

Global Assessment of Pediatric Patient Safety (GAPPS) Trigger Tool

Section 1. Basic Measure Information

1.A. Measure Name

Global Assessment of Pediatric Patient Safety (GAPPS) Trigger Tool

1.B. Measure Number

0143

1.C. Measure Description

Please provide a non-technical description of the measure that conveys what it measures to a broad audience.

This measure calculates the rate of adverse events (patient harm)—defined as unintended physical injury resulting from or contributed to by medical care—per 1,000 patient-days and per 100 hospitalizations for patients less than 18 years of age.

This measure also calculates the rate of preventable adverse events per 1,000 patient-days and per 100 hospitalizations for patients less than 18 years of age.

1.D. Measure Owner

Center of Excellence for Pediatric Quality Measurement

1.E. National Quality Forum (NQF) ID (if applicable)

Not applicable.

1.F. Measure Hierarchy

Please note here if the measure is part of a measure hierarchy or is part of a measure group or composite measure. The following definitions are used by AHRQ:

- 1. Please identify the name of the collection of measures to which the measure belongs (if applicable). A collection is the highest possible level of the measure hierarchy. A collection may contain one or more sets, subsets, composites, and/or individual measures.**

Not applicable.

2. **Please identify the name of the measure set to which the measure belongs (if applicable). A set is the second level of the hierarchy. A set may include one or more subsets, composites, and/or individual measures.**

Not applicable.

3. **Please identify the name of the subset to which the measure belongs (if applicable). A subset is the third level of the hierarchy. A subset may include one or more composites, and/or individual measures.**

Not applicable.

4. **Please identify the name of the composite measure to which the measure belongs (if applicable). A composite is a measure with a score that is an aggregate of scores from other measures. A composite may include one or more other composites and/or individual measures. Composites may comprise component measures that can or cannot be used on their own.**

Not applicable.

1.G. Numerator Statement

The number of adverse events found in a patient sample.

The number of preventable adverse events found in a patient sample.

1.H. Numerator Exclusions

Not applicable.

1.I. Denominator Statement

The two denominators that can be used are 1,000 patient-days and 100 hospitalizations for all sampled patients who meet the following criteria:

1. Patients < 18 years of age at discharge.
2. Patients with a length of stay (LOS) greater than or equal to 24 hours.
3. Patients admitted for acute care. Acute care does not include patients in rehabilitation and residential units, non-acute inpatient psychiatric units, and day treatment areas. If a patient is initially admitted acutely but subsequently transferred to inpatient psychiatric care, the acute portion of the hospitalization should be included.
4. Patients who were discharged from or transferred out of the inpatient observation stay and patients who died during the stay.

1.J. Denominator Exclusions

We exclude patients who meet the above inclusion criteria but fall into the following exclusion categories:

1. Patients discharged from the emergency department without admission to the hospital.
2. Patients in newborn nurseries.

1.K. Data Sources

Check all the data sources for which the measure is specified and tested.

Paper medical record; electronic medical record.

If other, please list all other data sources in the field below.

Not applicable.

Section 2: Detailed Measure Specifications

Provide sufficient detail to describe how a measure would be calculated from the recommended data sources, uploading a separate document (+ Upload attachment) or a link to a URL. Examples of detailed measure specifications can be found in the CHIPRA Initial Core Set Technical Specifications Manual 2011 published by the Centers for Medicare & Medicaid Services. Although submission of formal programming code or algorithms that demonstrate how a measure would be calculated from a query of an appropriate electronic data source are not requested at this time, the availability of these resources may be a factor in determining whether a measure can be recommended for use.

Please see Supporting Documents for detailed measure specifications.

Section 3. Importance of the Measure

In the following sections, provide brief descriptions of how the measure meets one or more of the following criteria for measure importance (general importance, importance to Medicaid and/or CHIP, complements or enhances an existing measure). Include references related to specific points made in your narrative (not a free-form listing of citations).

3.A. Evidence for General Importance of the Measure

Provide evidence for all applicable aspects of general importance:

- **Addresses a known or suspected quality gap and/or disparity in quality (e.g., addresses a socioeconomic disparity, a racial/ethnic disparity, a disparity for Children with Special Health Care Needs (CSHCN), a disparity for limited English proficient (LEP) populations).**
- **Potential for quality improvement (i.e., there are effective approaches to reducing the quality gap or disparity in quality).**

- **Prevalence of condition among children under age 21 and/or among pregnant women.**
- **Severity of condition and burden of condition on children, family, and society (unrelated to cost).**
- **Fiscal burden of measure focus (e.g., clinical condition) on patients, families, public and private payers, or society more generally, currently and over the life span of the child.**
- **Association of measure topic with children’s future health – for example, a measure addressing childhood obesity may have implications for the subsequent development of cardiovascular diseases.**
- **The extent to which the measure is applicable to changes across developmental stages (e.g., infancy, early childhood, middle childhood, adolescence, young adulthood).**

The Problem of Patient Harm

Remarkable advances in medicine over the past several decades have been accompanied by a potential for causing inadvertent harm while caring for patients. Unintended patient harm from medical care, also known as adverse events (AEs), can include, but is not limited to, medical errors—(defined as errors in the delivery of medical care, whether they have the potential to cause harm or not—(Landrigan, 2005), surgical complications, and health care-associated infections (HAIs).

In its groundbreaking 1999 report, *To Err is Human*, the Institute of Medicine (IOM) estimated that medical errors contribute up to 98,000 deaths and 1 million injuries each year (Kohn, Corrigan, Donaldson, 1999). In 2010, the U.S. Department of Health and Human Services’ Office of the Inspector General estimated that 180,000 deaths due partly to AEs occur among Medicare patients annually, making AEs the third leading cause of death in the United States, after heart disease and cancer (Levinson, 2010; Allen, 2013). A more recent estimate suggests that the true figure could be twice as high (James, 2013).

In the wake of *To Err is Human* and similar reports, regulators, payers, nonprofit organizations, governments, and hospitals have launched numerous initiatives and invested resources over the past 15 years to improve patient safety (Landrigan, Parry, Bones, et al., 2010). The Centers for Medicare & Medicaid Services (CMS) and The Agency for Healthcare Research and Quality (AHRQ) have dedicated resources to measure and improve health care quality and patient safety for over a decade. During that time, there have been important successes in reducing certain types of harms, such as surgical complications and catheter-related bloodstream infections (Haynes, Weiser, Berry, et al., 2009; Pronovost, Needham, Berenholtz, et al., 2006). However, overall rates of harm due to medical care have not improved since the release of the IOM report (Landrigan, et al., 2010).

Prevalence of Inpatient Pediatric Patient Harm

Hospitalized pediatric patients are at high risk for AEs (Agarwal, Classen, Larsen, et al., 2010; Bates, Boyle, Vander Vliet, et al., 1995; Kirkendall, Kloppenborg, Papp, et al., 2012; Takata, Mason, Taketomo, et al., 2008; Woods, Thomas, Holl, et al., 2005). A pediatric study found an average rate of 37 AEs per 100 admissions, which is comparable to the average rate for adults of 40 AEs per 100 admissions (Kirkendall, et al., 2012). In some settings, rates among children may even be higher than the adult rate (Kirkendall, et al., 2012; Miller, Elixhauser, Zhan, 2003). Studies in neonatal and pediatric intensive care units have identified, respectively, 74 and 203 AEs per 100 patients (Agarwal, et al., 2010; Sharek, Horbar, Mason, et al., 2006). AE rates are higher among surgical patients and more acutely ill medical patients compared with other patients, in part due to the need for sedation and the use of medical devices (Agarwal, et al., 2010; Larsen, Donaldson, Parker, et al., 2007; Miller, et al., 2003).

Disparities in Risk of Harm

Children with special health care needs experience elevated rates of medical errors (Ahuja, Zhao, Xiang, 2012). Among hospitalized pediatric patients, those with chronic conditions are at significantly higher risk for medical errors than those without chronic conditions (Ahuja, et al., 2012).

What is known about racial/ethnic disparities in patient safety, particularly among children, is limited (Flores, Ngui, 2006). Black and Hispanic newborns are at higher risk of birth trauma (Miller, et al., 2003). In addition, extrapolations from associations between race/ethnicity and known risk factors for harm suggest that black and Hispanic children are likely at greater risk of harm than white children. Because severity of illness and complexity increase the risk of errors, and black and Hispanic children are at higher risk for more complex conditions, these children are at greater risk for AEs (Flores, Ngui, 2006). Emergency room visits are also associated with increased risk of AEs, and black and Hispanic children are known to visit the emergency department more frequently (Flores, Ngui, 2006; Matlow, Baker, Flintoft, et al., 2012).

Clinical, Familial, and Societal Burdens of Patient Harm

Failures in patient safety lead to substantial morbidity and mortality and thus have grave ramifications for patients and families (Zhan, Miller, 2003). Deaths due to AEs or medical errors cause enormous suffering for families. Furthermore, temporary or long-term injuries, as well as additional medical or surgical interventions necessitated by AEs, place psychological and financial burdens on patients and families (National Quality Forum, 2007).

Fiscal Burden of Patient Harm

AEs are costly for both hospitals and patients (Scott, 2009; Stone, Glied, McNair, et al., 2010). AEs are associated with increased lengths of hospital stay and additional procedures and medications and thus result in increased health care expenditures (Agarwal, et al., 2010; Larsen, et al., 2007; Miller, et al., 2003; Miller, Elixhauser, Zhan, et al., 2001; Miller, Zhan, 2004; Trybou, Spaepen, Vermeulen, et al., 2013). Pediatric patient safety events in 2000 resulted in over \$1 billion in excess charges (Miller, Zhan, 2004). HAIs alone account for between \$36 and

\$45 billion annually in potentially preventable inpatient expenditures (Scott, 2009; Stone, et al., 2010).

Variation in Risk of Harm Across Developmental Stages

Risk of pediatric patient harm varies with age across the inpatient population. The youngest patients are at significantly greater risk for encountering preventable AEs (Flores, Ngui, 2006; Larsen, et al., 2007; Miller, et al., 2003, 2001; Miller, Zhan, 2004). Rates of medication errors and potential adverse drug events are highest among neonatal intensive care unit patients and exceed those of adult patients (Kaushal, Bates, Landrigan, et al., 2001).

Association of Patient Harm with Future Health

AEs due to health care can have lasting negative effects on children's health. Although AEs most often result in transient harms, some lead to permanent injury and death. Research is underway on the long-term outcomes of children who suffer from health care-associated harm. The GAPPS Trigger Tool is designed to identify permanent as well as transient harms and thus facilitate the identification of events that affect the future health of hospitalized children.

