Examining Health Disparities Using Data Science

The population of older adults in the United States is increasing in both size and racial and ethnic diversity. Research examining racial and ethnic disparities in care among older adults is essential to providing better quality care and improving patient outcomes. Yet, in the current climate of limited research funding, what efficient methods exist for gerontological nurse researchers to address these important health care issues among racially and ethnically diverse groups, groups typically underrepresented and difficult to access in research? We suggest data science—a rigorous, interdisciplinary philosophy and method to acquire, manage, explore, and interpret large data sets—using existing clinical and population data (Brennan & Bakken, 2015).

Clinical data sets provide insight into health trajectories across care settings and transitions. Examples such as electronic health records, the Minimum Data Set (MDS; Centers for Medicare and Medicaid Services [CMS], 2012a) from nursing homes, or the Outcomes Assessment Information Set (OASIS; CMS, 2012b) from home health, typically house information on thousands of patients across the country. Gerontological nurse researchers may consider partnering with health care agencies to sample from these larger clinical data sets to address research questions on differences in care practices, symptom science, and patient-centered outcomes across racially and ethnically diverse older adult populations. In addition, clinical data sets may contain information on health services and access to care. For example, data from the MDS and OASIS may be linked to the CMS claims data to address local and national level health services research questions. Costs to access clinical data sets may vary depending on the system level of data (e.g., local hospital data versus national Medicare data).

Population data sets also offer ample opportunity to study diverse racial and ethnic groups. These data often contain rich information on race and ethnicity as well as other characteristics interrelated with racial disparities, such as income and education. Population data sets hold important advantages. First, they are often nationally representative and their findings can be generalized to the U.S. population. Second, the large number of observations they contain often produce sufficient statistical power for a wide range of outcomes disaggregated by subpopulations. Many of these surveys oversample ethnic and racial minorities. Third, these data sources have typically been cleaned and are well documented, making them user-friendly. One noteworthy example is the Health and Retirement Study (HRS; 2012), which is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The HRS is a longitudinal survey conducted every 2 years since 1992 on the 50 and older population, which oversamples African American and Hispanic individuals. In addition to detailed demographic and health information, the HRS includes information for each of the respondent’s living children and up to seven caregivers. These data enable researchers to examine health and caregiving patterns across racial and ethnic groups. Researchers interested in studying older adults in other countries may consider using HRS international sister studies. These are surveys on aging and retirement conducted in 11 countries, including Mexico, Korea, and Ireland, which have been harmonized by the University of Southern California Center for Economic and Social Research for easy comparison.

Researchers interested in disparities in health behaviors may consider using the American Time Use Survey (ATUS; Hofferth, Flook, & Sobek, 2013). The ATUS is a cross-sectional study sponsored by the Bureau of Labor Statistics on time use among the 15 and older population in the United States. Researchers have used these data to
examine the time spent engaging in health-related behaviors including sleep, physical activity, and eldercare. Since 2011, the ATUS has implemented a separate eldercare module.

Working with large existing data sets presents unique challenges for nurse researchers. For example, clinical data sources may have limited demographic descriptors. In addition, accuracy of clinical data relies on data entry from multiple users whose primary focus may be clinical care and reimbursement, not research, resulting in potential measurement error and missing data issues. Information collected for clinical data sets may not have been initially intended for research, resulting in the need to clearly operationalize variables in the data set. Although most population databases are intended for research, they may not contain the specific variables researchers need. These databases often contain several thousands of variables, requiring greater computing power and special care in managing large data sets. Specific methodological considerations related to using large data sets include complex questionnaire skip patterns, survey weights, and multiple imputation techniques. Given these challenges, interdisciplinary collaboration is essential to ensuring the subject and methodological expertise required in data science. Furthermore, appropriate infrastructure is necessary to support potential issues with housing enormous quantities of potentially identifiable data.

Despite these challenges, using large, existing clinical and population data sets provides exceptional opportunities for gerontological nurse researchers to access under-represented groups in aging research. In capturing a large number of individuals from racial and ethnic minority groups, these data sources cast light on a rapidly growing population of increasing relevance in the United States.

REFERENCES


Jo-Ana D. Chase, PhD, APRN-BC
NRSA Postdoctoral Fellow
Center for Integrative Science in Aging

Alma Vega, PhD
Dean’s Distinguished Postdoctoral Fellow

NewCourtland Center for Transitions and Health
University of Pennsylvania School of Nursing
Philadelphia, Pennsylvania

The authors have disclosed no potential conflicts of interest, financial or otherwise.

doi:10.3928/19404921-20160404-01