STEFFEN KÜNN
Maastricht University, Netherlands, and IZA, Germany

I Z A
World of Labor
Evidence-based policy making

The challenges of linking survey and administrative data
Combining survey and administrative data is growing in popularity, even though data access is still highly restricted

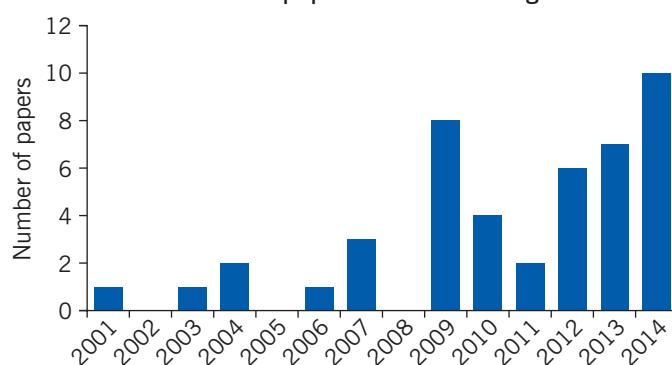Keywords: record linkage, administrative records, survey data

ELEVATOR PITCH

Using administrative records data and survey data to enhance each other offers huge potential for scientific and policy-related research. Two recent changes have expanded the potential for creating such linked data: the improved availability of data sources and progress in data-matching technology. These developments are reflected, among other ways, in the growing number of academic papers in labor economics that use linked survey and administrative data. While the number of studies using linked data is still small, the trend is clearly upward. Slowing the growth, however, are concerns about data security and privacy, which impede data access.

The use of linked administrative and survey data in labor economics papers has been rising

Source: Calculations based on author's analysis of the IZA discussion paper series.

I Z A
World of Labor

KEY FINDINGS

Pros

- Data linkage overcomes some of the shortcomings of the two separate data sources.
- Data linkage opens new research opportunities by combining highly reliable administrative data with detailed survey data.
- Administrative records, already collected routinely, are a cheap and authoritative source of data for enriching survey data.
- Data linkage can lower survey costs by requiring fewer questions.
- Data linkage enables sensitive data, such as wages, to be drawn from administrative records, reducing the burden on respondents and likely lessening survey dropout and item nonresponse rates.

Cons

- Linking data can be very costly and time-consuming, mainly because of drawn-out negotiations with data providers.
- Privacy concerns and resulting legal constraints and the need for data anonymization restrict data access and content.
- Requesting consent to use the linked data may introduce consent bias (consenters differ from non-consenters) or may reduce response rates, introducing yet another selection bias.
- Sound linkage requires a unique identifier for each individual; without such identifiers, linkage becomes burdensome and matching quality may suffer.

AUTHOR'S MAIN MESSAGE

Data linkage opens new research opportunities by combining highly reliable administrative records with detailed survey data. Researchers wishing to link the two data sources should establish that both data sources include unique personal identifiers and that the survey includes a properly worded consent request for respondents. Most important, any data security concerns of the data provider (typically a government institution) must be resolved in advance, to avoid having data security concerns lead to restrictions on access or to demands for strict anonymization of the data, which reduce its research potential.