Safety Outcomes in the United States: Trends and Challenges in Measurement

Michael D. Greenberg, Amelia M. Haviland, Hao Yu, and Donna O. Farley

Objective. To prepare Agency for Healthcare Research and Quality (AHRQ) for monitoring the impact of its own patient safety initiative, by exploring available outcomes data, assessing usability of measures, and estimating national trends in patient outcomes.

Data Sources. Annual summary data on incidence of Joint Commission Sentinel Events, MEDMARX medication error events, and MDS measures of falls and pressure ulcers in nursing home residents. HCUP National Inpatient Sample (NIS) administrative claims data.

Methods. Description and assessment of published summary data on selected safety measures. Analysis of selected Patient Safety Indicators (PSIs) and Utah–Missouri adverse event measures using HCUP NIS claims data (1994–2003).

Principal Findings. Interpretation of safety outcome trends requires close attention to the characteristics of underlying data sources and measures. Encounter-based measures have been affected by changes in definitions and ICD-9 coding, as well as by changes in the structure of administrative datasets like HCUP NIS. Historical trends are mixed for the safety outcome measures reviewed, with some measures showing improvement, others deterioration, and still others remaining fairly stable.

Conclusions. Constructing national trends of safety outcomes is difficult because of limitations in available data sources and measures. Tracking growth in the adoption of safe practices could offer an important strategy for complementing existing safety measurement capabilities.

Key Words. Patient safety, measurement, national trends, epidemiology, methods

One of the key components of any effort to track national trends in patient safety involves measuring patient outcomes. Ultimately, outcomes at the clinical level can be understood as the occurrence of adverse events that cause harm to patients. Actually measuring these sorts of adverse events, however, poses major challenges (Institute of Medicine 2000, 2004; Pronovost et al. 2005). The provision of health care in the United States involves enormous complexity, across widely ranging clinical settings, procedures, conditions,

and treatment interventions. There are many safety measures currently in use that apply to different aspects of this complex system, but the measures themselves vary considerably in the extent to which they have been validated, in the breadth and consistency of their use across facilities and regions, and in the purposes that they were designed to serve (Zhan et al. 2005). Drawing from this abundance of available measures to construct a big-picture view of patient safety in the United States is a difficult task. The task is nevertheless critical to understanding the impact of broad-based safety interventions, and to prioritizing resources and investment for patient safety in the future.

The current study involved a review of national data on safety outcomes, undertaken as part of a larger, CIPP-based evaluation of AHRQ's patient safety initiative (see Farley and Battles 2008, in this issue). Because the AHRQ safety initiative was not expected to have observable effects on national measures of safety outcomes or practices within the 4-year lifespan of our evaluation effort, our assessment of outcomes took on a developmental approach, intended to help AHRQ prepare for continued monitoring of its own safety initiative over time. In part, our aim was to explore baseline trends in outcomes on some widely available safety measures during the early years of this decade, predating any possible impacts generated by AHRQ's activities. More importantly, our aim was also to explore difficulties involved in using available patient safety measures in this way, and to examine the methodological hurdles and interpretive gaps involved in doing so. Safety measurement and data capabilities represent a fundamental part of the health care infrastructure, and a key precursor to efforts to improve safety within the system, whether on a local or national scale. Diagnosing this measurement infrastructure and its limitations, and suggesting ways to improve and better leverage it, have been basic priorities in our evaluation work for AHRQ.

It can be easy, in trying to look at national patterns in safety outcomes, either to become overwhelmed by methodological details or to reject any substantive findings outright, on grounds that the various measures are too flawed, too limited, or otherwise problematic. But for all of the legitimate criticisms of safety measures, we submit that this is an unsatisfactory answer. Patient safety is, and will remain, a major priority for health care reform and investment in the United States. Efforts are now ongoing throughout the country to try to improve safety performance in various ways, and AHRQ

Address correspondence to Michael D. Greenberg, J.D., Ph.D., RAND Health, 4570 Fifth Avenue, Pittsburgh, PA 15213, e-mail: michaelg@rand.org. Amelia M. Haviland, Ph.D., Hao Yu, Ph.D., and Donna O. Farley, Ph.D., are with RAND Health, Pittsburgh, PA.

