

---

## Editorial

---

# Enhancing the Value of Statewide Hospital Discharge Data: Improving Clinical Content and Race–Ethnicity Data

Statewide hospital discharge datasets are a commonly used source of data for health services research. For nearly 25 years, the Agency for Healthcare Research and Quality's (AHRQ's) Healthcare Cost and Utilization Project (HCUP) has partnered with state governments, hospital associations, and private data organizations to transform their databases into uniform, research-oriented datasets for use in health services research. These datasets are unique as a source of data on hospital services across all payers (including the uninsured), and almost all hospitals in the United States. The state discharge datasets complement payer-specific datasets that cover multiple care settings (beyond hospitals), including Medicare, Medicaid, and private payer datasets. They also complement more recent efforts to develop clinically rich datasets from electronic health record data in specific cohorts and populations.

Statewide discharge datasets support studies on a wide range of topics, including health policy, access to and utilization of care, clinical and quality aspects of care, economics and financing, and disparities based on race–ethnicity, insurance, and income. Although widely used, these data have limitations based on restricted service settings (hospital only), limited data element content, and concerns over accuracy of some data elements. Among the limitations most often mentioned are the lack of clinical detail and variation in the availability and accuracy of patient sociodemographic information (Schoenman et al. 2007; Andrews 2015). To foster successful approaches to enhancing state data, AHRQ awarded eight 3-year grants to explore methods to improve the clinical content and race/ethnicity information in these databases (Agency for Healthcare Research and Quality 2014; Andrews 2015).

This *HSR* Special Issue, sponsored by AHRQ, focuses on the products and lessons learned from these grant projects. Seven articles cover background information on statewide hospital discharge data and the grant program aims, challenges, and lessons learned, a case study related to one of the grantees' shared challenges (continuing funding after the grant), an innovative approach for linking data, a demonstration of the value of adding laboratory data, and two different approaches to assessing and improving race–ethnicity data quality.

The initial article by Andrews (2015) provides background information and context for the Special Issue articles. The overview of content, coverage, types of databases, and uses of the statewide discharge data offers a primer for readers not familiar with these data. This article also includes information about the evolution of these data since the 1970s and the data standards issues affecting the current content and quality of the data. The review of recent articles using statewide discharge databases, including the HCUP state and national versions of these data, demonstrates that these data are an important asset for health services research. Though the examples provided are from just one journal (*HSR*) for 3 years, they illustrate the range of topics examined and the specific datasets and linkages used by health services researchers. Limitations of the data are also noted, with a focus on the clinical and race–ethnicity data limitations that prompted the AHRQ Enhanced State Data grant program. Each of the grant projects is described, providing a useful prelude for the second article (Pine et al. 2015), which describes the challenges faced by the grantees.

The article by Pine et al. (2015) is a collaborative effort by the grantees to describe the five common challenges that they faced in their projects to improve the clinical content and race/ethnicity data: recruiting data partners, ensuring effective participation of data partners, ensuring high-quality data, linking data from multiple sources, and sustaining enhancements after the grants ended. States differ in the type of organization that collects the statewide discharge data (government, hospital associations, and other private entities), the size and characteristics of the hospital industry, the current capabilities of their discharge datasets, the current uses of the discharge data, and the interests of key stakeholders in improving various dimensions of their data. Given these variations, Pine and colleagues describe some typical approaches to address these challenges but also provide examples of unique approaches that addressed issues specific to certain states. These descriptions may be of interest both to those involved in statewide discharge data systems and to health services researchers involved in other types of projects involving data partnerships, linkages, and project continuity following grant funding.

The article by Salemi et al. (2015) is a case study of how one grant project dealt with one of the challenges mentioned by Pine et al. (2015)—sustainability of the improved data after the grant funding ended. Salemi and colleagues described how the enhanced data they created contributed to local health decision making as part of a community-based participatory research project concerning disparities in maternal and child health. Florida's grant project involved linking discharge data for newborns and their mothers with several datasets related to maternal and child health and/or the hospital stay. As part of the community-based participatory research, the enhanced data were used in conjunction with focus group and survey data to identify priority areas for intervention. As the authors discuss, use of the enhanced data in this way garnered community support for continued enhancements, multiyear funding to continue the enhancements, and thus a pathway for sustaining the project's accomplishments. These enhanced data should therefore be available in the future to support both the health information needs of the community and studies on maternal and infant health services.

