importance Rating	Median Validity Rating	20th Percentile Validity Rating	Percentage of Step 3 respondents who said this measure is not ready to be rated for validity
NSPM 9 - Ventilator-associated pneumonia (VAP) rate for intensive care unit (ICU) locations	8	6	6	5	.
NSPM 10 - Ventilator-associated pneumonia (VAP) rate for birthweight categories in the neonatal intensive care unit (NICU) locations	8	6	5	4	.
NY 12 - New acute pulmonary embolism	8	5.5	5	4	.
NQF 16 - Failure to treat hyperbilirubinimia	8	5	6	4	.
NQF 8 - Air embolism	8	5	5	4	.
Nonsurgical Inpatient					
NY 9 - Medication error occurred that resulted in permanent patient harm	9	8	5	4	.
NQF 9 - Infant discharged to the wrong person	9	7	8	4	.
NQF 12 - Death or serious disability due to medical error	9	7	6	4	.
NY 11 - Medication error occurred that resulted in patient death	9	7	5.5	4	.
NY 30 - Infant abduction	9	6.5	7	3	.
NY 31 - Infant discharged to the wrong family	9	6.5	6.5	4	.
NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	8.5	7	6	3	.
NY 32 - Rape of a patient	8.5	7	6	3	.
NQF 25 - Abduction of a patient of any age	8.5	6	7	4	.
NQF 17 - Severe pressure ulcers acquired after admission to a healthcare facility	8	7	6	4	.
NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	8	7	6	3	.
NY 10 - Medication error occurred that resulted in near-death event	8	7	5	3	.
NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or	8	7	5	3	.

Measure	Median Importance Rating	20th Percentile Importance Rating	Median Validity Rating	20th Percentile Validity Rating	Percentage of Step 3 respondents who said this measure is not ready to be rated for validity
other licensed health care provider					
MP 8 - Hospital-acquired bloodstream infections (BSIs)	8	6	7	5	.
NQF 19 - Patient harm from electric shock	8	6	6	3	.
NY 20 - Misadministration of radiation or radioactive material	8	6	5	4	.
PSI 2 - Death in low mortality DRGs	8	6	5	2	.
NQF 21 - Patient harm from burns	8	6	5	3	.
NY 15 - Patient burns	8	6	5	3	.
NY 8 - Malfunction of equipment with harm	8	6	4.5	3	.
NY 4 - Unexpected deaths	8	6	4	3	.
CP 15 - Administration of enteral medications/solutions intravenously	8	6	3	1	44.44
CP 7 - Hours of overwhelming pain in terminal patients	8	6	1	1	62.96
NQF 23 - Patient harm from use of restraints	8	5	6	4	.
PSI 4 - Failure to rescue	8	5	5	2	.
NQF 20 - Delivery of contaminated gas to a patient	8	5	5	3.5	.
NQF 11 - Patient suicide, or attempted suicide	8	5	5	3	.
NY 21 - Crime resulting in death or serious injury	8	5	5	3	.
CP 5 - Burns/scalding during bathing	8	5	2.5	1	46.15
NQF 7 - Patient harm from improperly used device	8	4	5	3	.
NY 6 - Loss of limb or organ	8	2	5	3	.
Ambulatory Care					
****AMB NQF 1 - Surgery performed on the wrong body part	9	8	6	3	12.00
AMB NQF 12 - Death or serious disability due to medical error	9	8	6	2.5	12.00
AMB NY 11 - Medication error occurred that resulted in patient death	9	8	6	3.5	8.00

Measure	Median Importance Rating	20th Percentile Importance Rating	Median Validity Rating	20th Percentile Validity Rating	Percentage of Step 3 respondents who said this measure is not ready to be rated for validity
****AMB NQF 3 - Wrong surgical procedure performed on the patient	9	8	5	3	12.00
****AMB NQF 2 - Surgery performed on the wrong patient	9	7	6	3.5	12.00
AMB NY 1 - Wrong patient, wrong site, surgical procedure	9	7	5.5	3	12.50
AMB NY 9 - Medication error occurred that resulted in permanent patient harm	8	8	5	3	8.00
AMB NY 10 - Medication error occurred that resulted in near-death event	8	8	5	2	12.00
AMB PSI 5 - Foreign body left in during procedure, Secondary diagnosis field	8	7	5.5	2	12.50
****AMB NQF 4 - Foreign object retention after surgery	8	7	5	4	8.70
AMB NS 3 - Organ/space surgical site infection	8	5	5	2	13.04
AMB CP 14 - Failure to follow-up test results	8	5	1	1	56.00
Long Term Care					
LTC NQF 12 - Death or serious disability due to medical error	9	8	5	2	16.67
LTC NY 11 - Medication error occurred that resulted in patient death	9	8	5	3	17.39
LTC NQF 27 - Death or significant injury of a patient or staff member resulting from a physical assault	9	8	5	3	17.39
LTC NY 9 - Medication error occurred that resulted in permanent patient harm	9	8	4.5	2	16.67
LTC NQF 17 - Severe pressure ulcers acquired after admission to a health care facility	9	7	6	3	8.70
LTC NY 32 - Rape of a patient	9	7	5.5	3	12.50
LTC NY 21 - Crime resulting in death or serious injury	9	6	4	3	12.50
LTC NY 10 - Medication error occurred that resulted in near-death event	8	8	4	2	13.04