(and others) will face the challenge of gauging the ultimate success of these efforts. Stated another way, the challenge is to make best and responsible use of imperfect measurement resources in tracking national safety progress, while simultaneously using the same information to help bootstrap measurement capabilities to higher levels of performance. This study, as well as our previous years of work in investigating clinical safety outcomes for AHRQ (Farley et al. 2005, 2007a, b), is intended to help support that end.

METHODS

Selection of Measures and Data

We began by selecting a set of clinical patient safety outcome measures for which national data were available at least for 2000–2003, guided by the following principles:

- Include measures covering a broad set of safety-relevant clinical events;
- Include measures that capture very serious adverse events;
- Cover at least some types of safety events relevant to settings beyond hospitals;
- Draw on a range of data sources, with different characteristics and limitations;
- Avoid duplicative coverage with the National Healthcare Quality Reports (NHQR) (AHRQ 2003, 2004, 2005).¹

Based on these criteria, we selected two distinct sets of safety outcome measures. The first set consisted of reporting-system measures for which aggregated data are regularly summarized by other organizations, including Joint Commission Sentinel Events, MEDMARX reports of adverse drug events, and MDS measures of falls and pressure ulcer prevalence in nursing homes. By drawing on published summaries of measures and data that were collected and analyzed elsewhere, we necessarily relied on the methods used by other agencies and organizations.

We also investigated a second set of administrative safety measures, which we constructed using a national sample of hospital encounter data from the Hospital Cost and Utilization Project (HCUP). Using administrative safety measures offers a number of analytical advantages, in part because the same measures can also be applied to other sets of administrative data (Miller and Zhan 2004; HealthGrades 2005; Sedman et al. 2005), and in part because

those measures offer a good opportunity to explore the technical challenges in building outcomes trends from source data.

Joint Commission Sentinel Events. The Joint Commission has established a set of definitions and guidelines concerning “Sentinel Events,” which are unexpected occurrences of death or serious injury, or risk thereof, within health care facilities. The Commission accredits hospitals throughout the United States and encourages them to report a subset of “reviewable” Sentinel Events back to the Commission. Since 1995, the Joint Commission has published annual summary statistics describing Sentinel Events reported to it.

MEDMARX Adverse Drug Events. MEDMARX is a voluntary reporting system for adverse medication events, which is operated by the U.S. Pharmacopeia (USP) through its Center for the Advancement of Patient Safety. MEDMARX employs the National Coordinating Council for Medication Error Reporting and Prevention classification scheme (NCC-MERP) for describing adverse drug events, and it tracks near misses, adverse events that do not harm patients, and adverse events that do harm patients. USP has periodically published reports from 1999 to 2003 that describe aggregated MEDMARX data and summarize broad trends in medication safety outcomes (Hicks et al. 2004).

Nursing Home Safety Outcome Measures. The Centers for Medicare and Medicaid Services (CMS) require all certified long-term care facilities in the United States to provide periodic reports on each of their residents, using the “Minimum Data Set” (MDS) data elements. CMS makes aggregated MDS data available for purchase on an annual basis and also publishes quarterly reports summarizing national and state-level trends on a number of MDS clinical measures (which we used in our analyses here). Two of the measures in the Medicare MDS reports—the prevalence of falls and the incidence of pressure ulcers—are particularly relevant to patient safety. Both measures reflect injuries that residents may experience as a result of inadequate vigilance or attention in nursing facility care.

Safety Outcome Measures Based on Discharge Data

We used the National Inpatient Sample (NIS) of the HCUP hospital discharge records to estimate annual, national rates of occurrence for selected safety measures that could be constructed using encounter data. These measures

included some of AHRQ's Patient Safety Indicators (PSIs) (AHRQ 2007) and some of the measures developed by the states of Utah and Missouri in AHRQ-funded patient safety projects (Utah Department of Health 2002). Notably, the PSIs have also been incorporated by AHRQ in its NHQR process.