Potential for Quality Improvement

Key strategies for reducing preventable harms in children include early detection and treatment of potential harm (Larsen, et al., 2007) and identification of potentially preventable AEs (McDonald, Davies, Haberland, et al., 2008). Use and further development of tools such as GAPPS to detect AEs are thus a critical part of efforts to improve patient safety (Agarwal, et al., 2010; Baker, Norton, Flintoft, et al., 2004; Takata, et al., 2008). By using tools that are more sensitive and more reliable, hospitals can increase their capacity to quantify inpatient AEs, identify priorities, and target available resources (Miller, et al., 2003).

3.B. Evidence for Importance of the Measure to Medicaid and/or CHIP

Comment on any specific features of this measure important to Medicaid and/or CHIP that are in addition to the evidence of importance described above, including the following:

- **The extent to which the measure is understood to be sensitive to changes in Medicaid or CHIP (e.g., policy changes, quality improvement strategies).**
- **Relevance to the Early and Periodic Screening, Diagnostic and Treatment benefit in Medicaid (EPSDT).**
- **Any other specific relevance to Medicaid/CHIP (please specify).**

Risk of Harm for Medicaid-Insured Children

Medicaid beneficiaries are at greater risk for nursing-related, surgery-related, and preventable AEs as compared with privately insured patients (Encinosa, Bernard, 2005). A recent study showed that in markets containing relatively few Medicaid payers and in hospitals with high proportions of Medicaid-insured patients, the chance of an AE is much greater than in markets with either greater competition among payers or hospitals with lower proportions of Medicaid-insured patients (Smith, Cheung, Owens, et al., 2007). In addition, patient harms occur more

often at financially stressed hospitals (Bernard, Encinosa, 2005). Rising costs and cuts to Federal Medicaid funding are associated with increases in AEs (Encinosa, Bernard, 2005).

Patient Safety and Medicaid/CHIP

CMS has made improving patient safety a priority for Medicaid and CHIP beneficiaries. CMS includes patient safety measures as part of voluntary quality measurement reporting for Medicaid and CHIP State agencies (Centers for Medicare & Medicaid Services [CMS], 2012).

CMS has also created incentives for reducing harms through reimbursement policies. In 2008, the agency published a list of “reasonably” preventable hospital-acquired conditions—such as hospital-acquired urinary tract infections and pressure ulcers—that qualify for reduced reimbursement (CMS, 2007). This policy raised awareness and increased resources for quality improvement and harm prevention activities in an unprecedented manner (Wald, Richard, Dickson, et al., 2012). In addition, the June 2011 rule on Provider Preventable Conditions (PPCs) prohibited Federal Medicaid reimbursement for medical expenditures associated with health care-acquired conditions and authorized States to implement non-payment policies for identified PPCs (CMS, 2011a, 2011b). The National Academy for State Health Policy has been working to align Federal and State policies around nonpayment or reduction in payment for preventable events and conditions, arguing that because Medicaid and CHIP programs account for one-sixth of all U.S. health care expenditures, aligning interests in nonpayment policies presents a crucial opportunity to advance patient safety (Rosenthal, Hanlon, 2009).

3.C. Relationship to Other Measures (if any)

Describe, if known, how this measure complements or improves on an existing measure in this topic area for the child or adult population, or if it is intended to fill a specific gap in an existing measure category or topic. For example, the proposed measure may enhance an existing measure in the initial core set, it may lower the age range for an existing adult-focused measure, or it may fill a gap in measurement (e.g., for asthma care quality, inpatient care measures).

Identification of Inpatient AEs

Various approaches exist for identifying AEs. Voluntary passive reporting systems are commonly employed but recognized to have low sensitivity (Sharek, et al., 2006). Although 20 States require mandatory reporting of serious AEs, no published studies appear to have cited these data (Zhan, Miller, 2003). This is at least in part due to questions about the reliability of voluntarily reported patient safety data. A more reliable, sensitive methodology for capturing data on the safety of hospital care is thus essential (Classen, Resar, Griffin, et al., 2011; Landrigan, et al., 2010).

In 2003, AHRQ released its Patient Safety Indicators (PSIs), developed in response to a congressional mandate to reduce medical errors (Miller, et al., 2003). PSIs are intended to identify events that most likely resulted from preventable medical errors (Dynan, Goudie, Smith, et al., 2013; Miller, et al., 2003; Miller, Zhan, 2004). PSIs have been used with some success but have a number of limitations, in part due to their reliance on administrative data. They have also been found to have low sensitivity (Classen, et al, 2011; Miller, et al., 2003).

Use of trigger tools has been shown to be a faster, more sensitive, and more reliable method of detection than other approaches (Agarwal, et al., 2010; Classen, et al., 2011; Resar, Rozich, Classen, 2003; Rozich, Haraden, Resar, 2003; Sharek, Classen, 2006; Takata, et al., 2008). “Triggers” are red flags in a medical record that may indicate the presence of an underlying AE and prompt further inspection to determine whether an AE has occurred (Resar, et al., 2003; Rozich, et al., 2003). An example of a trigger is the documented administration of an antidote-type medication (e.g., naloxone). Once a trigger is found, an in-depth review is undertaken to determine whether an AE occurred. In the case of naloxone, administration may indicate an AE occurred if the drug was given to counteract an overdose of opioids given in the hospital but, on the other hand, may not if the overdose occurred due to voluntary recreational opioid use. Trigger tools detect AEs in a high percentage of hospitalizations, ranging in published reports from 3 percent to 63 percent, and have evolved significantly over time (Baker, et al., 2004; Kennerly, Kudyakov, da Graca, et al., 2014; Kennerly, Saldana, Kudyakov, et al., 2013; Kirkendall, et al., 2012; Matlow, et al., 2012; Stockwell, Kirkendall, Muething, et al., 2013; Thomas, Studdert, Burstin, et al., 2000; Wilson, Runciman, Gibberd, et al., 1995).

The Global Trigger Tool for Measuring Adverse Events (GTT), developed by the Institute for Healthcare Improvement (IHI), has become widely accepted as an effective approach for identifying AEs in hospitalized adult patients (Agarwal, et al., 2010; Baker, et al., 2004; Kohn, et al., 1999; Sharek, et al., 2006; Takata, et al., 2008; Thomas, et al., 2000; Wilson, et al., 1995). The GTT approach identifies 10 times as many AEs as AHRQ’s PSIs and almost 100 times as many events as voluntary reporting (Classen, et al., 2011).

Need for a Pediatric Inpatient Trigger Tool

One study determined that a version of the GTT applied to the pediatric population was effective at identifying pediatric AEs (Takata, et al., 2008). However, the authors and other experts have called for development of a standardized pediatric tool that focuses more specifically on the problems of hospitalized children and that encompasses the breadth of inpatient pediatric care (Agarwal, et al., 2010; Kirkendall, et al., 2012; Sharek, et al., 2006; Takata, et al., 2008). The absence of a comprehensive pediatric trigger tool is a recognized limitation in quantifying the full scope of pediatric AEs. An early effort to develop a pediatric-focused trigger tool led to the development of the Canadian Pediatric Trigger Tool (Matlow, et al., 2012).

We developed GAPPS to meet the need for a comprehensive, sensitive measure of pediatric patient safety. We built GAPPS iteratively based on the Canadian Pediatric Trigger Tool and multiple other published tools. We used methods similar to those used for GTT, including manual medical record review by experts in patient safety, which has been demonstrated to be a crucial component of developing patient safety measures (Scanlon, Harris, Levy, et al., 2008).

GAPPS offers an enhancement in trigger tool methodology in that, unlike GTT, it requires reviewers to assess preventability. In the 5 years since IHI released the second edition of GTT, patient safety experts and national fiscal and quality improvement policies have increasingly focused on addressing preventable AEs. GAPPS uses the same approach to rate preventability as

the North Carolina Patient Safety study, which was found to ascertain preventability with a high degree of reliability (Sharek, Parry, Goldmann, et al., 2011).

Section 4. Measure Categories

CHIPRA legislation requires that measures in the initial and improved core set, taken together, cover all settings, services, and topics of health care relevant to children. Moreover, the legislation requires the core set to address the needs of children across all ages, including services to promote healthy birth. Regardless of the eventual use of the measure, we are interested in knowing all settings, services, measure topics, and populations that this measure addresses. These categories are not exclusive of one another, so please indicate "Yes" to all that apply.

Does the measure address this category?

- a. Care Setting – ambulatory: No.**
- b. Care Setting – inpatient: Yes.**
- c. Care Setting – other – please specify: No.**
- d. Service – preventive health, including services to promote healthy birth: No.**
- e. Service – care for acute conditions: Yes.**
- f. Service – care for children with acute conditions: Yes.**
- g. Service – other (please specify): No.**
- h. Measure Topic – duration of enrollment: No.**
- i. Measure Topic – clinical quality: Yes.**
- j. Measure Topic – patient safety: Yes.**
- k. Measure Topic – family experience with care: No.**
- l. Measure Topic – care in the most integrated setting: No.**
- m. Measure Topic other (please specify): No.**
- n. Population – pregnant women: Yes.**
- o. Population – neonates (28 days after birth) (specify age range): Yes; 0-28 days.**
- p. Population – infants (29 days to 1 year) (specify age range): Yes; 29-364 days.**
- q. Population – pre-school age children (1 year through 5 years) (specify age range): Yes; 1-5 years.**
- r. Population – school-aged children (6 years through 10 years) (specify age range): Yes; 6-10 years.**
- s. Population – adolescents (11 years through 20 years) (specify age range): Yes; 11-17 years.**
- t. Population – other (specify age range): No.**
- u. Other category (please specify): Not applicable.**

Section 5. Evidence or Other Justification for the Focus of the Measure

The evidence base for the focus of the measures will be made explicit and transparent as part of the public release of CHIPRA deliberations; thus, it is critical for submitters to

specify the scientific evidence or other basis for the focus of the measure in the following sections.

5.A. Research Evidence

Research evidence should include a brief description of the evidence base for valid relationship(s) among the structure, process, and/or outcome of health care that is the focus of the measure. For example, evidence exists for the relationship between immunizing a child or adolescent (process of care) and improved outcomes for the child and the public. If sufficient evidence existed for the use of immunization registries in practice or at the State level and the provision of immunizations to children and adolescents, such evidence would support the focus of a measure on immunization registries (a structural measure).