Measure	Median Importance Rating	20th Percentile Importance Rating	Median Validity Rating	20th Percentile Validity Rating	Percentage of Step 3 respondents who said this measure is not ready to be rated for validity
LTC NQF 13 - Administration of ABO-incompatible blood	8	7	5	1	26.09
LTC NQF 26 - Sexual assault on a patient within or on the grounds of the health care facility	8	7	5	3	12.50
LTC NQF 24 - Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	8	7	4	1	17.39
LTC CP 5 - Burns/scalding during bathing	8	7	2	1	50.00
LTC NQF 21 - Patient harm from burns	8	6	4.5	2	18.18
LTC NQF 7 - Patient harm from improperly used device	8	6	4	1	17.39
LTC NY 8 - Malfunction of equipment with harm	8	6	4	2	17.39
LTC CP 7 - Hours of overwhelming pain in terminal patients	8	6	1	1	58.33
****LTC NH 10 - Average risk residents with pressure ulcers	8	5	6	2	13.04
LTC MP 5 - Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	8	5	5	1	21.74
LTC UT 23 - Decubitus ulcer	8	5	5	1	21.74
LTC NQF 11 - Patient suicide, or attempted suicide	8	5	4	2	17.39
****LTC HHC 11 - Emergent care for improper medication administration, medication side effects	8	5	4	1	30.43
LTC NY 7 - Impairment of limb, organ, or body function	8	5	3	1	26.09
LTC MP 4 - Insertion-site infections associated with central vascular catheters (CVCs)	8	3	4	1	21.74

APPENDIX H.
COMMENTS FROM PARTICIPANTS FOR EACH CANDIDATE MEASURE

Surgical Inpatient

Cardiac

#	Measure Name	Comments
NS 20	Myocardial infarction intra/post-surgery	**General Comments:** 1. Can we be sure that everyone has a post-surgical follow-up at 30 days? What is the source of this info? Might it be biased to be picking up problems only for patients who conscientiously keep a follow-up appt? Or suppose they go to a different hospital or MD to address a serous infection, rather than to a hospital responsible for the problem? Will that be captured and attributed? 2. Not sure of rationale as to why followed out to 30 days post op and not inclusive of other validating data (echocardiogram or stress test etc). **Validity Comments**: 1. it seems that there is some judgment about the diagnosis of an MI - so it does not seem valid enough to rely upon untested measures 2. Risk adjustment concern: does not appear to control for co-morbidity and risk of procedure 3. I am concerned about the lack of denominator specification for NS19 and NS20. 4. sensitivity of measure would depend on universally accepted definition of "myocardial infarction" (enzymes, EKG; transmural or also including only subendocardial/non q wave). False negatives would definitely occur due to failure to diagnose.
NS 19	Cardiac arrest requiring CPR post surgery	**General Comments:** 1. Can we be sure that everyone has a post-surgical follow-up at 30 days? What is the source of this info? Might it be biased to be picking up problems only for patients who conscientiously keep a follow-up appt? Or suppose they go to a different hospital or MD to address a serous infection, rather than to a hospital responsible for the problem? Will that be captured and attributed? 2. non q wave infarctions—definitions not supplied. **Validity Comments**: 1. Cardiac arrest is probably well-documented (though some facilities call severe ineffective circulation an "arrest" even with continued cardiac pumping - so this might be an overstatement) 2. Numerator is more sensitive than the definition because the latter includes PEA. 3. I am concerned about the lack of denominator specification for NS19 and NS20 4. Failure to capture cases where CPR had been administered (false negatives) would definitely occur; depends on context in which CPR performed ie if "crash cart" not opened there would not be a pharmacy

#	Measure Name	Comments
		charge in some institutions and patient would not be billed therefore no administrative record of CPR-ACLS; a number of other cases in which CPR might not be formally recorded and coded administratively
STS 6	Risk-adjusted operative mortality for CABG	**Validity Comments:** 1. relies upon the adequacy of the risk-adjustment - I know that there is a lot of attention on these, but I don't believe that studies have included adequate numbers of very old or multiply co-morbid patients to assure adequate risk adjustment - most has been done on <75 year olds. 2. devil in details of risk adjustment 3. I'm confused about how these are risk-adjusted. 4. My scores for the STS measures arise from the lack of information on the methods to "risk-adjust" I'm not sure what is being done and in the absence of more information weight my score conservatively.
STS 7	Risk-adjusted operative mortality for aortic valve replacement	**Validity Comments:** 1. relies upon the adequacy of the risk-adjustment 2.devil in details of risk adjustment 3. I'm confused about how these are risk-adjusted. 4. My scores for the STS measures arise from the lack of information on the methods to "risk-adjust". I'm not sure what is being done and in the absence of more information weight my score conservatively.
STS 8	Risk-adjusted operative mortality for mitral valve replacement	**Validity Comments:** 1. relies upon the adequacy of the risk-adjustment. 2. devil in details of risk adjustment 3. I'm confused about how these are risk-adjusted. 4. My scores for the STS measures arise from the lack of information on the methods to "risk-adjust". I'm not sure what is being done and in the absence of more information weight my score conservatively.
STS 9	Risk-adjusted operative mortality for MVR + CABG	**Validity Comments:** 1. relies upon the adequacy of the risk-adjustment. 2. devil in details of risk adjustment 3. I'm confused about how these are risk-adjusted. 4. My scores for the STS measures arise from the lack of information on the methods to "risk-adjust". I'm not sure what is being done and in the absence of more information weight my score conservatively. 5. The "n" for any given hospital for valve replacements with CABG will be too small to allow a narrow confidence interval for these rates
STS 10	Risk-adjusted operative mortality for AVR + CABG	**Validity Comments:** 1. relies upon the adequacy of the risk-adjustment 2. devil in details of risk adjustment. 3. I'm confused about how these are risk-adjusted. 4. My scores for the STS measures arise from the lack of information on the methods to "risk-adjust" I'm not

