



PAA Applied Demography Conference 2021
February 4, 2021

Workshop: Comparing Data Sources 101

Collin McCarter, Trellance

Applied demographers routinely review and depend on high quality data sources to address a wide scope of questions for decision makers. During the data review process, researchers must weigh trade-offs inherent with data sources. Trade-offs can be defined as the frequency of data updates versus the granularity of characteristics, or even the geographic relevance versus adequate sample sizes. These trade-offs, along with many others, are represented in common demographer conversations like appropriate usage of Decennial Census versus American Community Survey data. While data input characteristics are always relevant, the latest trend of real-time insights using repurposed high frequency administrative data has put added emphasis on evaluating data characteristics for many new and experimental data sources. Given that research answers are always “needed yesterday”, many decision makers seem willing to sacrifice some components of a research insight if the conclusion is delivered on their rapid time schedule. Specifically, Covid-19 spread, unemployment projections, changes in media behavior, and school choice preferences are all challenging research topics which often demand instantaneous updates from stakeholders. While stakeholders can quickly understand the benefit of instantaneous insights, domain experts are often called upon to deliver the poorly understood sacrifices of those same instantaneous insights. With these trade-offs in mind, this presentation will share three key insights.

First, I will summarize key data source characteristics which have been helpful to my work at Trellance, Nielsen, and the Florida Legislature. These include but are not limited to: provider organization; provider dataset; cost; topic; topic granularity; unit of analysis; survey or census based; record size; geographic coverage; method of collection; update frequency; sustainability; ease of access; documentation; ethical collection; ethical storage; and ethical usage.

Second, I will highlight a few experimental data sources which may be of interest to other applied demographers. These include but are not limited to: Bureau of Transportation Statistics Air Traffic Volume; mobile based migration from Teralytics; television provider set-top boxes; credit union transactions; and public school Full-Time-Equivalent (FTE) enrollments.

Third, I will add use case context to the above sections to highlight trade-offs inherent with various data sources and stakeholder requirements.