Describe the nature of the evidence, including study design, and provide relevant citations for statements made. Evidence may include rigorous systematic reviews of research literature and high-quality research studies.

Use of Trigger Tools to Identify Patient Harm

The GAPPS measure uses a trigger tool to detect AEs in hospitalized children. Table 5.A.1 (see Supporting Documents) summarizes the key studies related to trigger tool identification of AEs and the importance of measuring patient safety. Multiple studies suggest that trigger tools are valid and reliable for tracking the incidence of patient harms in hospital settings, including within pediatric populations. Trigger tools, including automated approaches to trigger tools, are better able to detect AEs than other methods—for example, traditional voluntary incident reporting and detection tools used with administrative databases (Agarwal, et al., 2010; Call, Burlison, Robertson, et al., 2014; Classen, et al., 2011; Classen, Pestotnik, Evans, et al., 1991; Ferranti, Horvath, Cozart, et al., 2008; Griffin, Classen, 2008; Hooper, Tibballs, 2014; Kennerly, et al., 2014; Levinson, 2010; Rozich, et al., 2003; Stockwell, et al., 2013; Takata, et al., 2008). Trigger tool studies have identified specific triggers that have high yields for AE detection, such as “return to surgery,” “positive blood culture,” or “abrupt medication stop” (Naessens, O’Byrne, Johnson, et al., 2010; Resar, Rozich, Simmonds, et al., 2006; Rozich, et al., 2003). These findings were applied when developing the triggers used in the GAPPS measure.

In addition, nearly all studies evaluating mean medical record review times reported times under 30 minutes (Agarwal, et al., 2010; Lander, Roberson, Plummer, et al., 2010; Sharek, et al., 2006; Unbeck, Schildmeijer, Henriksson, et al., 2013), indicating that trigger tools can be used to adequately detect AEs with a reasonably small time burden. The GAPPS methodology likewise employs a maximum 30-minute timeframe. Since GAPPS can be applied using either a manual or automated approach (the difference is whether triggers are initially identified by a primary reviewer or by an algorithm programmed into an electronic health record [EHR] system), the automated approach to identifying AEs may decrease the time burden even further as shown in several studies (Bates, Evans, Murff, et al., 2003; Jha, Kuperman, Teich, et al., 1998; Kaiser, de Jong, Evelein-Brugman, et al., 2014; Kirkendall, Spires, Mottes, 2014).

Relationship Among Patient Safety and Other Aspects of Quality

Some studies in adult populations have shown that patient harm rates are associated with other aspects of clinical quality. These studies demonstrate that rates of patient harm directly correlate with other quality metrics, including performance on clinical processes of care and other health outcomes (Calder, Tierney, Jiang, et al., 2014; Joice, Deibert, Kates et al., 2013; Rosen, Geraci, Ash, et al., 1992; Rosen, Loveland, Shin, et al., 2013; Sukumar, Roghmann, Trinh., et al., 2013). For example, evidence shows that patients who experience health care-related harms have greater odds of in-hospital and 30-day mortality, as well as 30-day readmission (Joice, et al., 2013; Rosen, et al., 2013; Rosen, et al., 1992; Sukumar, et al., 2013). It is likely that the same general associations hold true for pediatric populations.

Measuring Patient Safety to Drive Quality Improvement

Patient safety is a core domain of health care quality and a major focus for quality improvement efforts (Kohn, et al., 1999; Levinson, 2010). Hospitals have been able to demonstrate that having a reliable means to track AEs leads to improvements in patient safety and associated clinical outcomes (Cohen, Kimmel, Benage, et al., 2005; Garrett Jr, Sammer, Nelson, et al., 2013; Evans, Pestotnik, Classen, et al., 1992; Kaufman, Rannie, Kahn, et al., 2012; Piontek, Kohli, Conlon, et al., 2010). For example, hospitals that institute real-time adverse drug event surveillance systems are able to intervene before AEs become severe, or they are able to prevent future AEs altogether (Evans, et al., 1992; Kaufman, et al., 2012; Piontek, et al., 2010). In addition, the ability to track AEs allows for the design, implementation, and evaluation of targeted interventions, resulting in fewer AEs and decreased mortality (Garrett, et al., 2013; Kaufman, et al., 2012; Piontek, et al., 2010).

5.B. Clinical or Other Rationale Supporting the Focus of the Measure (optional)

Provide documentation of the clinical or other rationale for the focus of this measure, including citations as appropriate and available.

See Section 5.A, above, for details regarding the association of patient harm rates and clinical outcomes.

Section 6. Scientific Soundness of the Measure

Explain the methods used to determine the scientific soundness of the measure itself. Include results of all tests of validity and reliability, including description(s) of the study sample(s) and methods used to arrive at the results. Note how characteristics of other data systems, data sources, or eligible populations may affect reliability and validity.

6.A. Reliability

Reliability of the measure is the extent to which the measure results are reproducible when conditions remain the same. The method for establishing the reliability of a measure will depend on the type of measure, data source, and other factors.

Explain your rationale for selecting the methods you have chosen, show how you used the methods chosen, and provide information on the results (e.g., the Kappa statistic). Provide appropriate citations to justify methods.

We tested 54 candidate triggers in a sample of 3,814 randomly selected pediatric medical records (i.e., for patients <18 years of age) from 16 hospitals evenly distributed among the four U.S. Census regions. The hospitals included equal numbers of teaching and non-teaching institutions, with at least one teaching and one non-teaching hospital in each region. For a description of how our candidate triggers were chosen, see Section 6.B. Records were eligible for inclusion if the patient had an inpatient length of stay between 24 hours and 6 months in an acute care area. Records were sampled at each site by taking a random sample of 10 medical records for each quarter between 2007 and 2012.

The candidate triggers were evaluated using three groups of medical records in each hospital (see Figure 6.A.1 in the Supporting Documents). Group A consisted of the total sample of medical records from each hospital that were reviewed by primary (nurse) and secondary (physician) reviewers internal to the hospital. Group B consisted of a random sub-sample of the records in Group A (i.e., 24 records per site) that were reviewed again by additional primary and secondary reviewers internal to each hospital. Group C consisted of a separate random sub-sample (i.e., 24 records per site) of the records reviewed in Group A that were reviewed again by primary and secondary reviewers external to each hospital (see Figure 6.A.1 in the Supporting Documents).

We used Groups A and B to evaluate the reliability of the measure. We compared ratings from pairs of independent secondary reviewers within Group A. We also compared primary reviewer findings (Group A) to a second primary reviewer's findings (Group B) for the same medical records. To assess reliability, we used a Kappa statistic for variables with only two possible outcomes and a weighted Kappa computed with Fleiss-Cohen weights for variables with more than two possible ordinal outcomes (Fleiss, Cohen, 1973, Landis, Koch, 1977). We used the categorization of Landis and Koch (1977) to interpret reliability for ranges of Kappa scores ($k < 0$: poor, $k = 0.00-0.20$: slight, $k = 0.21-0.40$: fair, $k = 0.41-0.60$: moderate, $k = 0.61-0.80$: substantial, $k = 0.81-1.00$: almost perfect).

The reliability for internal primary reviewers (Group A versus Group B, [$n = 379$]) was “substantial” for both determination of the total number of suspected AEs (Kappa = 0.73, 95 percent CI 0.62 – 0.85) and identification of at least one suspected AE (Kappa = 0.69, 95 percent CI 0.59 – 0.79). Both primary reviewers agreed on the total number of AEs 88 percent of the time, and they agreed that a record did or did not contain at least one AE 92 percent of the time. In some cases, however, the AEs identified in the medical records differed. From the total sample of records reviewed, primary reviewers identified the same AEs 62 percent of the time.

The two internal secondary reviewers in Group A independently determined the presence or absence of an AE among suspected AEs identified by the primary reviewer in Group A ($n = 617$). Internal secondary reviewers verified the same suspected AEs 92 percent of the time, with “almost perfect” reliability (Kappa = 0.81, 95 percent CI 0.76 – 0.86).

The reliability for internal secondary reviewers in Group A versus Group B ($n = 379$) was “substantial” for both determination of the total number of suspected AEs (Kappa = 0.73, 95

percent CI 0.57 – 0.89) and verification of at least one suspected AE (Kappa = 0.70, 95 percent CI 0.59 – 0.81). Internal secondary reviewers in Group A and Group B agreed on the total number of AEs 92 percent of the time, and they agreed that a record did or did not contain at least one AE 94 percent of the time. In some cases, however, the AEs identified in the medical records differed.

As reviewers gained experience over the course of the study, Group A primary reviewers' agreement with external expert primary reviewers increased. In particular, the level of their agreement on identifying any AE improved from 78 percent (Kappa = 0.27) for the first one-third of records reviewed to 84 percent (Kappa = 0.51) for the last one-third of records, and primary reviewers' judgments about the number of AEs they identified became significantly more reliable (Kappa = 0.21 vs. 0.51, $p=0.05$). Similarly, level of agreement between experienced hospital-based primary reviewers (i.e., those who had participated in a prior trigger tool study) and external primary reviewers was higher (83 percent [Kappa = 0.53]) than agreement between inexperienced hospital-based primary reviewers and external primary reviewers (69 percent [Kappa = 0.24]), suggesting that further training would improve agreement between hospital-based primary reviewers and expert reviewers.

Based on the frequency with which each trigger and AE was identified in medical records, two final approaches were created for the GAPPS Trigger Tool: a manual approach and an automated approach. The manual approach requires primary reviewers to look for and identify triggers in medical records, whereas the automated approach uses a hospital's EHR system to automatically flag medical records that contain triggers.

When determining which triggers should be included in the manual approach, we decided to include all triggers that led to detection of an AE greater than or equal to 10 percent of the time ($n=27$). Triggers that appeared fewer than 10 times in the sample were removed because we lacked adequate data to evaluate their performance. However, we made exceptions to this rule for three triggers for reasons specific to each trigger: "Naloxone administration," "Health care-associated infections: positive *C. difficile* test," and "Pressure ulcer documentation." These triggers were included because of the severity of the condition, preventability of the AEs, or high likelihood of the trigger being associated with an AE. One qualifying trigger – "low O2 saturation" – was removed because it was redundant with other triggers in detecting AEs.

When determining which triggers should be included in the automated approach, we included all triggers on the manual list that could be automated in an academic tertiary care hospital's EHR system, as well as any low frequency (fewer than 10 occurrences) triggers that could be automated in the academic tertiary care hospital's EHR system and that led to the detection of an AE greater than or equal to 10 percent of the time when further tested on more records. The final automated trigger list included 30 triggers.

