

Respondent Understanding of Disclosure Avoidance

Aleia Clark Fobia, Jennifer Hunter Childs, Mandi Martinez, and Casey Eggleston
Center for Behavioral Science Methods, U.S. Census Bureau
Challenges and New Approaches for Protecting Privacy in Federal Statistical
Programs.

June 6-7, 2019. NASEM



U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

“Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.”

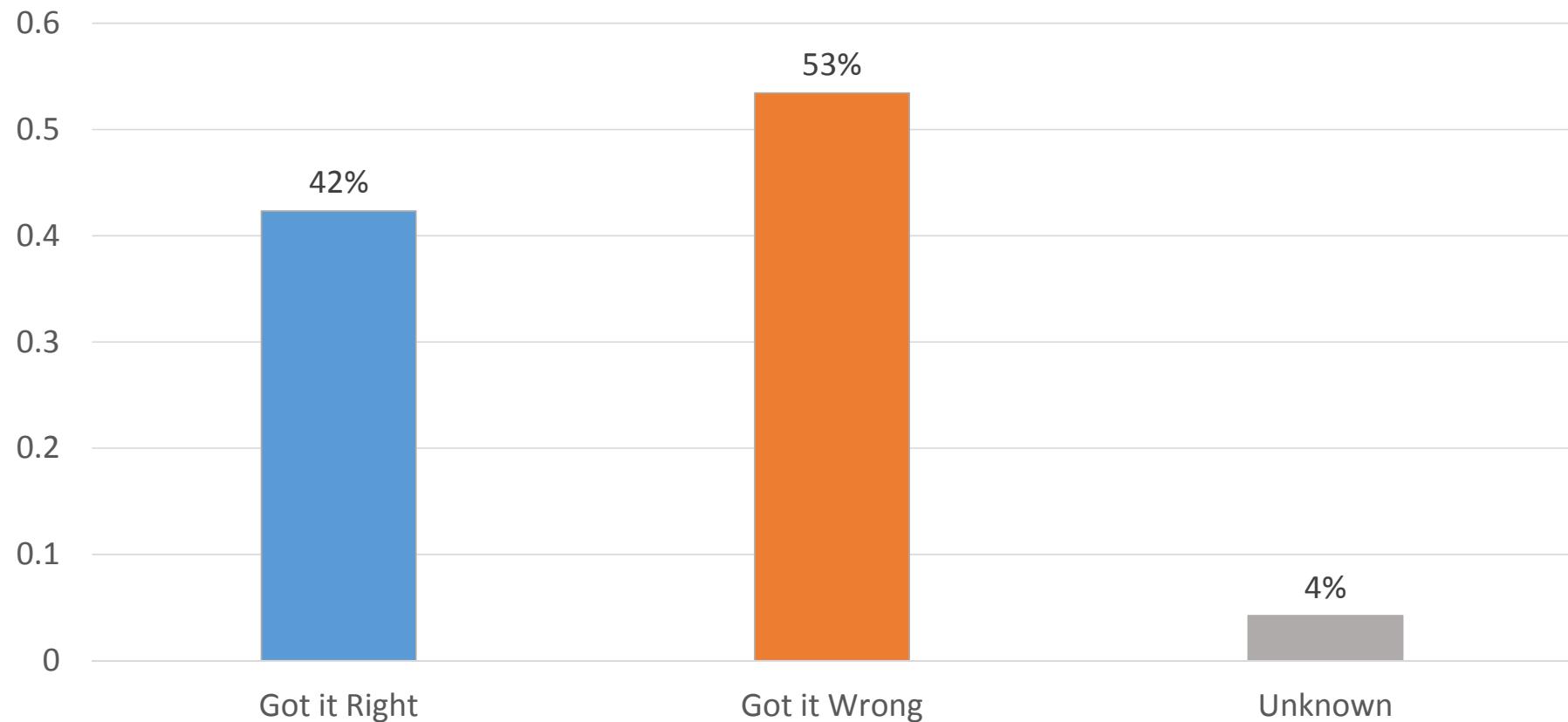
Background

- Asked to write questions about re-identification and privacy-accuracy trade-off
 - Are respondents worried about re-identification?
 - Do respondents prefer more privacy at the cost of less accuracy of publicly released data or are they willing to risk privacy for more accurate and useful data?
- The terms and concepts are familiar to economists, data scientists, and survey researchers but is not something respondents have had to think about