Researchers often link statewide hospital discharge datasets to other data to enhance their usefulness (Zingmond et al. 2004; Bradley et al. 2010; Andrews 2015). These linkages may use hospital identifiers, physician identifiers, patient zip code, or person identifiers. One common feature of all of the Enhanced State Data grant projects was linking the discharge data for individual patients to one or more outside datasets, either to augment the clinical content of the data (e.g., laboratory data or prehospital emergency services data) or to evaluate the accuracy of race and ethnicity data compared with self-reported information. The article by Naessens et al. (2015) explores opportunities for such person-level linkages using statewide discharge datasets when personal identifiers are not present in the data. Although personal identifiers, such as Social Security number (SSN) or name, are becoming more widely available, more than half the states still do not collect them, generally due to concerns about maintaining patient confidentiality or preventing identity theft. Naessens and colleagues provide evidence to support an alternative for linking patients for readmissions and vital statistics data using patient sex, date of birth, zip code, and the last four digits of the SSN. This approach yielded nearly perfect matching when compared to an algorithm using unique identifiers from a set of Minnesota hospitals. An algorithm using just three data elements that all states collect—patient sex, date of birth, and zip code—did reasonably well, but it was not as accurate as including the last four digits of the SSN. As the authors acknowledge, the tested approaches may not work as well in other states that differ on such characteristics as the population size of

zip codes. The type of linkage may matter as well; Naessens et al. focused on readmissions within Minnesota and with in-state death records. Thus, while their approach is promising, further testing would be advised prior to broader adoption.

The article by Lim et al. (2015) was based on the addition of laboratory data to Hawaii's statewide inpatient discharge data, and it adds to the evidence on the value of these data to increase the accuracy of risk adjustment for hospital-level reporting. Using data for heart failure and acute myocardial infarction patients, they examined whether a simple count of abnormal laboratory values would improve a widely used methodology (APR-DRG) to risk adjust for hospital-level mortality. In discussing their findings, Lim and colleagues describe some of the methodological issues that will confront health services researchers as they integrate laboratory results in their analyses. More sophisticated approaches, beyond a simple count of abnormal values, would incorporate the level and direction of abnormality and weight laboratory findings based on their contribution to risk. As laboratory results become more accessible through expanded electronic medical record systems and as statewide data organizations capture more clinical data, research will likely focus on how to take advantage of these data. As happened with ICD-9-CM coded diagnosis and procedure data, such research will likely yield analytic tools and standard methods that will facilitate wider use.

Lim et al. also tested the impact of adding race to the risk-adjustment model, which makes their article relevant to current policy and methodological discussions of how to handle differences in racial-ethnic mix across hospitals when measuring quality for hospital-level pay-for-performance or reporting efforts (Fiscella, Burstin, and Nerenz 2014; Jha and Zaslavsky 2014; Lipstein and Dunagan 2014). If there are unmeasured clinical and social factors beyond the hospital's control that affect outcomes of care, and if these factors vary by race-ethnicity, then hospitals that serve a higher proportion of minority patients could be unfairly penalized without adjustment for race-ethnicity (or these clinical or social factors). On the other hand, many publications have shown disparities in the care received by minorities (Mitchell et al. 1997; Institute of Medicine 2002; Agency for Healthcare Research and Quality 2015), so using race-ethnicity as a risk adjustor could mask disparities in the quality of care. The authors' finding of a modest racial effect suggests that either (1) unmeasured clinical or social factors are captured by race—even after controlling for laboratory values and other clinical information in the administrative data or (2) disparities in quality of care lead to differences in mortality across racial groups. These findings should be corroborated in other

states and with more sophisticated use of laboratory data to measure risk (beyond the simple count of abnormal values).

The article by Lim et al. also illustrates the value of collecting race and ethnicity using more granular categories than the standard OMB race–ethnicity categories. The Institute of Medicine (2009) has recommended use of more detailed categories, such as those in Hawaii’s data, but few states currently do so. Lim et al. found different inpatient mortality patterns for the different subgroups studied (Chinese, Filipino, Hawaiian, Japanese, other Pacific Islander) when compared to whites.