6.B. Validity

Validity of the measure is the extent to which the measure meaningfully represents the concept being evaluated. The method for establishing the validity of a measure will depend on the type of measure, data source, and other factors.

Explain your rationale for selecting the methods you have chosen, show how you used the methods chosen, and provide information on the results (e.g., R2 for concurrent validity).

We developed the draft trigger tool used in the GAPPS measure through the RAND/UCLA Appropriateness Method, which is a modified Delphi process (Brown, 1968; Fitch, Bernstein, Aguilar, et al., 2001; Sweidan, Williamson, Reeve, et al., 2010). We first compiled a set of 78 candidate triggers from a literature review of existing pediatric and adult trigger tools and input from trigger tool experts (Griffin, Resar, 2009; Kirkendall, et al., 2012; Stockwell, Bisarya, Classen, et al., 2015). We then recruited nine panelists from national pediatric and patient safety organizations and asked them to rate separately the validity and feasibility of the candidate triggers on a 9-point scale (where 1 is the least valid/feasible and 9 is the most valid/feasible). A trigger was considered valid if it was judged to be reasonably likely to identify an underlying AE, indicating that harm potentially occurred. A trigger was considered feasible if it was judged likely to be accurately and consistently documented in either paper or electronic medical records as part of patient care at a wide range of hospitals, from smaller community sites to larger tertiary care centers. Applying the RAND/UCLA Appropriateness Method, we accepted triggers that had both median validity and feasibility ratings greater than or equal to seven. This approach resulted in inclusion of 54 of the initial 78 candidate triggers in the draft GAPPS trigger list.

It is not possible to assess the performance of the GAPPS measure against a true “gold standard” for detection of AEs because such a gold standard does not yet exist. We therefore focused our validity testing on evaluation of how accurately and completely “typical reviewers” (i.e., clinicians who are trained in GAPPS methodology but not necessarily trigger tool experts) were able to identify AEs using the measure as compared to expert reviewers. The expert reviewers had extensive experience with using trigger tools for AE identification and consequently were most likely to identify AEs accurately and completely. To evaluate the validity of the GAPPS measure, we assessed the performance of the National Field Test’s hospitals’ internal reviewers relative to the performance of external expert reviewers in applying the measure (i.e., we compared findings of reviewers in Group A versus Group C, described in Section 6.A).

As summarized in Table 6.B.1 (see Supporting Documents), using the findings of the external reviewers as the standard of comparison, the specificity for identifying a record with one or more AEs was 0.91 for primary reviewers and 0.95 after taking secondary reviewer verification into account. The sensitivity was 0.40 for primary reviewers and 0.33 after taking secondary reviewer verification into account. The lower sensitivity is likely due in part to the novice reviewers’ lack of experience with the tool and their inability to make up for their inexperience by increasing the amount of time they took to perform their review, given that there was a 30-minute time limit per record.

Section 7. Identification of Disparities

CHIPRA requires that quality measures be able to identify disparities by race, ethnicity, socioeconomic status, and special health care needs. Thus, we strongly encourage nominators to have tested measures in diverse populations. Such testing provides evidence for assessing measure’s performance for disparities identification. In the sections below, describe the results of efforts to demonstrate the capacity of this measure to produce

results that can be stratified by the characteristics noted and retain the scientific soundness (reliability and validity) within and across the relevant subgroups.

7.A. Race/Ethnicity

We assessed differences in pediatric patient safety associated with race/ethnicity by evaluating whether the rate of AEs identified by reviewers varied among racial/ethnic groups.

Race/ethnicity was recorded in our national field test using the categories Alaska Native, American Indian, Asian, black, Hispanic, Native Hawaiian or other Pacific Islander, Other, and white. For our analysis, we combined Alaska Native, American Indian, Asian, Native Hawaiian or other Pacific Islander, and non-white, non-Hispanic Other patients into a single “Other” category because each of the categories was represented by a very small number of hospitalizations. In our adjusted analyses, we controlled for Chronic Condition Indicators (CCIs) and length of hospitalization when looking at differences in AE rates based on race/ethnicity since more CCIs and longer length of hospitalizations are associated with a greater number of AEs ($p < 0.001$) (AHRQ, 2017; Perrin, Newacheck, Pless, et al., 1993).

Two sites had missing race/ethnicity data for a large percentage of patients (57 percent and 97 percent, respectively) and were thus excluded from the analysis. Also, CCI (a covariate in the analysis) was missing for 85 percent of the hospitalizations at one site, and therefore the site was excluded from the analysis. This resulted in an analytical sample size of 2,834 hospitalizations. The distribution of race/ethnicity and AEs across sites is shown in Table 7.A.1 (see Supporting Documents).

Across all sites evaluated, we found that Hispanic patients had a higher unadjusted AE rate at 0.18, compared to white patients at 0.09 ($p < 0.001$).

We examined the association between race/ethnicity and AE risk using a Poisson multivariate regression model accounting for over-dispersion with a hospital random intercept, controlling for the number of chronic conditions a patient had (as classified in the CCI system) and length of hospitalization. We categorized CCIs into four levels: zero body system, one body system, two body systems, and three or more body systems.

After controlling for CCIs and length of hospitalization, race/ethnicity was significantly associated with AE rate ($p < 0.001$), with black patients (Relative Risk Estimate [RRE] 0.52, 95 percent CI: 0.37-0.73) having a significantly lower risk of AEs than white patients ($p < 0.001$). Hispanic and Other race/ethnicity patients were not significantly different from white patients (see Table 7.A.2 in the Supporting Documents).

The relationship between race/ethnicity and AE rates seems to vary among hospitals even when controlling for CCIs and length of hospitalization. Particularly, at one hospital, Hispanic patients had a seemingly high number of AEs, despite the overall Hispanic population in our study showing no differences when compared to white patients (see Table 7.A.2 in the Supporting Documents). However, this discrepancy in AE rates between Hispanics and whites was not the case for some other hospitals. This may suggest that race/ethnicity is not a uniform predictor of AEs across hospitals. It may therefore be important to also measure AE rates stratified by race/ethnicity and hospital to assess potential differences among hospitals in the quality of care provided for patients of different racial/ethnic groups. The limited literature dealing with the

relationship between patient safety and race/ethnicity seems to show contradicting associations between the two variables as well, with certain races/ethnicities having higher or lower AE rates than whites for various types of AEs (AHRQ, 2011; Gaskin, Spencer, Richard, et al., 2008; Metersky, Hunt, Kilman, et al., 2011; Perrin, et al., 1993).

7.B. Special Health Care Needs

We assessed differences in pediatric patient safety associated with special health care needs by evaluating whether the rate of AEs identified by reviewers varied based on the chronic conditions present among patients (as classified in the CCI system), controlling for length of hospitalization.

One site was missing diagnosis data in 85 percent of its hospitalizations and was therefore excluded from the analysis. This resulted in an analytical sample size of 3,442 hospitalizations. The distribution of numbers of CCIs per patient and AEs by site is shown in Table 7.B.1 (see Supporting Documents).

Overall, we found that patients with a body system affected by a chronic condition had higher unadjusted AE rates ($p < 0.001$) than those without any body system affected by a chronic condition. Particularly, patients with three or more body systems affected by a chronic condition had the highest unadjusted AE rate at 0.27 ($p < 0.001$) as shown in Table 7.B.1 (see Supporting Documents).

We examined the association between a patient's number of body systems affected by a chronic condition and AE risk using a Poisson multivariate regression accounting for over-dispersion with a hospital random intercept. After controlling for length of hospitalization, having at least one body system affected by a chronic condition was associated with a higher AE rate ($p < 0.001$). The relationship between a patient's number of CCIs and AE rate varied among hospitals ($p < 0.001$). Specifically, the expected number of AEs identified in patients with multiple body systems affected by a chronic condition is estimated to be approximately 1.7 times as great as the expected number of AEs identified in patients without any body system affected by a chronic condition, as shown in Table 7.B.2 (see Supporting Documents)—two body systems $p = < 0.001$; three or more body systems $p = 0.01$.

7.C. Socioeconomic Status

We assessed differences in pediatric patient safety associated with socioeconomic status (SES) by using insurance status as a proxy for SES and examining whether the rate of AEs identified by reviewers varied with insurance status. Insurance status was captured in our national field test using six non-mutually exclusive categories: Medicaid, Medicare, Private Insurance, Self-Pay, No Insurance, and Not Recorded. These sorted the cohort into eight unique categories (some of which indicate that a patient had multiple insurance types listed during the hospitalization included in our field test): no insurance; private insurance; public insurance; private insurance and self-pay; public and private insurance; public and no insurance; public, self-pay, and private insurance; and insurance not recorded. One site was excluded from the analysis because 85 percent of its hospitalizations had missing data on the CCI (covariate in the insurance analysis).

We chose to exclude hospitalizations for patients covered by Medicare from the analysis because pediatric eligibility for Medicare is based on having specific medical conditions rather than being based solely on family income (CMS, 2014). Insurance type was missing for all of the hospitalizations at one site because reviewers did not have access to this information; we therefore excluded this site from the analysis. Records at other sites were also excluded due to missing information on insurance and length of hospitalization, resulting in an analytical sample of 3,151 hospitalizations. Missing data for insurance appeared to be randomly distributed across all sites (except the site with missing insurance data). Therefore, we evaluated patients with public insurance (Medicaid), private insurance, and no insurance. Patients who were recorded to have both private insurance and public insurance were categorized as patients with private insurance.

Table 7.C.1 (see Supporting Documents) shows the distribution of the insurance types included in our analysis: 38.6 percent of the patients had Medicaid, 58.2 percent had private insurance, and 3.2 percent did not have insurance. Patients with private insurance (AE rate 0.10) and no insurance (AE rate 0.04) had lower crude AE rates across sites than patients with Medicaid insurance (AE rate 0.14).

We examined the association between insurance type and AE risk using an over-dispersion corrected Poisson multivariate regression model with a hospital random intercept, controlling for patient's number of body systems affected by a chronic condition and length of hospitalization. When controlling for CCIs and length of hospitalization, insurance type was not significantly associated with AE risk ($p=0.68$, see Table 7.C.2 in the Supporting Documents).

7.D. Rurality/Urbanicity

The measure cannot be used to evaluate differences in patient safety associated with residence in rural versus urban settings. This measure relies solely on medical records for such information, and unfortunately, patients' residences are not reliably or consistently categorized in hospital records.

The measure potentially could be modified to incorporate richer clinical information from EHR data, once these data are more widely available.

