

University of Massachusetts Medical School

eScholarship@UMMS

Center for Health Policy and Research (CHPR)
Publications

Commonwealth Medicine

2014-05-16

Judy Savageau and Kathy Muhr on Working with Different Data Sources


Judith A. Savageau

University of Massachusetts Medical School

Et al.

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/healthpolicy_pp

 Part of the [Health Services Administration Commons](#), [Health Services Research Commons](#), and the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

Repository Citation

Savageau JA, Muhr K. (2014). Judy Savageau and Kathy Muhr on Working with Different Data Sources. Center for Health Policy and Research (CHPR) Publications. Retrieved from https://escholarship.umassmed.edu/healthpolicy_pp/170

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Center for Health Policy and Research (CHPR) Publications by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Working with Different Data Sources

Hi. We're Judy Savageau and Kathy Muhr from the University of Massachusetts Medical School's [Center for Health Policy and Research](#). Within our Research and Evaluation Unit, we work on a number of projects using qualitative and quantitative methods as well as primary and secondary data sources. We've come to appreciate that different types of data from different sources need varying levels of data management and quality oversight.

One of our current projects is evaluating a screening program that requires primary care providers to screen children for potential behavioral health conditions. Among a random sample of 4000 children seen for a well child visit during one of two study years, we collected data both from medical records (primary data source: both quantitative and qualitative chart notes) as well as administrative/claims data (secondary data source: solely quantitative). Given the nature of data from the two sources, we implemented different data quality checks and cross-checks between them.

Lessons Learned:

- Claims data comes from the insurance payer having already gone through its own internal data cleaning and data management processes. However, much of the patient demographic data comes at the time of insurance enrollment and not updated at the time of a clinical visit. Some data elements are often incomplete and not updated even after numerous clinical encounters, especially data such as gender, race, ethnicity and primary language. While a provider might 'know' this information when seeing a patient, it's not necessarily updated in administrative datasets.
- Many practices don't necessarily collect demographic data in a uniform manner unless they're required to report on this data. Primary care providers are well connected to their patient's demographics in terms of needs for interpreters, cultural health beliefs, and age- or gender-specific anticipatory guidance needs. Unfortunately, medical records data often had nearly as much missing data as did the administrative claims data!
- Cross-checking data between these two sources was an important step for us to take in this project as we hypothesized that there might be differences in screening children for behavioral health needs. Wanting to assess potential health service disparities was an important factor in this evaluation given the interest in vulnerable populations.
- While electronic medical records (EMRs) were evident in at least 60% of practices where charts were abstracted, it was no surprise to find that EMRs vary practice to practice. It was clear that projects such as this one might then need to use text-based data within the chart notes to obtain vital information in order to assess potential disparities.

Hot Tip: Although data quality is key, find a balance between budgetary and personnel resources and the time required to cross-check data through multiple sources and/or impute missing data using a variety of techniques.