As Andrews (2015) notes in the opening article, almost all states now collect race and ethnicity on their statewide discharge data. However, questions remain about the accuracy of the coding of this information. The special issue includes articles from two of the grant projects that focused on evaluating and improving race and ethnicity data. Zingmond et al. (2015) developed and tested candidate audit measures to estimate the accuracy of hospital-level race–ethnicity data reporting. Although it would be ideal to use self-reported data as the gold standard against which to judge accuracy of hospital-reported race–ethnicity, statewide data organizations often cannot do so because of the costs of obtaining and analyzing these types of data, difficulties in accessing self-report data on a regular basis, and patient privacy issues. Their aim was to develop simpler methods that would be more feasible for statewide data organizations to implement. In essence, their method involves comparing the distribution of racial–ethnic groups in an individual hospital with the racial–ethnic distribution of the hospital’s catchment area (based on a proportional weighting of the zip codes in which the patients reside). They demonstrate a moderate correlation of this proxy measure of accuracy to “gold standard” approaches (discharge record comparisons to self-report on birth certificates and cancer registries). One shortcoming of the proxy approach is an underlying assumption that the hospitalization rates are equal across racial groups. Taking this into consideration, it may be possible for statewide data organizations to implement the proposed auditing approach using rules of thumb to identify patterns that are highly unexpected based on knowledge of the local population and their health care patterns. Thus, the approach offers statewide data organizations something beyond the common race–ethnicity data edit checking practice of looking solely at percent missing or coded “other.” In states that collect additional information about the patient’s residence beyond zip code, for example, patient address or nine-digit zip code, the approach could be refined for smaller geographic areas, likely resulting in more accuracy for auditing.

In addition to providing an assessment of this new auditing method, the article provides evidence about the quality of race and ethnicity coding on California discharge data for two patient cohorts—mothers delivering newborns and hospitalized cancer patients. In both cohorts, the American Indian/Alaska Native (AIAN) patients have the worst underreporting problem, as demonstrated by very poor specificity (only 10–30 percent of AIAN patients were coded as AIAN in the hospital data). This finding is consistent with other research that indicates that the AIAN coding is substantially undercoded (Korenbrodt, Ehlers, and Crouch 2003) in statewide discharge data, including this special issue's article focused on AIAN coding in Oregon and Washington, described below (Bigback et al. 2015).

Bigback et al. (2015) evaluate the extent of misclassification of AIAN patients in Oregon and Washington hospital discharge data by linking these data to the Northwest Tribal Registry. Through this linkage, they discovered that 45–55 percent of the hospitalized individuals in the Northwest Tribal Registry were miscoded as not being AIAN in the discharge data. By adding these miscoded discharges to the count of AIAN records, they increased the hospitalization rate for AIAN by about one-third in each state. Their undercount of AIAN is likely a conservative estimate, as their registry underrepresents urban AIAN populations and those who do not use IHS or tribal clinics. Nonetheless, their study adds to the evidence that there is substantial undercoding of AIAN in statewide hospital discharge data, while showing how to mitigate this undercoding through linkage to tribal registries. Furthermore, their article serves as another example of the power of data linkages to both improve discharge data and evaluate its quality.

Together, these articles provide a pathway to enhancing the value of the HCUP and statewide datasets for health services research by improving their clinical content and race–ethnicity information. These articles also highlight the challenges of improving a near-universal data source such as HCUP. HCUP and other statewide data remain an extremely useful resource for exploratory analyses; student, fellow, and trainee projects; and basic epidemiologic studies of access, hospital quality, or cost. While efforts to expand the value of the HCUP and statewide data are challenging, we hope that the articles in this issue will inspire more collaborative efforts among researchers, statewide data organizations, and hospitals to explore these and other useful advancements to ensure that these datasets remain a valuable research resource for years to come.

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* Roxanne Andrews was a senior health services researcher at AHRQ for most of her term as Guest Editor for this Special Issue. She is now retired from Federal service. The views expressed in this editorial are those of the authors and do not represent those of the AHRQ or the US Department of Health and Human Services. This Special Issue was supported by AHRQ. The manuscripts submitted for the issue have undergone the usual *HSR* peer review process. The findings and views expressed in the Special Issue articles are those of the authors and do not represent those of AHRQ or the US Department of Health and Human Services.

*Disclosures:* None.

*Disclaimers:* None.