7.E. Limited English Proficiency (LEP) Populations

Because the measure uses medical record data, which do not include information on English proficiency for patients or their families, the measure cannot be used to identify differences in patient safety based on limited English proficiency. The measure could potentially be modified to incorporate language proficiency information from EHR data, once these data are more widely available and language proficiency is documented in EHRs in a more standardized and consistent way. The measure might then be used to assess differences associated with limited English proficiency.

Section 8. Feasibility

Feasibility is the extent to which the data required for the measure are readily available, retrievable without undue burden, and can be implemented for performance measurement. Using the following sections, explain the methods used to determine the feasibility of implementing the measure.

8.A. Data Availability

1. What is the availability of data in existing data systems? How readily are the data available?

The GAPPS measure uses data obtained from electronic and/or paper patient medical records. The triggers used in the measure rely on clinical information, such as clinical notes, laboratory results, and medication order histories. This information is routinely recorded during the course of inpatient care and thus is readily available to providers and hospitals.

2. If data are not available in existing data systems or would be better collected from future data systems, what is the potential for modifying current data systems or creating new data systems to enhance the feasibility of the measure and facilitate implementation?

While the measure can be applied to either EHRs or paper records, EHRs are being adopted at increasing rates by hospitals and offer multiple benefits. First, EHRs are readily accessible, whereas paper records may be missing or stored in relatively inaccessible locations, such as off-site warehouses. In addition, as methods for querying EHRs become more sophisticated, detection of triggers that are part of the GAPPS measure could potentially be more thoroughly automated. Automated trigger detection allows for much more rapid and consistent identification of triggers and makes the AE detection process less labor-intensive. Manual review would only be needed for assessing the presence of AEs once triggers have already been identified, and eventually, even that process could be automated to some degree.

8.B. Lessons from Use of the Measure

1. Describe the extent to which the measure has been used or is in use, including the types of settings in which it has been used, and purposes for which it has been used.

We tested implementation of the GAPPS measure in eight teaching and eight non-teaching U.S. hospitals that are well-distributed among the Midwest, Northeast, South, and West. The goal of this national field test was to evaluate how reliably the measure detects AEs across a variety of settings. The hospitals received the Manual of Operations (see Supporting Documents), which provides detailed guidelines on using the GAPPS Trigger Tool to measure pediatric AEs. The manual outlines step-by-step instructions for collecting, reviewing, and managing the medical record review process.

Prior to the review process, experts in trigger tool methodology led three sessions of a Web-based training with the reviewers/trainees. During the first session, the experts discussed the trigger tool method and the concept of AEs. Between the first and second sessions, trainees reviewed 10 redacted example charts; during the second session, experts presented findings from these charts. Between the second and third sessions, trainees reviewed 10 medical records from

their own hospitals. In the third session, they presented their findings from these records and asked any remaining questions about identifying AEs and determining their severity and preventability.

2. If the measure has been used or is in use, what methods, if any, have already been used to collect data for this measure?

In each of the test hospitals, 240 unique hospitalizations were examined. Primary reviewers first reviewed the medical records looking for triggers and suspected AEs. Secondary reviewers then evaluated all suspected AEs to confirm whether an AE had indeed occurred and, if so, to rate its severity and preventability. Information about triggers, suspected AEs, and associated details were recorded in a standardized Web-based case report form.

3. What lessons are available from the current or prior use of the measure?

Feedback from the 16 hospitals indicated that the measure is straightforward to use and easily understandable. Based on helpful suggestions from the hospitals, we improved the clarity of the Manual of Operations. To make the measure more efficient for reviewers, we removed triggers that performed poorly in the detection of AEs. The GAPPS National Field Test also demonstrated that it is crucial to provide rigorous training and feedback to reviewers on practice cases prior to their use of the measure in order to achieve optimal standardization in AE detection.

Section 9. Levels of Aggregation

CHIPRA states that data used in quality measures must be collected and reported in a standard format that permits comparison (at minimum) at State, health plan, and provider levels. Use the following table to provide information about this measure's use for reporting at the levels of aggregation in the table.

For the purpose of this section, please refer to the definitions for provider, practice site, medical group, and network in the Glossary of Terms.

If there is no information about whether the measure could be meaningfully reported at a specific level of aggregation, please write "Not available" in the text field before progressing to the next section.

Level of aggregation (Unit) for reporting on the quality of care for children covered by Medicaid/ CHIP†:

State level Can compare States*

**Intended use: Is measure intended to support meaningful comparisons at this level?
(Yes/No)**

No.

Data Sources: Are data sources available to support reporting at this level?

No.

Sample Size: What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?

Not applicable.

In Use: Have measure results been reported at this level previously?

No.

Reliability & Validity: Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?

No.

Unintended consequences: What are the potential unintended consequences of reporting at this level of aggregation?

Not applicable.

Other geographic level: Can compare other geographic regions (e.g., MSA, HRR)

Intended use: Is measure intended to support meaningful comparisons at this level? (Yes/No)

No.

Data Sources: Are data sources available to support reporting at this level?

No.

Sample Size: What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?

Not applicable.

In Use: Have measure results been reported at this level previously?

No.

Reliability & Validity: Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?

No.

Unintended consequences: What are the potential unintended consequences of reporting at this level of aggregation?

Not applicable.

Medicaid or CHIP Payment model: Can compare payment models (e.g., managed care, primary care case management, FFS, and other models)

Intended use: Is measure intended to support meaningful comparisons at this level? (Yes/No)

No.

Data Sources: Are data sources available to support reporting at this level?

No.

Sample Size: What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?

Not applicable.

In Use: Have measure results been reported at this level previously?

No.

Reliability & Validity: Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?

No.

Unintended consequences: What are the potential unintended consequences of reporting at this level of aggregation?

Not applicable.

Health plan*: *Can compare quality of care among health plans.*

Intended use: Is measure intended to support meaningful comparisons at this level?

(Yes/No)

No.

Data Sources: Are data sources available to support reporting at this level?

No.

Sample Size: What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?

Not applicable.

In Use: Have measure results been reported at this level previously?

No.

Reliability & Validity: Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?

No.

Unintended consequences: What are the potential unintended consequences of reporting at this level of aggregation?

Not applicable.

Provider Level

Individual practitioner: *Can compare individual health care professionals*

***Intended use:* Is measure intended to support meaningful comparisons at this level?
(Yes/No)**

No.

***Data Sources:* Are data sources available to support reporting at this level?**

No.

***Sample Size:* What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

Not applicable.

***In Use:* Have measure results been reported at this level previously?**

No.

***Reliability & Validity:* Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**

No.

***Unintended consequences:* What are the potential unintended consequences of reporting at this level of aggregation?**

Not applicable.

Provider Level

Hospital: Can compare hospitals

***Intended use:* Is measure intended to support meaningful comparisons at this level?
(Yes/No)**

Yes.

***Data Sources:* Are data sources available to support reporting at this level?**

No.

***Sample Size:* What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

To achieve adequate reliability for estimates of hospital-level AE incidence, we recommend a minimum sample size of 60 records per quarter, collected by selecting 20 records from each month in the quarter. This sample size is based on the assumption that the trigger tool will be used in an improvement setting for which the aim is to detect trends in the data that show meaningful change over time. According to Perla and colleagues, to plot the data quarterly, the appropriate sample size of medical records is given by $9/R$, where R is the average number of AEs per person (Perla, Provost, Murray, 2013). Assuming an AE rate of at least 0.15, the recommended sample size computes to $9/0.15 = 60$.

***In Use:* Have measure results been reported at this level previously?**

No.

Reliability & Validity: Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?

No.

Unintended consequences: What are the potential unintended consequences of reporting at this level of aggregation?

Implementation of GAPPS could substantially increase the detection of inpatient AEs, which would support quality improvement efforts but could also have unintended consequences. Because of programs that reduce reimbursement for preventable health care-acquired conditions, increased AE detection could have negative financial ramifications for hospitals, creating an incentive to underreport events. We suggest establishing a system to audit hospital GAPPS reports if the measure is ultimately used to identify events that affect hospital reimbursement. As the GAPPS measure is implemented, it will be important for payers such as Medicaid and CHIP to carefully consider the implications of introducing new types of “never” events or otherwise altering payment structures in response to broader reporting of inpatient AEs. We recommend careful consideration and monitoring of this issue on a small scale before a large-scale rollout of the measure.

Provider Level

Practice, group, or facility: Can compare: (i) practice sites; (ii) medical or other professional groups; or (iii) integrated or other delivery networks**

Intended use: Is measure intended to support meaningful comparisons at this level? (Yes/No)

No.

Data Sources: Are data sources available to support reporting at this level?

No.

Sample Size: What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?

Not applicable.

In Use: Have measure results been reported at this level previously?

No.

Reliability & Validity: Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?

No.

Unintended consequences: What are the potential unintended consequences of reporting at this level of aggregation?

Not applicable.

Section 10. Understandability

CHIPRA states that the core set should allow purchasers, families, and health care providers to understand the quality of care for children. Please describe the usefulness of this measure toward achieving this goal. Describe efforts to assess the understandability of this measure (e.g., focus group testing with stakeholders).

The 1999 IOM report *To Err is Human* received extensive media coverage, garnering attention from the public, the Federal Government, and the health care industry (Stelfox, Palmisani, Scurlock, et al., 2006). Blendon and colleagues estimate that about half of the U.S. population followed the coverage (Blendon, DesEoches, Brodie, et al., 2002), and the report was thought to be the most closely followed health policy story of the year (Altman, Clancy, Blendon, 2004). Indeed, the IOM report prompted the Federal Government to appropriate \$50 million annually for patient safety research (Leape, Berwick, 2005). Further, it spurred an ongoing national conversation about patient injuries, and the subject of patient safety became a focus for journalists, health care leaders, researchers, and citizens (Leape, Berwick, 2005), fueling demands for improved health care quality, accountability, and reliability among an outraged public. A subsequent study found that 35 percent of physicians and 42 percent of the public reported experiencing a medical error in either their care or care received by a family member (Blendon, et al., 2002). Awareness of the frequency and magnitude of patient harm has helped foster widespread understanding that the domain of patient safety outcomes is a crucial component of health care quality (Mitchell, 2008). Patient safety surveys and literature show continued public concern about medical errors, with 88 percent of patients wanting to know if a medical error occurred during the course of their hospitalization, and 63 percent of the general public believing that hospital reports of errors should be made public (Gandhi, Weingart, Borus, et al., 2003; Hasnain-Wynia, Baker, 2006; Hobgood, Peck, Gilbert, et al., 2002; Weingart, Pagovich, Sands, et al., 2005; Weissman, Schneider, Weingart, et al., 2008).