#	Measure Name	Comments
		sure what is being done and in the absence of more information weight my score conservatively. 5. The "n" for any given hospital for valve replacements with CABG will be too small to allow a narrow confidence interval for these rates

Pulmonary

#	Measure Name	Comments
SC 5	Patients diagnosed with post-operative ventilator-associated pneumonia (VAP) during index hospitalization	**Validity Comments:** 1. I'm a bit more comfortable making assumptions about the denominator in this case, but more specificity would be helpful to the measure. 2. There continues to be controversy regarding making the diagnosis of VAP. 3. Preventive interventions require large NNT 4. Great measure to look at but really need a denominator to make sense of this measure. 5. Rate highly dependent on surveillance intensity and criteria used to define 6. Pronovost's work and the IHI work has given substantial practical implementation of this measure. While diagnosis of a VAP still requires a little judgment, at least the definition has been worked out, the benchmarks, and the possibility of improvement. So, even with no direct evidence on point, I'd give it a "moderate"

PE/DVT

#	Measure Name	Comments
SC 3	Intra/post-operative pulmonary embolism diagnosed during index hospitalization and within 30 days of surgery	**Validity Comments:** 1. Subject to substantial challenges in including very high risk patients in uncertain numbers. The measures should probably be stratified, or use exclusions to get a group with similar risk. There is also a problem in inadequate and variable diagnosis - and in how to handle death within the period that is relevant to the measure. 2. Depends on surveillance details. 3. I'm concerned about the indicator not catching all of the events because of the nature of the event. 4. Input of clinical experts indicated re: sensitivity and specificity of diagnostic tests, especially ability to differentiate between preexisting and new disease 5. Very dependent on how hard DVT/PE is looked for. The most diligent providers, who aggressively screen/diagnose may look worse than the less diligent provider who never considers the possibility of these AEs.

#	Measure Name	Comments
PSI 12	Postoperative pulmonary embolism or deep vein thrombosis (DVT)	**Validity Comments:** 1. The PSI 12 has a thoughtful list of exclusions. 2. Depends on surveillance details 3. SC3- given the lack of details I needed to weight this one downward 4. Input of clinical experts indicated re: sensitivity and specificity of diagnostic tests, especially ability to differentiate between preexisting and new disease 5. Very dependent on how hard DVT/PE is looked for. The most diligent providers, who aggressively screen/diagnose may look worse than the less diligent provider who never considers the possibility of these AEs.
MP 1	Postoperative venous thromboembolic events	**Validity Comments:** 1. Subject to substantial challenges in including very high risk patients in uncertain numbers. The measures should probably be stratified, or use exclusions to get a group with similar risk. There is also a problem in inadequate and variable diagnosis - and in how to handle death within the period that is relevant to the measure. 2. Depends on surveillance details 3. I'm concerned about the indicator not catching all of the events because of the nature of the event. 4. Input of clinical experts indicated re: sensitivity and specificity of diagnostic tests, especially ability to differentiate between preexisting and new disease 5. Very dependent on how hard DVT/PE is looked for. The most diligent providers, who aggressively screen/diagnose may look worse than the less diligent provider who never considers the possibility of these AEs.

Labor and Delivery

#	Measure Name	Comments
NQF 14	Maternal harm in labor and delivery of low-risk pregnancy	**General Comments:** 1. Why do we specify “while being cared for in a health care facility”? Licensed nurse midwifes who perform home deliveries should be held to the same standard. Also, why do we exclude “death from pulmonary embolism” when it is otherwise considered a quality measure to provide appropriate prophylaxis from VTE to hospitalized patients? **Validity Comments:** 1. Measures would benefit from added specificity (e.g. definition of "maternal harm" in NQF 14, definition of "low risk pregnancy," definition of "injuries to neonate", definition of "obstetric trauma." 2. Risk adjustment concern 3. NQF 14 deals with bad enough events that there may be little importance to having a fuzzy definition -

#	Measure Name	Comments
		cardiomyopathy bad enough to diagnose is bad for a young woman! But it is not clear that cardiomyopathy is regularly preventable. The events noted are so rare that it seems that they are challenging for "rates," instead of case-by-case review.
PSI 17	Birth trauma - Injury to neonate	**Validity Comments:** 1. Measures would benefit from added specificity (e.g. definition of "maternal harm" in NQF 14, definition of "low risk pregnancy," definition of "injuries to neonate", definition of "obstetric trauma." 2. These are highly subject to coding differences, though I expect these differences will decrease should the measure be put into use. 3. Maybe obstetricians have strong definitions of these events, but they are not in our material. Is it a trauma to have a hematoma? It's not always wrong to have a trauma when that reduces risks of anoxic damage. I'm not sure that the exclusions in PSI 17 should not apply to PSI 18 and 19. 4. The events noted are so rare that it seems that they are challenging for "rates," instead of case-by-case review.
PSI 18	Obstetric trauma - vaginal delivery with instrumentation	**Validity Comments:** 1. Measures would benefit from added specificity (e.g. definition of "maternal harm" in NQF 14, definition of "low risk pregnancy," definition of "injuries to neonate", definition of "obstetric trauma." 2. These are highly subject to coding differences, though I expect these differences will decrease should the measure be put into use. 3. Maybe obstetricians have strong definitions of these events, but they are not in our material. Is it a trauma to have a hematoma? It's not always wrong to have a trauma when that reduces risks of anoxic damage. I'm not sure that the exclusions in PSI 17 should not apply to PSI 18 and 19. 4. The events noted are so rare that it seems that they are challenging for "rates," instead of case-by-case review.
PSI 20	Obstetric trauma - cesarean delivery	**Validity Comments:** 1. Measures would benefit from added specificity (e.g. definition of "maternal harm" in NQF 14, definition of "low risk pregnancy," definition of "injuries to neonate", definition of "obstetric trauma." 2. These are highly subject to coding differences, though I expect these differences will decrease should the measure be put into use. 3. Very small numbers 4. Maybe obstetricians have strong definitions of these events, but they are not in our material. Is it a trauma to have a hematoma? It's not always wrong to have a trauma when that reduces risks of anoxic damage. I'm not sure that the exclusions in PSI 17 should not apply to PSI 18 and 19. 5. The events noted are so rare that it seems that they are challenging for "rates," instead of case-by-case review.