Both sets of measures were originally generated through expert consensus processes, although the specifications for constructing the PSIs are more technically detailed than those for the UT-MO measures. In addition, the PSI specifications identify denominators that define unique at-risk populations for each measure, whereas the denominators for the UT-MO measures correspond either to all hospitalized patients or all surgical patients.

By leveraging administrative encounter data, analyses of safety outcomes using the NIS capitalize on a nationally representative sample, and on a data collection mechanism that is effectively mandatory throughout the health care system. On the downside, administrative data are subject to variations in local coding practices and are collected for purposes primarily unrelated to clinical outcomes tracking. In consequence, administrative measures of patient safety offer an important, but limited, perspective on nationwide safety outcomes.

We undertook our own independent analyses of PSIs for several reasons. First, we wanted to examine multiyear trends in annual incidence rates for the PSIs, which AHRQ had not reported in the NHQRs. Second, to explore the robustness of PSI estimation methodology, we sought to develop independent PSI estimates that replicated AHRQ's estimates, before analyzing time trends. Again, one of our chief aims was to identify and consider technical challenges involved in estimating multiyear trends, using the PSIs.

We ultimately chose a total of 13 measures to test and trend. These included five PSIs (*Postoperative Hemorrhage or Hematoma, Death in Low Mortality DRGs, Postoperative Hip Fracture, Postoperative Pulmonary Embolism or Deep Vein Thrombosis, Failure to Rescue*). They also included eight UT-MO patient outcome measures (*Accidental Cut During Procedure, Respiratory Arrest, Poisoning by Medication, Complications Affecting Specific Body Systems, Alterations in Mental Status, Accidental Falls, Other Complications of Procedures, Other Misadventures of Medical and Surgical Care*).

We selected measures taking into consideration (1) clinical judgments from physician colleagues about the importance of each measure, and the strength of its association with clinical error, and (2) available empirical information about the validity and reliability of each measure. For the PSIs, we considered the technical literature describing the PSIs (McDonald et al. 2002), and we discarded measures that were flagged as rare, having unclear

preventability, coding problems, or low sensitivity. For the UT–MO measures, we considered findings from two empirical validation studies of the UT–MO measures (Tuinen et al. 2005; Hougland et al. 2006). Notably, the Utah–Missouri set includes several measures relating to specific categories of poisoning by medication: For our purposes, we combined all of those categories into a single aggregate measure of medication poisonings.

We computed trends for our five selected PSI measures, using the HCUP NIS data for the period from 1994 to 2003. Further, we also calculated the selected PSIs using two different version of the official AHRQ algorithms for the measures (the “Old Code” version of 2004 and the “New Code” version of 2006). AHRQ typically revises its PSI definitions and algorithms each year, in part to incorporate annual updates to the ICD-9 diagnostic codes and DRG codes (which define key data elements used in calculating PSI rates). In our analyses, we applied both versions of the PSI algorithms in calculating outcome rates, to assess the extent to which annual changes in PSI coding practice affect observed outcome rates.

We also computed outcome trends for the selected UT–MO measures using the HCUP NIS data for 1994–2003. In 2003, AHRQ changed the way that ICD-9 E-codes (which denote external causes of injury) are incorporated in the HCUP NIS dataset. Because the E-codes are used for computing rates for several of the UT–MO safety measures, this change created some challenges for computing the measures. Previously, the HCUP NIS dataset had associated E-codes with primary and secondary diagnosis fields, but the dataset in 2003 ceased to make those associations. The distinction between primary and secondary diagnoses in claims data are particularly important for computing safety measures, because a primary diagnosis presumably reflects the medical condition for which a person was admitted to the hospital, and by inference, cannot reflect an injury sustained during the course of a hospitalization.

Given the ambiguity of the 2003 HCUP NIS E-code definitions, we calculated two separate 2003 rates for each of the UT–MO measures. The first rate was an upper-bound estimate, for which we assumed that all observed E-codes reflected secondary diagnoses, and the second was a lower-bound estimate, for which we assumed that the first observed E-code for each hospitalization was a primary diagnosis. Here again, our twofold aim was to construct estimates for national trends on the UT–MO measures, and to explore the impact of structural changes in HCUP data on the calculation of the measures. For purposes of this study, note that all of the UT–MO measures are computed as rates, based either on the denominator of the entire

population of persons in the HCUP NIS dataset who received medical services, or on the population of persons who received surgical care (for the two UT–MO measures relating to surgery).