Because patient safety has become an area of increasing national concern, stakeholders readily recognize the importance and usefulness of measures of patient safety. The push for quality measurement across a broad range of populations, health care settings, and aspects of care delivery stems from the notion that measures can enable patients, providers, and payers to understand what high quality care entails, recognize existing gaps in quality, and demand improvements (Hasnain-Wynia, Baker, 2006). Within pediatrics, there is a significant need to identify, quantify, and rectify AEs through targeted quality initiatives for hospitalized children. The GAPPS Trigger Tool offers the ability to detect pediatric patient harm and, in so doing, presents the possibility for devising and implementing strategies to improve outcomes.

We have tested GAPPS in a National Field Test consisting of 16 academic and non-academic hospitals across the country. Trained reviewers from the participating sites reported understanding GAPPS, and they were able to identify triggers in medical records, use them to detect AEs, and to assess severity and preventability. Next, we presented our candidate measure to our Scientific Advisory Board, consisting of representatives from Boston Children's Hospital, the larger Harvard community, and organizations such as the National Initiative for Children's Healthcare Quality, as well as to our National Stakeholder Panel, which includes representatives from diverse national organizations that represent patients and families, providers, payers, and

health services researchers. Comments and feedback from these various users and stakeholders indicate that they find the concept of a pediatric trigger tool understandable and believe such a tool is useful to pediatric medical settings.

Section 11. Health Information Technology

Please respond to the following questions in terms of any health information technology (health IT) that has been or could be incorporated into the measure calculation.

11.A. Health IT Enhancement

Please describe how health IT may enhance the use of this measure.

Our candidate measure relies on manual data abstraction from patient medical records. While the GAPPS Trigger Tool can be applied to either paper charts or EHRs, EHR use offers multiple benefits. First, EHRs are readily accessible, whereas paper records may be missing or stored in relatively inaccessible locations, such as off-site warehouses. Second, data in EHRs can be more easily subjected to validity checks to assure good data quality. EHR use also eliminates the difficulty of deciphering illegible handwriting, which may be encountered when reviewing paper records. In addition, as methods for querying EHRs become more sophisticated, more triggers may be added to the automated trigger list, and the detection of triggers on the automated trigger list may become more feasible at more hospitals. Automated trigger detection allows for more rapid and consistent identification of triggers and makes the AE detection process less labor-intensive. Manual review would only be needed for assessing the presence of AEs once triggers have already been identified, and eventually, even that process could be automated to some degree due to advances in natural language processing and in other EHR system capabilities.

11.B. Health IT Testing

Has the measure been tested as part of an electronic health record (EHR) or other health IT system?

Yes.

If so, in what health IT system was it tested and what were the results of testing?

Hospitals in the GAPPS National Field Test implemented the measure using various public and private EHR systems, including those provided through Cerner, Eclipsys, Epic, McKesson, Meditech, and Siemens. Additionally, one site implemented the measure using a homegrown EHR system. Participants indicated that GAPPS can be used successfully to detect AEs in a variety of EHRs.

11.C. Health IT Workflow

Please describe how the information needed to calculate the measure may be captured as part of routine clinical or administrative workflow.

The GAPPS measure is designed to use clinical information that is routinely recorded in medical records during the course of inpatient care, such as clinical notes, laboratory results, and

medication order histories. Use of GAPPS does not require capture of any additional data. Manual review of medical records for triggers and AEs does require some work beyond the routine clinical and administrative work involved in patient care, but the time burden (maximum review time of 30 minutes per record) is much less than the burden of conducting a comprehensive, unstructured medical record review.

11.D. Health IT Standards

Are the data elements in this measure supported explicitly by the Office of the National Coordinator for Health IT Standards and Certification criteria (see healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__standards_ifr/1195)?

Yes.

If yes, please describe.

The GAPPS measure uses data elements for which the Office of the National Coordinator for Health Information Technology (ONC) has established a set of standards and certification criteria. For example, many triggers in GAPPS rely on data regarding procedures, medications, and laboratory test results for which the ONC has adopted standard code sets, terminology, and nomenclature that are used when representing this information in EHR systems.

11.E. Health IT Calculation

Please assess the likelihood that missing or ambiguous information will lead to calculation errors.

Missing or ambiguous data within medical records may result in inaccurate estimates of AE incidence. This problem is not exclusive to the use of EHRs and might be even greater in paper records. However, many of the triggers in the GAPPS measure rely on data elements that are typically recorded with high accuracy and reliability in EHRs, such as lab results, medication orders, and information about procedures. Therefore, it is unlikely that medical records will have a sizable amount of missing or ambiguous information leading to significant calculation errors.

11.F. Health IT Other Functions

If the measure is implemented in an EHR or other health IT system, how might implementation of other health IT functions (e.g., computerized decision support systems in an EHR) enhance performance characteristics on the measure?

Incorporation of computerized decision support in an EHR system could facilitate institutional-level reductions in patient harm. For example, clinical decision support could incorporate real-time alerts regarding medication allergies, drug interactions, abnormal laboratory values, or other potential issues, thus allowing these issues to be addressed more promptly and avoiding AEs. Decision support functions can also be used to promote the delivery of guideline-adherent care. For example, Lucile Packard Children's Hospital implemented an intervention whereby a checklist of pediatric-specific central line insertion guidelines was incorporated into the EHR for each patient to help prevent central line-associated bloodstream infections (Pageler, Longhurst, Wood, et al., 2014). Furthermore, EHR functionalities, such as computerized provider order

entry, can help ensure completeness in medication prescribing fields, reducing the potential for medication errors that could lead to the occurrence of AEs.

Section 12. Limitations of the Measure

Describe any limitations of the measure related to the attributes included in this CPCF (i.e., availability of measure specifications, importance of the measure, evidence for the focus of the measure, scientific soundness of the measure, identification of disparities, feasibility, levels of aggregation, understandability, health information technology).

Reviewer and Institutional Variability

The GAPPS measure relies on manual assessment of medical records by clinician reviewers, which is an inherently subjective process. While studies suggest that experienced reviewers make fairly consistent judgments about whether harm due to health care has occurred, AE identification is vulnerable to biases that may affect outcomes measurement (Resaar, et al., 2003; Resar, et al., 2006; Sharek, et al., 2006; Takata, et al., 2008). For example, determining whether a hypotensive event is due to medication or disease progression relies on individual judgment. Reviewers' conclusions using trigger tools are also affected by training and clinical expertise, although training in trigger tool methodology eliminates some variability (Resar, et al., 2003; Rozich, et al., 2003; Sharek, et al., 2006; Sharek, et al., 2011; Takata, et al., 2008).

In addition, there are many factors that may affect how institutions recognize and record medical events, leading to variability in the information in medical records and in performance among sites (Kirkendall, et al., 2012; Rozich, et al., 2003). Internal pressures, such as a punitive culture that leads to underreporting of medical errors, may influence details in health records that are relevant to AE detection (James, 2013). These sources of variability across institutions apply to all methods of detecting harm, but trigger tools are widely available, their implementation is feasible, and they detect the largest number of AEs with the most consistent AE detection (Agarwal, et al., 2010; Classen, et al., 2011; Griffin, Classen, 2008; Rozich, et al., 2003; Sharek, et al., 2011; Sharek, et al., 2006; Takata, et al., 2008). Moreover, having a consistent, systematic approach to AE surveillance should help reduce variability.

Challenges of Assessing Preventability

GAPPS captures all AEs and assesses AE preventability. Although we include guidelines for evaluating preventability, reviewers will inevitably have different thresholds when determining preventability. In addition, perceptions of what is preventable evolve over time, so that events that are initially considered non-preventable can later be classified as preventable. Overall, preventability is controversial enough that some users of trigger tools elect not to examine it (Resar, et al., 2006; Sharek, et al., 2006). However, preventability assessment allows institutions to identify AEs that are most amenable to immediate intervention and to prioritize their quality improvement efforts. With policies like the 2008 CMS initiative to decline payment to hospitals for preventable hospital-acquired infections, determining preventability is increasingly important (Hoff, Hartmann, Soerensen, et al., 2011; Wald, et al., 2012).

Health Information Technology Restrictions

The automated GAPPS approach would improve reliability and efficiency by eliminating human error involved in finding triggers and decreasing reviewers' time burden. However, not all hospitals use EHRs, and existing EHR systems vary based on the information they contain and the document formats they use. For example, some hospitals do not document patient falls in a readily extractable manner in their EHRs.

Time Burden

Manual record review is labor-intensive and time consuming, although the automated approach would ease some of that burden (Bates, et al., 2003; Classen, et al., 2011; Cullen, Bates, Small, et al., 1995; Jha, et al., 1998; Kaiser, et al., 2014; Kirkendall, et al., 2014). Trigger tool methodology is more efficient than comprehensive record review and still is reliable in detecting harm (Agarwal, et al., 2010; Sharek, et al., 1006; Takata, et al., 2008).

Section 13. Summary Statement

Provide a summary rationale for why the measure should be selected for use, taking into account a balance among desirable attributes and limitations of the measure. Highlight specific advantages that this measure has over alternative measures on the same topic that were considered by the measure developer or specific advantages that this measure has over existing measures. If there is any information about this measure that is important for the review process but has not been addressed above, include it here.

A major focus of patient safety improvement efforts over the past 15 years has been the development and implementation of interventions to reduce health care-associated preventable AEs. A challenge in these efforts, particularly in pediatrics, has been the lack of a widely accepted, standardized method to measure AEs.

We developed the GAPPS measure to evaluate harm in all settings for hospitalized pediatric patients. The GAPPS Trigger Tool expedites identification of AEs by using red flags, or triggers, in a medical record that may indicate the presence of an underlying AE. This methodology is state-of-the-art in the field of patient safety. We compiled a draft set of triggers relevant to pediatric inpatients through a comprehensive literature review, a review of prior pediatric trigger tools, and extensive input from an expert group of diverse stakeholders using the RAND/UCLA Appropriateness Method.