Infections

#	Measure Name	Comments
STS 2	Deep sternal wound infection rate (for CABG surgical patients)	**Validity Comments:** 1. STS 2 is probably a serious enough complication to have good sensitivity - but there is a problem of classifying marginal wounds as "superficial."
MP 4	Insertion-site infections associated with central vascular catheters (CVC's)	**General Comments:** 1. Consider deleting. Not as clinically significant as catheter related BSI. **Validity Comments:** 1. MP4 is a measure that has had substantial practical experience in the IHI networks and thus probably has ways of being implemented with substantial validity. 2. In the absence of positive blood cultures, how would this be universally reliably diagnosed? What is an "insertion site" infection? Is this the same as a positive culture from the tip of a central venous catheter, or just a superficial or deep skin infection?
PSI 13	Postoperative sepsis	**Validity Comments:** 1. PSI 13 is at least replicable in the data, but there are all sorts of influences upon reporting this as a code - so it is probably thin on validity
NS 25	Systemic Sepsis (SIRS) post-surgery	**General Comments:** 1. Not useful measures if looking to improve care, more useful to look at the care issues that led the patient to be at risk, ie. surgical site infection. 2. SIRS is not defined. **Validity Comments:** 1. Relies upon a syndrome that is often masked in elderly and sick people and would be expected to have varying interpretation
NS 26	Systemic Sepsis (Sepsis) post-surgery	**General Comments:** 1. Not useful measures if looking to improve care, more useful to look at the care issues that led the patient to be at risk, ie. surgical site infection. 2. Why aren't blood cultures an important qualifier? **Validity Comments:** 1. NS 26 at least has clear criteria so that one would know what one had - those criteria have some controversy, but they are clear and thus valid in that sense. 2. Covariation of temperature and pulse confound the definition-- eg, a patient with viral pharyngitis and a fever will have a HR>90, but not sepsis

#	Measure Name	Comments
NS 27	Systemic Sepsis (Septic Shock) post-surgery	**General Comments:** 1. Not useful measures if looking to improve care, more useful to look at the care issues that led the patient to be at risk, ie. surgical site infection. 2. Why aren't blood cultures an important qualifier?
MP 5	Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	
NSPM 7	Central line-associated bloodstream infection (CLABSI) rate for intensive care unit locations	
NSPM 8	Central line-associated bloodstream infection (CLABSI) rate for birthweight categories in neonatal intensive care unit locations	
NS 3	Organ/Space surgical site infection (SSI)	

Procedural

#	Measure Name	Comments
PSI 5	Foreign body left in during procedure, Secondary diagnosis field	**Validity Comments:** 1. Clear enough - but rare event - may be valid and useless except in case review 2. Broad denominator allows med/surg ratio to skew data
NQF 4	Foreign object retention after surgery	**Validity Comments:** 1. don't really know how you are going to reliably and uniformly capture these adverse event rates (foreign body left in). ? voluntary reporting: highly influenced by local culture and in many ways best hospitals will have higher rates. 2. Title is "after surgery" - numerator is "after surgery or other procedure" - this complicates determining the denominator - but seems plausible to do - and valid - just low rate

#	Measure Name	Comments
		3. These are not really measures or rates, but "never" events
NY 3	Unintentionally retained foreign body	**Validity Comments:** 1. Same issues, less precise development all the rest share the same problems and potentials - they are probably easy to identify, hard to report, and rare. May require reporting system attention to encourage reports (and to encourage "near misses" reports) - but reports are probably valid. Will need work on denominators**.**
NY 1	Wrong Patient, wrong site surgical procedure	**General Comments:** 1.This is a good example of aggregating something that was disaggregated in the NQF measures-ie. wrong site and wrong patient are together here. 2. Too rare **Validity Comments**: 1. For the wrong patient/wrong site, the number may be relatively easy to determine, but the determination of a "rate" with comparisons between hospitals will be difficult due to the rarity of the events
NQF 1	Surgery performed on the wrong body part	**General Comments:** 1. Why does exigency mitigate concerns about wrong body part? 2. What if laterality on consent is incorrect? **Validity Comments:** 1. These are not really measures or rates, but "never" events
NQF 2	Surgery performed on the wrong patient	**Validity Comments:** 1. These are not really measures or rates, but "never" events
NQF 3	Wrong surgical procedure performed on a patient	**General Comments:** 1. Why does exigency mitigate concerns about wrong body part? **Validity Comments:** 1. These are not really measures or rates, but "never" events
NY 2	Incorrect procedure or treatment - invasive	