RESULTS

Joint Commission Sentinel Events

As of June 2006, the Joint Commission had reviewed 3,811 Sentinel Events because it began tracking them in 1995. Of these, > 80 percent occurred in general hospital or psychiatric hospital settings, and about 73 percent of the events resulted in the death of a patient (Joint Commission 2006). The frequency of Sentinel Events rose over the decade, with approximately 150 events reported in 1997, approximately 350 reported in 2000, and approximately 580 reported in 2005 (Joint Commission 2006).

The Joint Commission also publishes a more detailed breakdown of Sentinel Events data by type of event. Figure 1 shows trends in the occurrence of the four most frequent categories of Sentinel Events, collectively comprising about 50 percent of all such events reported. These categories include patient suicides in hospital settings, operative or postoperative complications, wrong-site surgeries, and medication errors. In 2005, the frequency of reported events declined somewhat for medication errors and increased for the other three types of Sentinel Events. Notably, wrong-site surgeries and operative or

Figure 1: Numbers of Sentinel Events Reported to the Joint Commission, for Top Four Types of Events, 1995–2005

postoperative complications both reached new highs in 2005, consonant with the increasing frequency for Sentinel Events overall.

Data on Joint Commission Sentinel Events are striking in showcasing both the promise, and the limitations, of safety reporting systems more generally. Notably, Sentinel Events correspond to some of the most serious iatrogenic incidents in health care, and related data (including root cause analyses) can offer some insight into the factors that give rise to these events, as well as to changes in their occurrence over time. However, Sentinel Events data also have some major drawbacks as an epidemiological tool. In particular, it is likely that the frequency statistics for Sentinel Events are underestimates, because of incomplete reporting. This is a limitation for any external reporting system, and particularly so for voluntary reporting in the absence of specific legal protections against discovery. In addition, because health care facilities establish their own definitions for what constitutes a Sentinel Event, there is also likely to be institutional inconsistency in how such events are classified.

The MEDMARX Reporting System for Medication Events

Figure 2 shows trends in the aggregate number of medication errors and potential errors captured by the MEDMARX system, from 1999 to 2003. The trends show rapidly increasing numbers of potential medication errors, actual errors, and errors involving patient harm, over the time period. These results basically reflect growth in the number of participating facilities, from 56 facilities in 1999 to 570 facilities in 2003 (Hicks et al. 2004). In every year during the period, the number of medication errors involving harm to patients

Figure 2: Frequency of Medication Events Reported to MEDMARX, by Type of Event, 1999–2003

has been very small relative to the number of errors without harm (Hicks et al. 2004).

More important for monitoring purposes are the trends in reported medication errors captured by MEDMARX, on an average-per-facility basis. In general, the frequency of reported medication errors per facility *involving no harm to patients* trended upwards from 2000 to 2003, from approximately 229 such events per facility reported in 2000, to approximately 341 such events per facility reported in 2003. By contrast, the frequency of MEDMARX events that *did involve harm to patients* actually trended downwards over the same interval, from approximately 7.5 events per facility reported in 2000, to approximately 6.2 events per facility reported in 2003.

Here again, MEDMARX is a voluntary reporting system, and its data are subject to a number of corresponding limitations, including the potentials for underreporting and inconsistent reporting (although in this regard, the confidentiality of the MEDMARX system limits the institutional incentives toward reporting bias). Further, the MEDMARX sample of subscribing facilities is not representative of the United States as a whole, and this sample has in fact grown and changed over time. These caveats acknowledged, the MEDMARX sample nevertheless offers an important window onto the occurrence of adverse medication events in hospitals.

Minimum Data Set Measures for Nursing Home Care (MDS)

Our outcome analysis tracked two MDS measures—the prevalence of falls, and the incidence of pressure ulcers in nursing homes. The national rates of both falls and pressure ulcers among nursing home residents have remained stable since late 2001. The rate of falls has been between 12 and 13 percent of residents each quarter, and the rate of pressure ulcers has been between 10 and 11 percent of residents per quarter (Centers for Medicare and Medicaid Services 2007). These trends underscore the fact that falls and pressure ulcers affect a very significant number of nursing home residents each year, and that no broad changes in the frequency of these events have been detected in the MDS tracking mechanism in recent years.