We tested the measure on 3,814 medical records from a geographically distributed sample of 16 teaching and non-teaching U.S. medical centers. The GAPPS process consists of an initial medical record review by a nurse primary reviewer followed by verification of any suspected AEs by two physician secondary reviewers. To assess the measure's reliability, a second primary reviewer from each site reviewed a random 10 percent subset of the records, with secondary reviewer verification when necessary. Within hospitals, both primary reviewers agreed that a record did or did not contain at least one AE 92 percent of the time. From the total sample of records reviewed, primary reviewers identified the same AEs 62 percent of the time. Secondary reviewers within hospitals verified the same suspected AEs 92 percent of the time, with an almost perfect reliability (Kappa = 0.81). They agreed that a record did or did not contain at least

one AE 94 percent of the time. A team of external primary and secondary auditors with extensive trigger tool experience reviewed a random 10 percent subset of each site's records using GAPPS. Using the external expert reviewers' findings as the standard of comparison, the review process, as conducted by internal primary reviewers, had a sensitivity and specificity of 40 percent and 91 percent, respectively. The reliability of reviewers' findings improved significantly over time. Primary reviewers reported that they felt capable of identifying both triggers and AEs when using the tool, and that they understood the process of assessing severity and preventability after the required training sessions.

We believe that the GAPPS measure is a powerful tool that could allow an individual hospital to track patient safety and to evaluate the effects of its improvement efforts. Although the measure is more labor-intensive than voluntary reporting or administrative code-based tools, the trigger tool approach offers the advantage of a far more complete view of patient safety within hospitals. When using the GAPPS measure for quality comparisons across hospitals, a robust training protocol will help achieve consistent use of the tool, and further assessment will determine whether such training improves reliability. An external auditing process would verify that the measure is properly applied over time. In addition, case-mix adjustment to account for differences in patient populations across institutions would be beneficial. Compared with the manual approach, we anticipate that the GAPPS automated approach will increase both the efficiency and reliability of the medical record review process. This is because a computer search decreases the risk of error in identifying triggers.

References

Agarwal S, Classen D, Larsen G, et al. Prevalence of adverse events in pediatric intensive care units in the United States. *Pediatr Crit Care Med* 2010; 11(5):568-78.

Agency for Healthcare Research and Quality. Chronic Condition Indicator (CCI) for ICD-10-CM. HCUP Web site. Available at <https://hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp>. Accessed October 19, 2017.

Agency for Healthcare Research and Quality. National Healthcare Disparities Report 2010. Rockville, MD: Agency for Healthcare Research and Quality; 2011. Available at <http://www.ahrq.gov/research/findings/nhqrdr/nhqrdr10/qdr10.html>. Accessed October 19, 2017.

Ahuja N, Zhao W, Xiang H. Medical errors in U.S. pediatric inpatients with chronic conditions. *Pediatrics* 2012; 130(4):e786-93.

Allen M. How Many Die From Medical Mistakes In U.S. Hospitals? NPR Health News; 2013. Available at <http://www.npr.org/blogs/health/2013/09/20/224507654/how-many-die-from-medical-mistakes-in-u-s-hospitals>. Accessed October 17, 2017.

Altman DE, Clancy C, Blendon RJ. Improving patient safety — Five years after the IOM report. *N Engl J Med* 2004; 351(20):2041-3.

Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: The incidence of adverse events among hospital patients in Canada. *Can Med Assoc J* 2004; 170(11):1678-86.

Bates DW, Boyle DL, Vander Vliet MB, et al. Relationship between medication errors and adverse drug events. *J Gen Intern Med* 1995; 10(4):199-205.

Bates DW, Evans RS, Murff H, et al. Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003; 10(2):115-28.

Bernard D, Encinosa WE. Financial and demographic influences on Medicare patient safety events. In: Henriksen K, Battles JB, Marks ES, et al., eds. *Advances in Patient Safety: From Research to Implementation (Volume 1: Research Findings)*. Rockville (MD): Agency for Healthcare Research and Quality; 2005. Available at <https://www.ahrq.gov/professionals/quality-patient-safety/patient-safety-resources/resources/advances-in-patient-safety/index.html>. Accessed October 18, 2017.

Blendon RJ, DesRoches CM, Brodie M, et al. Views of practicing physicians and the public on medical errors. *N Engl J Med* 2002; 347(24):1933-40.

Brown B. Delphi process: A methodology used for the elicitation of opinions of experts. Santa Monica, CA: RAND; 1968, p. 1-14.

Calder L, Tierney S, Jiang Y, et al. Patient safety analysis of the ED care of patients with heart failure and COPD exacerbations: A multicenter prospective cohort study. *Am J Emerg Med* 2014; 32(1):29- 35.

Call RJ, Burlison JD, Robertson JJ, et al. Adverse drug event detection in pediatric oncology and hematology patients: Using medication triggers to identify patient harm in a specialized pediatric patient population. *J Pediatr* 2014; 165(3):447-52.

Centers for Medicare & Medicaid Services. Medicare program; Changes to the hospital inpatient prospective payment systems and fiscal year 2008 rates. *Fed Regist* 2007; 72(162):47129-48175.

Centers for Medicare & Medicaid Services. Medicare Web site. 2014. Available at <http://www.cms.gov/Medicare/Medicare.html>. Accessed October 19, 2017.

Centers for Medicare & Medicaid Services. Medicaid.gov Web site, Provider Preventable Conditions. 2011b. Available at <https://www.medicaid.gov/medicaid/financing-and-reimbursement/provider-preventable-conditions/index.html>. Accessed October 18, 2017.

Centers for Medicare & Medicaid Services. Payment adjustment for provider-preventable conditions including health care-acquired conditions. *Fed Regist* 2011a; 76(108). Available at <https://www.federalregister.gov/documents/2011/06/06/2011-13819/medicaid-program-payment-adjustment-for-provider-preventable-conditions-including-health>. Accessed October 18, 2017.

Centers for Medicare & Medicaid Services. Quality of Care Patient Safety. Web site; 2012. Available at <https://www.medicare.gov/medicaid/quality-of-care/improvement-initiatives/patient-safety/index.html>. Accessed October 18, 2017.

Classen DC, Pestotnik SL, Evans RS, et al. Computerized surveillance of adverse drug events in hospital patients. *JAMA* 1991; 266(20):2847-51.

Classen DC, Resar R, Griffin F, et al. Global Trigger Tool shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff* 2011; 30(4):581-9.

Cohen MM, Kimmel NL, Benage MK, et al. Medication safety program reduces adverse drug events in a community hospital. *Qual Saf Health Care* 2005; 14(3):169-74.

Cullen DJ, Bates DW, Small SD, et al. The incident reporting system does not detect adverse drug events: A problem for quality improvement. *Jt Comm J Qual Improv* 1995; 21(10):541-8.

Dynan L, Goudie A, Smith RB, et al. Differences in quality of care among non-safety-net, safety-net, and children's hospitals. *Pediatrics* 2013; 131(2):304-11.

Encinosa WE, Bernard DM. Hospital finances and patient safety outcomes. *Inquiry* 2005; 42(1):60-72.

Evans RS, Pestotnik SL, Classen DC, et al. Prevention of adverse drug events through computerized surveillance. *Proc Annu Symp Comput Appl Sic Med Care* 1992; 437-41.

Ferranti J, Horvath MM, Cozart H, et al. A multifaceted approach to safety: The synergistic detection of adverse drug events in adult inpatients. *J Patient Saf* 2008; 4(3):184-90.

Fitch K, Bernstein S, Aguilar MD, et al. The RAND/UCLA appropriateness method user's manual. Santa Monica, CA: RAND; 2001.

Fleiss J, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973; 33:613-9.

Flores G, Ngui E. Racial/ethnic disparities and patient safety. *Pediatr Clin North Am* 2006; 53(6):1197-1215.

Gandhi TK, Weingart SN, Borus J, et al. Adverse drug events in ambulatory care. *N Engl J Med* 2003; 348(16):1556-64.

Garrett Jr PR, Sammer C, Nelson A, et al. Developing and implementing a standardized process for global trigger tool application across a large health system. *Jt Comm J Qual Patient Saf* 2013; 39(7):292-7.

Gaskin DJ, Spencer CS, Richard P, et al. Do hospitals provide lower-quality care to minorities than to whites? *Health Aff* 2008; 27(2):518-27.

Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. *Qual Saf Health Care* 2008; 17(4):253-8.

Griffin FA, Resar RK. IHI global trigger tool for measuring adverse events (2nd Ed). Cambridge, MA: Institute for Healthcare Improvement; 2009.

Hasnain-Wynia R, Baker DW. Obtaining data on patient race, ethnicity, and primary language in health care organizations: Current challenges and proposed solutions. *Health Serv Res* 2006; 41(4 Pt 1):1501-18.

Haynes AB, Weiser TG, Berry WR, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *N Engl J Med* 2009; 360(5):491-9.

Hobgood C, Peck CR, Gilbert B, et al. Medical errors—what and when: What do patients want to know? *Acad Emerg Med* 2002; 9(11):1156-61.

Hoff T, Hartmann CW, Soerensen C, et al. Making the CMS payment policy for healthcare-associated infections work: Organizational factors that matter. *J Healthc Manag* 2011; 56(5):319-35; discussion 335-6.

Hooper AJ, Tibballs J. Comparison of a Trigger Tool and voluntary reporting to identify adverse events in a paediatric intensive care unit. *Anaesth Intensive Care* 2014; 42(2):199- 206.

James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf* 2013; 9(3):122-8.

Jha AK, Kuperman GJ, Teich JM, et al. Identifying adverse drug events: Development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Assoc* 1998; 5(3):305-14.

Joice GA, Deibert CM, Kates M, et al. “Never events”: Centers for Medicare & Medicaid Services complications after radical cystectomy. *Urology* 2013; 81(3):527-32.

Kaiser AM, de Jong E, Evelein-Brugman SF, et al. Development of trigger-based semi-automated surveillance of ventilator-associated pneumonia and central line-associated bloodstream infections in a Dutch intensive care. *Ann Intensive Care* 2014; 4:40.

Kaufman J, Rannie M, Kahn MG, et al. An interdisciplinary initiative to reduce unplanned extubations in pediatric critical care units. *Pediatrics* 2012; 129(6):e1594-600.

Kaushal R, Bates DW, Landrigan C, et al. Medication errors and adverse drug events in pediatric inpatients. *JAMA* 2001; 285(16):2114-20.

Kennerly DA, Kudyakov R, da Graca B, et al. Characterization of adverse events detected in a large health care delivery system using an enhanced global trigger tool over a five-year interval. *Health Serv Res* 2014; 49(5):1407-25.

Kennerly DA, Saldaña M, Kudyakov R, et al. Description and evaluation of adaptations to the global trigger tool to enhance value to adverse event reduction efforts. *J Patient Saf* 2013; 9(2):87-95.

Kirkendall ES, Kloppenborg E, Papp J, et al. Measuring adverse events and levels of harm in pediatric inpatients with the Global Trigger Tool. *Pediatrics* 2012; 130(5):e1206-14.

