

Quality of linked data: Linking the National Hospital Care Survey Data to the National Death Index

Mirel, L¹, Resnick, D², Campbell, S², and Golden, C¹

¹CDC/NCHS/OAE/SPB

²NORC at the University of Chicago

Introduction

Data linkages can produce rich data resources to address a variety of research topics. However, assessing linkage quality can be challenging given that there are many approaches and no clear best practices.

Objectives and Approach

Through its Data Linkage Program, the National Center for Health Statistics (NCHS) links national survey data with vital and administrative records. A recent linkage of the National Hospital Care Survey data with the National Death Index employed a new linkage methodology, which included a first time approach for validating the results within the linkage algorithm.

Results

The new methodology includes two passes: a deterministic linkage, followed by a probabilistic approach based on the Fellegi-Sunter methodology. In the second pass, a key identifier, Social Security Number (SSN), was not used as a linkage variable but instead to determine link accuracy, when available on the patient record. A model was then built to predict link accuracy status according to the computed Fellegi-Sunter total pair weight and then used to estimate it for those patient records without an SSN. Results from this new approach were compared with results from prior linkage methodologies and generated higher match rates and lower error rates. The linkage methodology designed for this study is now being tested on other types of input data such as data from household surveys.

Conclusion/Implications

The linkage approach may be incorporated into additional linkages conducted by NCHS. This talk will describe the input sources for this linkage, the methodology used, the error rate assessment and then discuss conclusions and implications for precision and efficiency.