MDS data on patient safety are particularly interesting for several reasons. At heart, MDS involves an administrative framework under Medicare, rather than an adverse event reporting system. On the plus side, this means that MDS involves a national, mandatory data pipeline, and that MDS safety data should be representative of the United States as a whole. Even more important, MDS captures data from one component of the long-term care

setting (nursing homes), and thereby addresses one of the gaps in current resources for measuring safety outcomes. On the minus side, MDS potentially involves some of the drawbacks both of reporting systems data (e.g., in the potential for systemic bias toward underreporting) and of administrative data (e.g., in the potential for inconsistent local coding practices). A number of existing research studies have suggested that MDS coding practices are inconsistent across facilities and states (Mor et al. 2003a, b, 2004; Roy and Mor 2005)—a finding that may limit the accuracy of national outcome statistics based on MDS data, and that demands caution in any comparative analysis of MDS statistics across smaller geographic regions.

Safety Outcome Measures Based on Hospital Discharge Data

Trends for Selected PSI Measures. Findings from our trend analyses for the selected PSIs are presented in Figure 3. Two different trend lines for rates are shown, based on two different versions of the official AHRQ algorithms for the measures (the Old Code version of 2004 and the New Code version of 2006). Again, we generated the two trend lines to assess the impact of annual changes in PSI definitions on observed outcomes.

Two of the selected PSIs, *Postoperative Hip Fracture* and *Postoperative PE or DVT*, showed fairly flat occurrence rates over the 1994–2003 interval. By contrast, the rate for *Failure to Rescue* showed a downward trend over the interval, as did the rate for *Postoperative Hemorrhage or Hematoma*. Two of the PSIs, *Failure to Rescue* and *Postoperative PE or DVT* occur at far greater (population-adjusted) rates than do the other three PSI measures. In none of these four cases was there much difference between the Old Code and New Code versions of the PSIs.

By contrast, the PSI *Death in Low Mortality DRGs* reflected a striking difference in its trend across the Old Code and New Code versions of the indicator. Using the New Code version of *Death in Low Mortality DRGs*, rates remained stable in 2002 and 2003, but they spiked sharply upwards during those 2 years when we used the Old Code version of the indicator. This spike for the Old Code version was due in large part to changes in observed frequencies for the underlying ICD-9 codes and DRGs used to calculate rates for the indicator.

Trends for Selected UT–MO Measures. Estimated trends for the UT–MO measures are shown in Figure 4. Differences in the upper-bound estimates for 2003 (marked by triangles) and lower-bound estimates (marked by

Figure 3: Trends for Selected PSI Measures, 1994–2003

Note: The “Old Code” refers to PSI Version 2.1 (revision 2), October 2004, while the “New Code” refers to PSI Version 3.0, February 2006.

diamonds) show the effect of the change that AHRQ made in 2003 to the way that E-codes are incorporated into the HCUP NIS dataset.

In general, rates of *Alterations in Mental Status* and *Other Complications of Procedures* showed the strongest increasing trends over the 1994–2002 interval, while *Respiratory Arrest* and *Complications Affecting Specified Bodily Systems* showed decreasing trends over the same interval. The findings suggest patterns of improvement in some patient safety measures, and deterioration in others, over the period. As with the PSI trends, some of the UT–MO measures (e.g., *Accidental Falls*) show far more frequent rates of occurrence than do others (e.g., *Accidental Cuts*).

For 2003, the upper- and lower-bound estimates for five of the eight UT–MO measures are very similar and appear to represent continuations of previous trends over time.² For three of the UT–MO measures (*Accidental*

Figure 4: Trends for Selected UT–MO Measures, 1994–2003

Falls, Accidental Cuts, and Poisonings by Medication), however, the upper- and lower-bound estimates for 2003 are quite different. We suspect that the *upper-bound* estimates may be more accurate for accidental falls and accidental cuts, in that those estimates result in smooth continuation of existing trends. The trends also reflect the likelihood that few patients are admitted to hospitals with accidental falls or accidental cuts listed as the primary diagnosis. By contrast, it seems likely that the *lower-bound* estimate is

more plausible for poisonings by medication, reflecting a greater likelihood that patients are sometimes admitted to hospitals with a primary diagnosis of medication poisoning.