Other

#	Measure Name	Comments
NQF 5	Intraoperative or immediately postoperative death in an ASA Class 1 patient	**General Comments:** 1. Anesthesiologists often disagree as to whether a patient is ASA 1 or 2. ASA 2 patients are usually very healthy—why not include? **Validity Comments:**

#	Measure Name	Comments
		1. Need specified denominators! 2. These have the same problems as those before - rare, hard to report, easy to identify for those involved. Usually, more sensitivity to good practices is found in tracking rates of "near misses" if one can engineer a system to get them reported.
NQF 13	Administration of ABO-incompatible blood	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. blood accidents are highly indicative of errors but also will be rare if examining facility level data. 2. need specified denominators! 3. suspect some ABO mismatch errors will not be properly diagnosed or reported. 4. These have the same problems as those before - rare, hard to report, easy to identify for those involved. Usually, more sensitivity to good practices is found in tracking rates of "near misses" if one can engineer a system to get them reported.

Surgical and Non-surgical Inpatient

#	Measure Name	Comments
NSPM 9	Ventilator-associated pneumonia (VAP) rate for intensive care unit (ICU) locations	**Validity Comments:** 1. VAP is having substantial experience in the IHI work, so lots of conventions have been established - and the measure is sensitive to gains I'm not sure how that applies to neonates
NSPM 10	Ventilator-associated pneumonia (VAP) rate for birthweight categories in the neonatal intensive care unit (NICU) locations	
NY 12	New acute pulmonary embolism	**Validity Comments:** 1. New PE is often missed - but there is no better way to deal with it - so - might work well
NQF 8	Air embolism	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. Air embolism can be very difficult to diagnose, and easily confused with other shock states in complex patients. That said, sudden death in an otherwise stable ICU patient who has a cardiac arrest immediately after simple removal of a large bore catheter from the central circulation has likely suffered an air

#	Measure Name	Comments
		embolism - infrequent clear situation, much more clear than other air embolism scenarios - but just as preventable as wrong ABO blood unit transfused...
NQF 16	Failure to treat hyperbilirubinimia	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. Not risk adjusted

Non-surgical Inpatient

#	Measure Name	Comments
MP 8	Hospital-acquired bloodstream infections (BSIs)	**General Comments:** 1. Consider deleting. The most likely to be AEs are covered in MP 5.
NY 20	Misadministration of radiation or radioactive material	**Validity Comments:** 1. Depends on quality of surveillance measures (have to know it happened.) 2. Like the other rare errors - valid, but useful only as events (one hopes!)
NQF 7	Patient harm from improperly used device	**General Comments:** 1. How is serious disability defined? 2. Too imprecise of a definition to be useful. **Validity Comments:** 1. Vague 2. Too vague 3. NQF 7 and NY8 - judging malfunction and improper use invite interpretation and variation - should be validated in QI work at least
NY 8	Malfunction of equipment with harm	**Validity Comments:** 1. Will need denominator - otherwise valid 2. NQF 7 and NY8 - judging malfunction and improper use invite interpretation and variation - should be validated in QI work at least
NQF 17	Severe pressure ulcers acquired after admission to a health care facility	**Validity Comments:** 1. Denominator needs to be keyed to exposure - e.g., patients at risk per unit time - and probably has to be stratified - but will be valid except for PUs that are evolving during transfer (and that can take a few days to delineate)
NY 4	Unexpected deaths	**General Comments:** 1. How do we define “unexpected”? 2. Why is transfusion-related death excluded?

#	Measure Name	Comments
		Validity Comments: 1. Too vague 2. Having "Unexpected deaths" be valid will require having an excluded category of "people for whom death would not be a surprise" - or stratifying around this population - it is very different to have a person in otherwise good health to die unexpectedly than to have a person with an established serious illness who has a minor complication and dies quickly. In order to have this be a valid measure, there would need to be a way in which patients are prospectively categorized (at least as to long-term expected, or 'no surprise')
PSI 2	Death in low mortality DRGs	**Validity Comments:** 1. Doesn't control for heterogeneous severity 2. PSI 2 and 4 are likely to be valid. The testing shows a need for risk-adjustment - but the definitions and exclusions seem to handle most - the PSI 4 may have more problem with being intelligible - but it is probably fairly valid
PSI 4	Failure to rescue	**Validity Comments:** 1. PSI 2 and 4 are likely to be valid. The testing shows a need for risk-adjustment - but the definitions and exclusions seem to handle most - the PSI 4 may have more problem with being intelligible - but it is probably fairly valid
NQF 19	Patient harm from electric shock	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. Another rare, but valid, measure - unless obfuscated. Might again want to go for "near misses" not just actual harm.

NQF 20	Delivery of contaminated gas to a patient	**Validity Comments:** 1. Except for NY 6, these all have the same problem of being rare events at some risk of being hidden, and being more suitable for event monitoring than for rates (except as a part of a class of hazards and for "near miss" measures) - but they are probably valid overall
NQF 21	Patient harm from burns	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. Except for NY 6, these all have the same problem of being rare events at some risk of being hidden, and being more suitable for event monitoring than for rates (except as a part of a class of hazards and for "near miss" measures) - but they are probably valid overall
NY 15	Patient burns	**General Comments:** 1. Too rare.