Kirkendall ES, Spires WL, Mottes TA, et al. Development and performance of electronic acute kidney injury triggers to identify pediatric patients at risk for nephrotoxic medication-associated harm. *Appl Clin Inform* 2014; 5(2):313-33.

Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: Institute of Medicine Committee on the Quality of Health Care in America; 1999.

Lander L, Roberson DW, Plummer KM, et al. A trigger tool fails to identify serious errors and adverse events in pediatric otolaryngology. *Otolaryngol Head Neck Surg* 2010; 143(4):480-6.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1):159.

Landrigan CP. The safety of inpatient pediatrics: Preventing medical errors and injuries among hospitalized children. *Pediatr Clin North Am* 2005; 52(4):979-93.

Landrigan CP, Parry GJ, Bones CB, et al. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010; 363(22):2124-34.

Larsen GY, Donaldson AE, Parker HB, et al. Preventable harm occurring to critically ill children. *Pediatr Crit Care Med* 2007; 8(4):331-6.

Leape LL, Berwick DM. Five years after to err is human: What have we learned? *JAMA* 2005; 293(19):2384-90.

Levinson DR. *Adverse Events in Hospitals: Methods for Identifying Events*. Washington, DC: Department of Health and Human Services; 2010.

Levinson DR. *Adverse Events in Hospitals: National Incidence Among Medicare Beneficiaries*. Washington, DC: U.S. Department of Health and Human Services, Office of the Inspector General; November 2010.

Matlow AG, Baker GR, Flintoft V, et al. Adverse events among children in Canadian hospitals: The Canadian Paediatric Adverse Events Study. *Can Med Assoc J* 2012; 184(13):e709-18.

McDonald KM, Davies SM, Haberland CA, et al. Preliminary assessment of pediatric health care quality and patient safety in the United States using readily available administrative data. *Pediatrics* 2008; 122(2):e416-25.

Metersky ML, Hunt DR, Kliman R, et al. Racial disparities in the frequency of patient safety events: Results from the National Medicare Patient Safety Monitoring System. *Med Care* 2011; 49(5):504-10.

Miller MR, Elixhauser A, Zhan C. Patient safety events during pediatric hospitalizations. *Pediatrics* 2003; 111(6):1358-66.

Miller MR, Elixhauser A, Zhan C, et al. Patient Safety Indicators: Using administrative data to identify potential patient safety concerns. *Health Serv Res* 2001; 36(6 Pt 2):110-32.

Miller MR, Zhan C. Pediatric patient safety in hospitals: A national picture in 2000. *Pediatrics* 2004; 113(6):1741-6.

Mitchell PH. Defining patient safety and quality care. In: Hughes RG, ed. *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*. Rockville, MD: Agency for Healthcare Research and Quality; 2008. Available at <http://www.ncbi.nlm.nih.gov/books/NBK2681/>. Accessed October 19, 2017.

Naessens JM, O'Byrne TJ, Johnson MG, et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care* 2010; 22(4):266-74.

National Quality Forum. *Serious Reportable Events in Healthcare--2006 Update*. Washington, DC: NQF; 2007. Updated report available at http://www.qualityforum.org/projects/hacs_and_sres.aspx. Accessed October 17, 2017.

Pageler NM, Longhurst CA, Wood M, et al. Use of electronic medical record-enhanced checklist and electronic dashboard to decrease CLABSIs. *Pediatrics* 2014; 133(3):e738-46.

Perla RJ, Provost LP, Murray SK. Sampling considerations for health care improvement. *Qual Manag Health Care* 2013; 22(1):36-47.

Perrin EC, Newacheck P, Pless IB, et al. Issues involved in the definition and classification of chronic health conditions. *Pediatrics* 1993; 91(4):787-93.

Piontek F, Kohli R, Conlon P, et al. Effects of an adverse-drug-event alert system on cost and quality outcomes in community hospitals. *Am J Health Syst Pharm* 2010; 67(8):613-20.

Pronovost P, Needham D, Berenholtz S, et al. An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU. *N Engl J Med* 2006; 355(26):2725-32.

Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Health Care* 2003; 12(suppl 2):ii39-45.

Resar RK, Rozich JD, Simmonds T, et al. A trigger tool to identify adverse events in the intensive care unit. *Jt Comm J Qual Patient Saf* 2006; 32(10):585-90.

Rosen AK, Loveland S, Shin M, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: The case of readmissions. *Med Care* 2013; 51(1):37-44.

Rosen AK, Geraci JM, Ash AS, et al. Postoperative adverse events of common surgical procedures in the Medicare population. *Med Care* 1992; 30(9):753-65.

Rosenthal J, Hanlon C. Nonpayment for preventable events and conditions: Aligning State and Federal policies to drive health system improvement. Washington, DC: National Academy for State Health Policy; 2009. Available at <http://www.nashp.org/sites/default/files/PatientSafety.pdf>. Accessed October 18, 2017.

Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: A practical methodology for measuring medication related harm. *Qual Saf Health Care* 2003; 12(3):194-200.

Scanlon MC, Harris JM, Levy F, et al. Evaluation of the Agency for Healthcare Research and Quality Pediatric Quality Indicators. *Pediatrics* 2008; 121(6):e1723-31.

Scott RD. The direct medical costs of healthcare-associated infections in U.S. hospitals and the benefits of prevention. Atlanta, GA: Centers for Disease Control and Prevention; 2009. Available at http://www.cdc.gov/hai/pdfs/hai/scott_costpaper.pdf. Accessed October 17, 2017.

Sharek PJ, Classen D. The incidence of adverse events and medical error in pediatrics. *Pediatr Clin North Am* 2006; 53(6):1067-77.

Sharek PJ, Horbar JD, Mason W, et al. Adverse events in the neonatal intensive care unit: Development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. *Pediatrics* 2006; 118(4):1332-40.

Sharek PJ, Parry G, Goldmann D, et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res* 2011; 46(2):654-78.

Smith RB, Cheung R, Owens P, et al. Medicaid markets and pediatric patient safety in hospitals. *Health Serv Res* 2007; 42(5):1981-98.

Stelfox HT, Palmisani S, Scurlock C, et al. The “To Err is Human” report and the patient safety literature. *Qual Saf Health Care* 2006; 15(3):174-8.

Stockwell D, Bisarya H, Classen D, et al. A trigger tool to detect harm in pediatric inpatient settings. *Pediatrics* 2015; 135(6):1036-42.

Stockwell DC, Kirkendall E, Muething SE, et al. Automated adverse event detection collaborative: Electronic adverse event identification, classification, and corrective actions across academic pediatric institutions. *J Patient Saf* 2013; 9(4):203-10.

Stone PW, Glied SA, McNair PD, et al. CMS changes in reimbursement for HAIs: Setting a research agenda. *Med Care* 2010; 48(5):433-9.

Sukumar S, Roghmann F, Trinh VQ, et al. National trends in hospital- acquired preventable adverse events after major cancer surgery in the USA. *BMJ Open* 2013; 3(6).

Sweidan M, Williamson M, Reeve JF, et al. Identification of features of electronic prescribing systems to support quality and safety in primary care using a modified Delphi process. *BMC Med Inform Decis Mak* 2010; 10(1):21.

Takata GS, Mason W, Taketomo C, et al. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in U.S. children's hospitals. *Pediatrics* 2008; 121(4):e927-35.

Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000; 38(3):261-71.

Trybou J, Spaepen E, Vermeulen B, et al. Hospital-acquired infections in Belgian acute-care hospitals: Financial burden of disease and potential cost savings. *Acta Clin Belg* 2013; 68(3):199-205.

Unbeck M, Schildmeijer K, Henriksson P, et al. Is detection of adverse events affected by record review methodology? An evaluation of the "Harvard Medical Practice Study" method and the "Global Trigger Tool." *Patient Saf Surg* 2013; 7(1):7-10.

Wald H, Richard A, Dickson VV, et al. Chief nursing officers' perspectives on Medicare's hospital-acquired conditions non-payment policy: Implications for policy design and implementation. *Implement Sci* 2012; 7:78.

Weingart SN, Pagovich O, Sands DZ, et al. What can hospitalized patients tell us about adverse events? Learning from patient-reported incidents. *J Gen Intern Med* 2005; 20(9):830-6.

Weissman JS, Schneider EC, Weingart SN, et al. Comparing patient-reported hospital adverse events with medical record review: Do patients know something that hospitals do not? *Ann Intern Med* 2008; 149(2):100-8.

Wilson RM, Runciman WB, Gibberd RW, et al. The Quality in Australian Health Care Study. *Med J Aust* 1995; 163(9):458-71.

Woods D, Thomas E, Holl J, et al. Adverse events and preventable adverse events in children. *Pediatrics* 2005; 115(1):155-60.

Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003; 290(14):1868-74.

Section 14: Identifying Information for the Measure Submitter

First Name: Mark
Last Name: Schuster, MD, PhD
Title: William Berenberg Professor of Pediatrics, Harvard Medical School
Chief of General Pediatrics and Vice-Chair for Health Policy
Department of Medicine, Boston Children's Hospital
Organization: Center of Excellence for Pediatric Quality Measurement
Mailing Address: Boston Children's Hospital, 300 Longwood Avenue
City: Boston
State: MA
Postal Code: 02115
Telephone: 617-355-5859
Email: cepqm@childrens.harvard.edu

The CHIPRA Pediatric Quality Measures Program (PQMP) Candidate Measure Submission Form (CPCF) was approved by the Office of Management and Budget (OMB) in accordance with the Paperwork Reduction Act.

The OMB Control Number is 0935-0205 and the Expiration Date is December 31, 2015.

Public Disclosure Requirements

Each submission must include a written statement agreeing that, should U.S. Department of Health and Human Services accept the measure for the 2014 and/or 2015 Improved Core Measure Sets, full measure specifications for the accepted measure will be subject to public disclosure (e.g., on the Agency for Healthcare Research and Quality [AHRQ] and/or Centers for Medicare & Medicaid Services [CMS] websites), except that potential measure users will not be permitted to use the measure for commercial use. In addition, AHRQ expects that measures and full measure specifications will be made reasonably available to all interested parties. "Full measure specifications" is defined as all information that any potential measure implementer will need to use and analyze the measure, including use and analysis within an electronic health record or other health information technology. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure. This statement must be signed by an individual authorized to act for any

holder of copyright on each submitted measure or instrument. The authority of the signatory to provide such authorization should be described in the letter.

AHRQ Publication No. 18-P001-EF
April 2018