Impact of Data and Measures Changes on Safety Outcomes. Perhaps the most interesting observation that comes from our work with the PSIs and UT-MO measures is the simple fact that there have been major, documented changes in the structure of the NIS dataset and the defining algorithms for the PSI measures. These changes have had a material impact on the calculation of these safety measures over time, and in some instances, the changes potentially render earlier summary statistics using the measures noncomparable to later summary statistics. They also illustrate the substantial technical complexities involved in using administrative data for safety purposes: complexities that may not be apparent purely from reviewing national summary statistics or trends themselves. We reflect on some implications in the discussion that follows.

DISCUSSION

The findings described above reflect the reality that patient safety is a complex phenomenon that crosses multiple clinical settings, therapeutic interventions, and types of risk to patients. Consequently, efforts to track national trends in safety outcomes need to tap a range of measures and different data sources. Similar to results reported in the NHQRs, we generally observe that recent historical trends in national safety outcomes have been mixed: Some adverse outcomes have become more frequent over time, others less so, while still others appear to have remained fairly stable. Interestingly, we see at least one example of convergent results across different safety measures: both MEDMARX harmful medication events (on an average per-facility basis) and UT-MO medication poisoning events appear to have declined in frequency over recent years of available data.

Any interpretation of the clinical meaning of these results is challenging and depends on underlying observations about the nature of the datasets and measures that are involved. Notably, reporting systems data and measures like Sentinel Events and MEDMARX have the virtue of being designed to capture safety outcomes, but they exist in a world where adverse events have negative repercussions for providers and institutions, and where reporting is at best likely to be incomplete. Moreover, these reporting systems were not formed with the primary function of supporting national epidemiology (e.g., by

incorporating well-defined population denominators). In consequence, we view these sorts of reporting measures as offering a trace on the occurrence of important categories of safety events, even though they are ill suited for monitoring changes in outcome rates over time. Finally, we note that changes in reporting system measures likely reflect a combination of actual changes in the frequency of adverse events, as well as in the fidelity of the reporting systems that capture them. We are unable to distinguish between these two types of effects given the data here—another reason why reviewing administrative outcome measures is important as a complement.

Our analysis of trends for selected PSIs and UT–MO measures also reveals a series of technical and interpretive challenges. In particular, periodic changes in PSI definitions mean that earlier computations of the safety indicators may not match later computations of the same indicators, even when using the same years of data. Similarly, changes made to the structure of the HCUP NIS dataset in 2003 had the effect of making it impossible to compute unambiguously some other measures of major safety outcomes, such as medication poisoning events. These kinds of challenges highlight the importance of careful documentation in presenting administrative trends data, regarding exactly how measures are computed and adjusted. They also illustrate the fact that periodic updates to measurement definitions and dataset architectures can have a substantial impact on computation (and observed trends) for safety measures: a point that deserves careful consideration by AHRQ, both in protecting the robustness of its administrative data, and in avoiding any misinterpretation of outcomes trends.

A significant implication of our review of trends in patient safety outcomes is that currently available measures are fundamentally limited in their ability to generate a composite national picture. Notably, several of the major limitations we have discussed here also apply to measures included in AHRQ's NHQR effort. The results of our analyses lead us to conclude that, to be maximally effective, the future monitoring process for AHRQ's safety initiative should track trends both in patient outcome measures and in the implementation of evidence-based safe practices. As growth in the adoption of safe practices is observed over time, one would infer that those practices are likely to lead to improved patient outcomes, some of which may be difficult to detect through conventional outcome measures.