		Validity Comments: 1. Except for NY 6, these all have the same problem of being rare events at some risk of being hidden, and being more suitable for event monitoring than for rates (except as a part of a class of hazards and for "near miss" measures) - but they are probably valid overall
NY 6	Loss of limb or organ	**Validity Comments:** 1. NY 6 has a different set of problems in that differentiating between losses that are intrinsic to the illness and those that arise from treatment or error seems quite challenging in many situations. Would take some experience to see how this worked and what definitions and clarifications might make it work well - but there is very little evidence.
NQF 11	Patient suicide, or attempted suicide	**Validity Comments:** 1. except for NY 6, these all have the same problem of being rare events at some risk of being hidden, and being more suitable for event monitoring than for rates (except as a part of a class of hazards and for "near miss" measures) - but they are probably valid overall

Medication

#	Measure Name	Comments
NQF 12	Death or serious disability due to medical error	**General Comments:** 1. How is serious disability defined? 2. Data collection issues. Many more events will be detected by hospitals using automated pharmacy systems and those simply looking harder. Hospital-to-hospital comparisons are therefore meaningless, as will be time trends as more hospitals adopt these systems. 3. You might consider including "omitted medication" in the list of examples. Many researchers have included omissions. For example, this might include death from infection after a failure to provide timely antibiotics. **Validity Comments:** 1. Exclusion criteria too vague
NY 9	Medication error occurred that resulted in permanent patient harm	**General Comments:** 1. Data collection issues. Many more events will be detected by hospitals using automated pharmacy systems and those simply looking harder. Hospital-to-hospital comparisons are therefore meaningless, as will be time trends as more hospitals adopt these systems. 2. I would consider combing NY 9, 10, 11 with NQF 12—I'm not sure what they add independently. **Validity Comments:** 1. Exclusion would be better stated "any non-preventable adverse drug reaction"
NY 10	Medication error occurred that resulted	**General Comments:** 1. Data collection issues. Many more events will be detected by hospitals using automated pharmacy

#	Measure Name	Comments
	in near-death event	systems and those simply looking harder. Hospital-to-hospital comparisons are therefore meaningless, as will be time trends as more hospitals adopt these systems. 2. Some exclusions could be "near death event" - which itself is vague. For example, a patient with blood sugar of 25 due to medication error involving insulin could have just had a seizure, or be in near shock with altered mental status etc. The other exclusion example, narcotic overdose, could be a patient with very low respiratory rate and high carbon dioxide - would require ACLS in another 10 minutes but was rescued. **Validity Comments:** 1. "near death event" is unacceptably vague
NY 11	Medication error occurred that resulted in patient death	**General Comments:** 1. Data collection issues. Many more events will be detected by hospitals using automated pharmacy systems and those simply looking harder. Hospital-to-hospital comparisons are therefore meaningless, as will be time trends as more hospitals adopt these systems.
All		**Validity Comments:** 1. For all of these, the obvious source of error is in the attribution of causation. It is hard enough to get reporting of medication errors and near misses - but then to attribute harm is often challenging because the patients were very ill anyway and the adverse effects are not usually atypical of the course. There should be more data on this, given all the work on medication errors. Are there algorithms and rules that are working reliably and that leave a rate that is sensitive to quality and to improvements? My guess is that there are some, and that would leave these a bit better off as to validity. 2. often difficult to definitively attribute adverse outcomes to adverse drug reaction

Physical Security

#	Measure Name	Comments
NQF 23	Patient harm from use of restraints	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. "harm" from use of restraints is vague - prefer more specification; assume false negatives would be a problem (institutional recognition and compliance issues unless case obvious, significant, widely known) 2. NQF 23 was a very important measure in nursing homes until recently, but now there are exceedingly few hazardous restraints in use in that setting. Thus, the measure itself may become, in that setting, too rare to be of help except in event monitoring. However, in hospitals, it is probably still a major issue. The attribution of causation here is usually fairly clear, so this is an unusually valid measure, even without much formal testing

#	Measure Name	Comments
NQF 9	Infant discharged to the wrong person	**General Comments:** 1. Consider defining infant. I believe that this measure is intended to be applied to infant discharges after birth, but this is not clear. It might cause confusion because infants have variable age definitions, (up to 3-6 months old) and are commonly seen in emergency departments and outpatient facilities after their initial birth hospitalization. Further it is not uncommon for infants to be brought to EDs and outpatient facilities by grandparents or other family members (or in the case of the ED, babysitters). In these cases discharging the infant with the (non-parent) person who brought them in is not an adverse event. **Validity Comments:** 1. The abduction and wrong discharge measures seem quite valid, even without testing, but rare.
NY 31	Infant discharged to the wrong family	**Validity Comments:** 1. The abduction and wrong discharge measures seem quite valid, even without testing, but rare.
NY 30	Infant abduction	**Validity Comments:** 1. The abduction and wrong discharge measures seem quite valid, even without testing, but rare.
NQF 25	Abduction of a patient of any age	**Validity Comments:** 1. The abduction and wrong discharge measures seem quite valid, even without testing, but rare.
NQF 24	Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	**Validity Comments:** 1. NQF 24-- measuring "instance of care" vs number of "individuals" impersonating will skew results since this is a rare occurrence.
All		**Validity Comments:** 1. My concerns about these are specifically with the very low frequency, making it impossible to compare "rates". Most will be 0. 2. All very rare- so if combined in one category would suggest lack of security. 3. No denominators specified.
NQF 26	Sexual assault on a patient within or on the grounds of the health care facility	
NY 32	Rape of a patient	
NQF 27	Death or significant injury of a patient or	

	staff member resulting from a physical assault	
NY 21	Crime resulting in death or serious injury	**Validity Comments:** 1. 'crime' too vague
All		**Validity Comments:** 1. Patients alleging the first three are often impaired by mental or cognitive challenges, so proof that an event occurred is often very difficult. Rape should be provable, provided there is any indication early on. But there are no data. So, my hunch is that these are challenging on validity ground. However, death or injury to a staff member and serious crime are usually well-investigated and probably as valid as anything civil society can offer (e.g., proof in the courts) 2. literature abounds with failure to detect some of the more egregious instances of these occurrences 3. Valid, reliable, but "never" events, not rates or measures. 4. again, all rare events- may need to be combined into a broader category