Roxanne M. Andrews

Kevin A. Schulman

## REFERENCES

- Agency for Healthcare Research and Quality. 2014. "Data Innovations. Healthcare Cost and Utilization Project (HCUP)" [accessed on April 27, 2015]. Available at <https://www.hcup-us.ahrq.gov/datainnovations/grants.jsp>
- Agency for Healthcare Research and Quality. 2015. *2014 National Healthcare Quality and Disparities Report*. Rockville, MD: Agency for Healthcare Research and Quality. AHRQ Pub. No. 15-0007.
- Andrews, R. M. 2015. "Statewide Hospital Discharge Data: Collection, Uses, Limitations, and Improvements." *Health Services Research* 50 (4 Pt 2): 1273–99.
- Bigback, K. M., M. Hoopes, J. Dankovchik, E. Knaster, V. Warren-Mears, S. Joshi, and T. Weiser. 2015. "Using Record Linkage to Improve Race Data Quality for American Indians and Alaska Natives in Two Pacific Northwest State Hospital Discharge Databases." *Health Services Research* 50 (4 Pt 2): 1390–402.
- Bradley, C. J., L. Penberthy, K. J. Devers, and D. J. Holden. 2010. "Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future." *Health Services Research* 45 (5 Pt 2): 1468–88.
- Fiscella, K., H. R. Burstin, and D. R. Nerenz. 2014. "Quality Measures and Sociodemographic Risk Factors: To Adjust or Not to Adjust." *Journal of the American Medical Association* 312 (24): 2615–6.
- Institute of Medicine. 2002. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, edited by B. D. Smedley, A. Y. Stith, and A. R. Nelson. Washington, DC: National Academies Press.
- Institute of Medicine. 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: National Academies Press.

- Jha, A. K., and A. M. Zaslavsky. 2014. "Quality Reporting that Addresses Disparities in Health Care." *Journal of the American Medical Association* 312 (3): 225–6.
- Korenbrodt, C. C., S. Ehlers, and J. A. Crouch. 2003. "Disparities in Hospitalizations of Rural American Indians." *Medical Care* 41 (5): 626–36.
- Lim, E., Y. Cheng, C. Reuschel, O. Mbowe, H. J. Ahn, D. T. Juarez, J. Miyamura, T. B. Seto, and J. J. Chen. 2015. "Risk-Adjusted In-Hospital Mortality Models for Congestive Heart Failure and Acute Myocardial Infarction: Value of Clinical Laboratory Data and Race/Ethnicity." *Health Services Research* 50 (4 Pt 2): 1351–71.
- Lipstein, S. H., and W. C. Dunagan. 2014. "The Risks of Not Adjusting Performance Measures for Sociodemographic Factors." *Annals of Internal Medicine* 161 (8): 594–6.
- Mitchell, J. M., K. R. Meehan, J. Kong, and K. A. Schulman. 1997. "Access to Bone Marrow Transplantation for Leukemia and Lymphoma: The Role of Sociodemographic Factors." *Journal of Clinical Oncology* 15 (7): 2644–51.
- Naessens, J. M., S. L. Visscher, S. M. Peterson, K. M. Swanson, M. G. Johnson, P. A. Rahman, J. Schindler, M. Sonneborn, D. E. Fry, and M. Pine. 2015. "Incorporating the Last Four Digits of Social Security Numbers Substantially Improves Linking Patient Data from De-identified Hospital Claims Databases." *Health Services Research* 50 (4 Pt 2): 1339–50.
- Pine, M., N. M. Kowlessar, J. L. Salemi, J. Miyamura, D. S. Zingmond, N. E. Katz, and J. Schindler. 2015. "Enhancing Clinical Content and Race/Ethnicity Data in Statewide Hospital Administrative Databases: Obstacles Encountered, Strategies Adopted, and Lessons Learned." *Health Services Research* 50 (4 Pt 2): 1300–21.
- Salemi, J. L., A. A. Salinas-Miranda, R. E. Wilson, and H. M. Salihu. 2015. "Transformative Use of An Improved All-Payer Hospital Discharge Data Infrastructure for Community-Based Participatory Research: A Sustainability Pathway." *Health Services Research* 50 (4 Pt 2): 1322–38.
- Schoenman, J. A., J. P. Sutton, A. Elixhauser, and D. Love. 2007. "Understanding and Enhancing the Value of Hospital Discharge Data." *Medical Care Research and Review* 64 (4): 449–68.
- Zingmond, D. S., Z. Ye, S. L. Ettner, and H. Liu. 2004. "Linking Hospital Discharge and Death Records—Accuracy and Sources of Bias." *Journal of Clinical Epidemiology* 57 (1): 21–9.
- Zingmond, D. S., P. Parikh, R. Louie, D. Y. Lichtensztajn, N. Ponce, R. Hasnain-Wynia, and S. L. Gomez. 2015. "Improving Hospital Reporting of Patient Race and Ethnicity—Approaches to Data Auditing." *Health Services Research* 50 (4 Pt 2): 1372–89.

## SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.