Beyond the above, we would highlight three additional issues that should be central considerations, for AHRQ and others, in refining a monitoring system to assess safety outcomes trends on a national basis:

- *Sources of patient safety outcome data are currently limited, and coverage is skewed toward hospital settings.* Both our own work, and the NHQRs, have

been limited in this regard: New data resources and measurements covering ambulatory and long-term care settings are strongly needed.

- *Selection of safety measures on which to focus involves difficult judgments.* Formal expert consensus processes may be the best way to settle on a limited number of key measures for any future summary national tracking efforts.
- *Interpretation of safety measures depends on consistency of data definitions and coding, and characteristics of data collection mechanisms.* An important way to address these sorts of technical issues is to present them to readers, together with trends in outcomes, in order clearly to qualify any interpretation of results.

In sum, given the respective strengths and limitations of the various measures and data sources now available to AHRQ, the agency's challenge for the future will be to make best use of imperfect resources in tracking national safety performance, while simultaneously drawing on those resources as a platform: namely, to refine and build new and better measurement capabilities in the future.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was funded by the Agency for Healthcare Research and Quality (AHRQ) under contract No.290-02-0010. We would like to acknowledge Scott Ashwood for his assistance and support.

Disclosures: None.

Disclaimers: The research described in this manuscript was undertaken as one part of a larger evaluation project funded under contract with AHRQ. Pursuant to that contract, AHRQ had the right to review and comment on this paper prior to its publication.

This manuscript was developed with data collected and/or analyzed under contract with AHRQ. The information and opinions expressed herein reflect solely the position of the authors. Nothing herein should be construed to indicate AHRQ support or endorsement of its contents.

NOTES

1. The NHQRs tend to focus on only a small number of safety measures in any given year, drawing both on administrative data and the Patient Safety Indicators (PSIs), and

also the National Nosocomial Infections Survey (NNIS), the Medicare Expenditure Panel Survey (MEPS), and the Medicare Patient Safety Monitoring System (MPSMS).

2. These five measures are not based on the use of E-codes or use E-codes only minimally.

REFERENCES

- AHRQ. 2003. *2003 National Healthcare Quality Report*. Rockville, MD: AHRQ, 201pp.
- . 2004. *2004 National Healthcare Quality Report*. Rockville, MD: AHRQ, 104pp.
- . 2005. *2005 National Healthcare Quality Report*. Rockville, MD: AHRQ, 88pp.
- . 2007. *AHRQ Quality Indicators — Guide to Patient Safety Indicators, Version 3.1*. Rockville, MD: AHRQ, 70pp.
- Centers for Medicare and Medicaid Services. 2007. “MDS Public Quality Indicator and Resident Reports” [accessed on February 27, 2007, 2007]. Available at http://www.cms.hhs.gov/MDSPubQIandResRep/03_qireports.asp#TopOfPage
- Farley, D. O., and J. B. Battles. 2008. “Evaluation of the AHRQ Patient Safety Initiative: Framework and Approach.” *Health Services Research*, DOI: 10.1111/j.1475-6773.2008.00931.x
- Farley, D. O., C. L. Damberg, M. S. Ridgely, M. E. S. Sorbero, M. D. Greenberg, A. M. Haviland, R. C. Meili, S. S. Teleki, L. Bradley, J. W. Dembosky, A. Fremont, T. K. Nuckols, R. Shaw, S. L. Taylor, and H. Yu. 2007a. *Assessment of the AHRQ Patient Safety Initiative: Focus on Implementation and Dissemination, Evaluation Report III (2004–2005)*. Santa Monica, CA: RAND Corporation.
- Farley, D. O., S. C. Morton, C. L. Damberg, A. Fremont, S. H. Berry, M. D. Greenberg, M. E. S. Sorbero, S. S. Teleki, K. Ricci, and N. Pollock. 2005. *Assessment of the National Patient Safety Initiative: Context and Baseline Evaluation Report*. Santa Monica, CA: RAND Corporation.
- Farley, D. O., S. C. Morton, C. L. Damberg, M. S. Ridgely, A. Fremont, M. D. Greenberg, M. E. S. Sorbero, S. S. Teleki, and P. Mendel. 2007b. *Assessment of the AHRQ Patient Safety Initiative: Moving from Research to Practice, Evaluation Report II (2003–2004)*. Santa Monica, CA: RAND Corporation.
- HealthGrades. 2005. *HealthGrades Quality Study: Second Annual Patient Safety in American Hospitals Report*. Golden, CO: Health Grades, 24pp.
- Hicks, R. W., J. P. Santell, D. D. Cousins, and R. L. Williams. 2004. *MEDMARX 5th Anniversary Data Report: A Chartbook of 2003 Findings and Trends 1999–2003*. Rockville, MD: USP Center for the Advancement of Patient Safety.
- Houglund, P., W. Xu, S. Pickard, C. Masheter, and S. D. Williams. 2006. “Performance of International Classification of Diseases, 9th Revision, Clinical Modification Codes as an Adverse Drug Event Surveillance System.” *Medical Care* 44 (7): 629–36.
- Institute of Medicine. 2000. *To Err Is Human: Building A Safer Health System*. Edited by L. T. Kohn, J. M. Corrigan, and M. S. Donaldson. Washington, DC: National Academy Press.
- . 2004. *Patient Safety: Achieving a New Standard for Care*. Washington, DC: The National Academies Press.