Concept Measures

#	Measure Name	Comments
CP 5	Burns/scalding during bathing	**Validity Comments:** 1. Burns and errors in giving enteral solutions in the wrong way are generally valid - but have the problems of under-reporting, rare events, and need for near-misses. 2. Burns and wrong-route errors are 'never' events.
CP 7	Hours of overwhelming pain in terminal patients	**Validity Comments:** 1. CP 7 - very sympathetic to this highly undesirable event, however, is vague, and has the characteristic of the kind of event that would be reported by a whistleblower - how would this be clinically and administratively recorded and reliably measured? 2. The CP-7 measure is very important but is not specified in the materials. There are measures of serious pain that are specified and for which population-based rates are reported (e.g., whether the patient has moderate to severe pain all or most of the time, in response to an interview). These data have been reviewed by the So Cal EPC recently. However, there is little evidence that good practices actually reduce the rates - experts believe it, but the rates have been surprisingly resistant to improvement. Also - "terminal" patients would need identified. There is some work on pain measurement being done by CMS and by NHPCO at present. Contact Melanie Merriman or Jean Kutner. 3. Pain is very important, a very good internal monitor, but probably not practical for use across sites yet
CP 15	Administration of enteral medications/solutions	**Validity Comments:** 1. Burns and errors in giving enteral solutions in the wrong way are generally valid - but have the problems of under-reporting, rare events, and need for near-misses.

#	Measure Name	Comments
	intravenously	2. Burns and wrong-route errors are 'never' events.
All		**Validity Comments:** 1. None of these are adequately defined

Ambulatory

Ambulatory-Surgical

#	Measure Name	Comments
AMB NS 3	Organ/space surgical site infection	**Validity Comments:** 1. problems will often be well after D/C, as well as low frequency
AMB PSI 5	Foreign body left in during procedure, Secondary diagnosis field	**Validity Comments:** 1. problems will often be well after D/C, as well as low frequency
AMB NQF 4	Foreign object retention after surgery	**Validity Comments:** 1. problems will often be well after D/C, as well as low frequency
AMB NY 1	Wrong patient, wrong site surgical procedure	**Validity Comments:** 1. problems will often be well after D/C, as well as low frequency
AMB NQF 1	Surgery performed on the wrong body part	
AMB NQF 2	Surgery performed on the wrong patient	
AMB NQF 3	Wrong surgical procedure performed on a patient	
All		**Validity Comments:** 1. These seem valid - just rare, with the usual problems. 2. again too rare- suggest combining into one measure wrong pt, site, surgery, etc.

Ambulatory- Non-Surgical

#	Measure Name	Comments
AMB NQF 12	Death or serious disability due to medical error	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. Med errors important but difficult to measure reliably across sites.

#	Measure Name	Comments
		2. Exclusions too broad 3. The medication errors raise the same problems as before - with challenges in finding the errors and getting them reported, and challenges in attribution. Probably will end up with "sure thing" and "possible" causation chain (here and in hospital list). For those with "sure thing" - validity probably good, but the boundary will be hard to regulate
AMB NY 9	Medication error occurred that resulted in permanent patient harm	**Validity Comments:** 1. Med errors important but difficult to measure reliably across sites. 2. The medication errors raise the same problems as before - with challenges in finding the errors and getting them reported, and challenges in attribution. Probably will end up with "sure thing" and "possible" causation chain (here and in hospital list). For those with "sure thing" - validity probably good, but the boundary will be hard to regulate
AMB NY 10	Medication error occurred that resulted in near-death event	**Validity Comments:** 1. "near death" is too vague 2. Med errors important but difficult to measure reliably across sites. 3. “near-death event" too vague 4. The medication errors raise the same problems as before - with challenges in finding the errors and getting them reported, and challenges in attribution. Probably will end up with "sure thing" and "possible" causation chain (here and in hospital list). For those with "sure thing" - validity probably good, but the boundary will be hard to regulate
AMB NY 11	Medication error occurred that resulted in patient death	**Validity Comments:** 1. Med errors important but difficult to measure reliably across sites. 2. The medication errors raise the same problems as before - with challenges in finding the errors and getting them reported, and challenges in attribution. Probably will end up with "sure thing" and "possible" causation chain (here and in hospital list). For those with "sure thing" - validity probably good, but the boundary will be hard to regulate
AMB CP 14	Failure to follow-up test results	**Validity Comments:** 1. AMB CP 14 - good but requires more specification; obviously "no followup" is ratable but much of the harm is "delayed followup;" would prefer time frame but that would have to be contextualized ie failure to follow up a + breast cancer biopsy for 4 months is very different from failure to follow up an elevated calcium value of 11.5 for three weeks in someone with minor clinical problems who develops additional issues. 2. The concept of failure to follow up is important but needs an endpoint such as death or other adverse outcome. 3. Follow up problems very important for more work. 4. Failure to follow up seems too unspecified to know - there is probably a good measure in this - with

#	Measure Name	Comments
		delayed follow-up being commonplace and trackable. But the measure would have to be specified to know what is meant.

Long Term Care

#	Measure Name	Comments
LTC MP4	Insertion-site infections associated with central vascular catheters (CVC's)	**General Comments:** 1. Not as clinically significant as catheter related BSI. **Validity Comments:** 1. While such things as CVCs and transfusions are plausible in some LTC settings, they are quite rare. On the other hand, devices are everywhere. There needs to be some experience with these in this setting before one could be sure to know what one is measuring. 2. Please define insertion-site infections more specifically
LTC MP 5	Bloodstream infections (BSIs) associated with central vascular catheters (CVCs)	
LTC NQF 7	Patient harm from improperly used device	**Validity Comments:** 1. "improperly used device" is vague
LTC NY 8	Malfunction of equipment with harm	
LTC NQF 13	Administration of ABO-incompatible blood	**General Comments:** 1. How is serious disability defined? 2. I was puzzled to see that administration of ABO incompatible blood was not rated as an important measure for surgical inpatients but that it was for long term care patients, where blood is not likely to be administered nearly as often as with surgical inpatients; concern for a preventable major transfusion error should be identical irrespective of setting and based on frequency of possible occurrence as well as severity.

LTC UT 23	Decubitus ulcer	**General Comments:** 1. Change wording to Pressure Ulcer (the recommended term by the National Pressure Ulcer Advisory Panel). **Validity Comments:** 1. On pressure ulcers - the new MDS work is getting better measures in hand - check with Deb Saliba at

		RAND. 2. The UT measure is foolish- misses a lot, fails to use the MDS. The NQF is better, but needs to be stratified. 3. Requires more specification
#	**Measure Name**	**Comments**
LTC NQF 17	Severe pressure ulcers acquired after admission to a health care facility	**Validity Comments:** 1. On pressure ulcers - the new MDS work is getting better measures in hand - check with Deb Saliba at RAND. 2. The UT measure is foolish- misses a lot, fails to use the MDS. The NQF is better, but needs to be stratified.
LTC NH 10	Average risk residents with pressure ulcers	**Validity Comments:** 1. On pressure ulcers - the new MDS work is getting better measures in hand - check with Deb Saliba at RAND. 2. The NH 10 measure includes stages 1 and 2 - but there is little evidence that those correlate with shortcomings in care.
LTC NQF 21	Patient harm from burns	**General Comments:** 1. How is serious disability defined? **Validity Comments:** 1. The burns and injury are just rare - event only
LTC NY 7	Impairment of limb, organ, or body function	**Validity Comments:** 1. The burns and injury are just rare - event only. 2. Too vague ie define "impairment"

LTC CP 5	Burns/scalding during bathing	
LTC CP 7	Hours of overwhelming pain in terminal patients	**Validity Comments:** 1. Highly sympathetic, however, is vague, and has the characteristic of the kind of event that would be reported by a whistleblower - how would this be clinically and administratively recorded and reliably measured? 2. severe pain that is untreated whether or not the person is terminal would be important
LTC NQF 11	Patient suicide, or attempted suicide	**Validity Comments:** 1."attempted suicide" too open to interpretation
LTC NQF	Death or serious disability due to	**General Comments:** 1. How is serious disability defined?

12	medical error	**Validity Comments:** 1."serious disability" requires specification 2. Exclusion too vague
LTC NY 9	Medication error occurred that resulted in permanent patient harm	**General Comments:** 1. I would consider combining NY 9, 10, 11 with NQF 12—I'm not sure what they add independently.
All		**Validity Comments:** 1. These have the same shortcomings as when they arose in other contexts

LTC NY 10	Medication error occurred that resulted in near-death event	**Validity Comments:** 1. "near death event" is unacceptably vague 2. "near death event" too vague
LTC NY 11	Medication error occurred that resulted in patient death	
LTC HHC 11	Emergent care for improper medication administration, medication side effects	**Validity Comments:** 1. LTC HHC 11 has promise; however "emergent care" needs to be defined and "improper" is open to wide interpretation 2. The emergent care will be more valid - but more rare.
LTC NQF 24	Any instance of care ordered or provided by someone impersonating a physician, nurse, pharmacist, or other licensed health care provider	**Validity Comments:** 1. NQF 24-- instance of care vs individual

LTC NQF 26	Sexual assault on a patient within or on the grounds of the health care facility	**Validity Comments**: 1. Patients alleging are often impaired by mental or cognitive challenges, so proof that an event occurred is often very difficult. Rape should be provable, provided there is any indication early on. But there are no data. So, my hunch is that these are challenging on validity ground.
LTC NY 32	Rape of a patient	**Validity Comments:** 1. Patients alleging are often impaired by mental or cognitive challenges, so proof that an event occurred is often very difficult. Rape should be provable, provided there is any indication early on. But there are

		no data. So, my hunch is that these are challenging on validity ground.
LTC NQF 27	Death or significant injury of a patient or staff member resulting from a physical assault	**Validity Comments**: 1. Death or injury to a staff member and serious crime are usually well-investigated and probably as valid as anything civil society can offer (e.g., proof in the courts)
LTC NY 21	Crime resulting in death or serious injury	**Validity Comments:** 1. Death or injury to a staff member and serious crime are usually well-investigated and probably as valid as anything civil society can offer (e.g., proof in the courts)