- Joint Commission. 2006. "Sentinel Event Statistics as of June 30, 2006" [accessed on February 27, 2007, 2006]. Available at http://www.jointcommission.org/NR/rdonlyres/74540565-4D0F-4992-863E-8F9E949E6B56/0/se_stats_6_30_06.pdf
- McDonald, K. M., P. S. Romano, J. Geppert, S. M. Davies, B. W. Duncan, and K. G. Shojania. 2002. *Technical Review 5 – Measures of Patient Safety Based on Hospital Administrative Data – The Patient Safety Indicators*. Rockville, MD: AHRQ.
- Miller, M. R., and C. Zhan. 2004. "Pediatric Patient Safety in Hospitals: A National Picture in 2000." *Pediatrics* 113 (6): 1741–6.
- Mor, V. 2004. "A Comprehensive Clinical Assessment Tool to Inform Policy and Practice: Applications of the Minimum Data Set." *Medical Care* 42 (4, suppl): III50–9.
- Mor, V., J. Angelelli, R. Jones, J. Roy, T. Moore, and J. Morris. 2003a. "Inter-Rater Reliability of Nursing Home Quality Indicators in the U.S." *BMC Health Services Research* 3 (1): 20.
- Mor, V., K. Berg, J. Angelelli, D. Gifford, J. Morris, and T. Moore. 2003b. "The Quality of Quality Measurement in U.S. Nursing Homes." *Gerontologist* 43 Spec. No. 2: 37–46.
- Pronovost, P. J., D. A. Thompson, C. G. Holzmueller, L. H. Lubomski, and L. L. Morlock. 2005. "Defining and Measuring Patient Safety." *Critical Care Clinics* 21 (1): 1–19, vii.
- Roy, J., and V. Mor. 2005. "The Effect of Provider-Level Ascertainment Bias on Profiling Nursing Homes." *Statistics in Medicine* 24 (23): 3609–29.
- Sedman, A., J. M. Harris II, K. Schulz, E. Schwalenstocker, D. Remus, M. Scanlon, and V. Bahl. 2005. "Relevance of the Agency for Healthcare Research and Quality Patient Safety Indicators for Children's Hospitals." *Pediatrics* 115 (1): 135–45.
- Tuinen, M. V., S. Elder, C. Link, S. Li, J. H. Song, and T. Pritchett. 2005. *Surveillance of Surgery-Related Adverse Events in Missouri Using ICD-9-CM Codes*. Rockville, MD: AHRQ, 14pp.
- Utah Department of Health. 2002. "The 2002 Report on the Findings of Rating the Utah/Missouri ICD-9-CM Adverse Event Codes" [accessed on March 13, 2007, 2002]. Available at <http://health.utah.gov/psi/pubs/Expertpanel.pdf>
- Zhan, C., E. Kelley, H. P. Yang, M. Keyes, J. Battles, R. J. Borotkanics, and D. Stryer. 2005. "Assessing Patient Safety in the United States: Challenges and Opportunities." *Medical Care* 43 (3, suppl): I42–7.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.