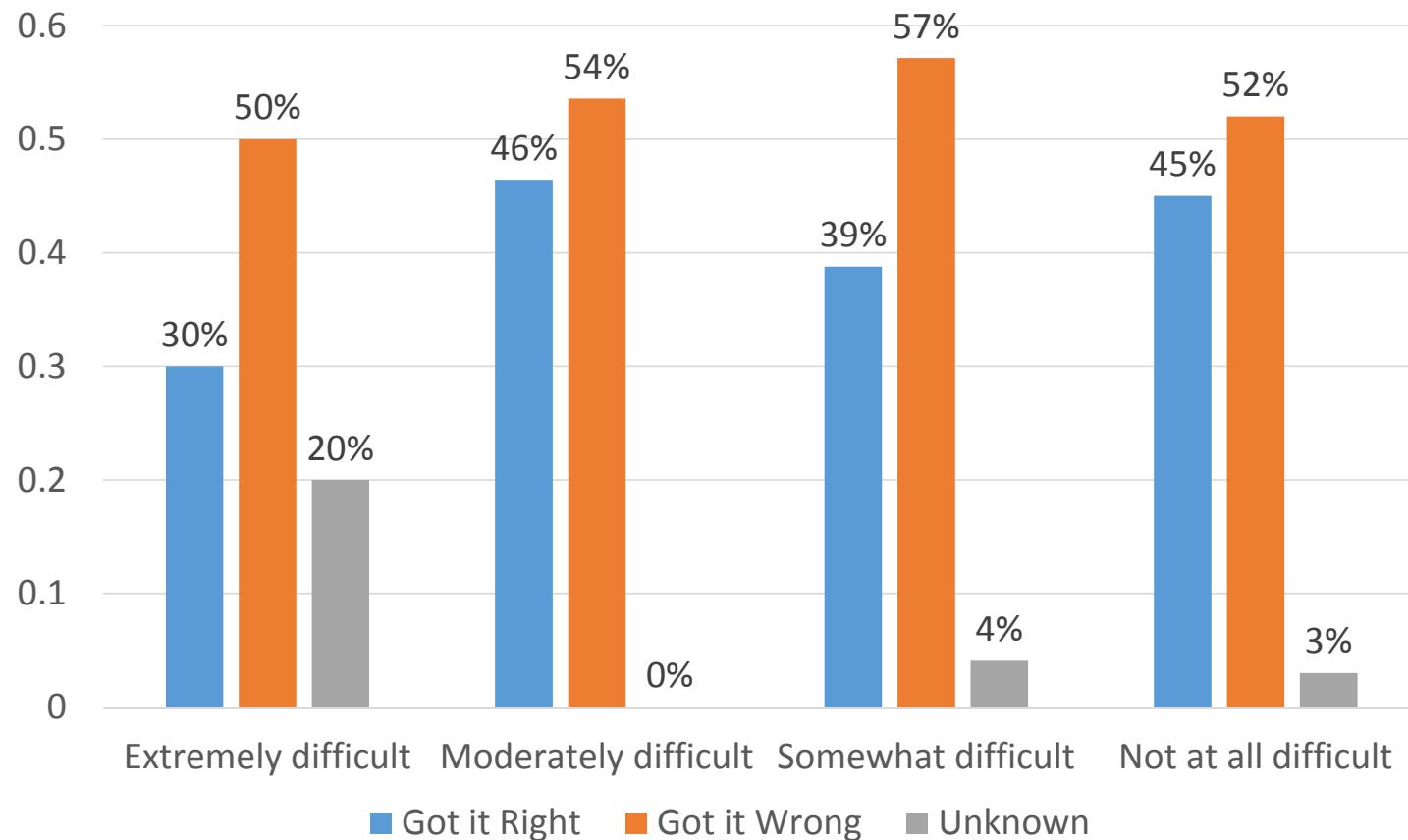
Methods

- Goal: Large, nationally representative sample survey
- In-person cognitive testing, think-aloud with intermittent probing
 - Round 1: 27 interviews
 - Round 2: 17 interviews
- Proof of concept
 - Qualtrics instrument
 - Sample: 20,000 households, up to 3 emails per household, randomized national sample
 - 727 responses after cleaning
 - Half of sample had web probes

Frequency of comprehension of reidentification question

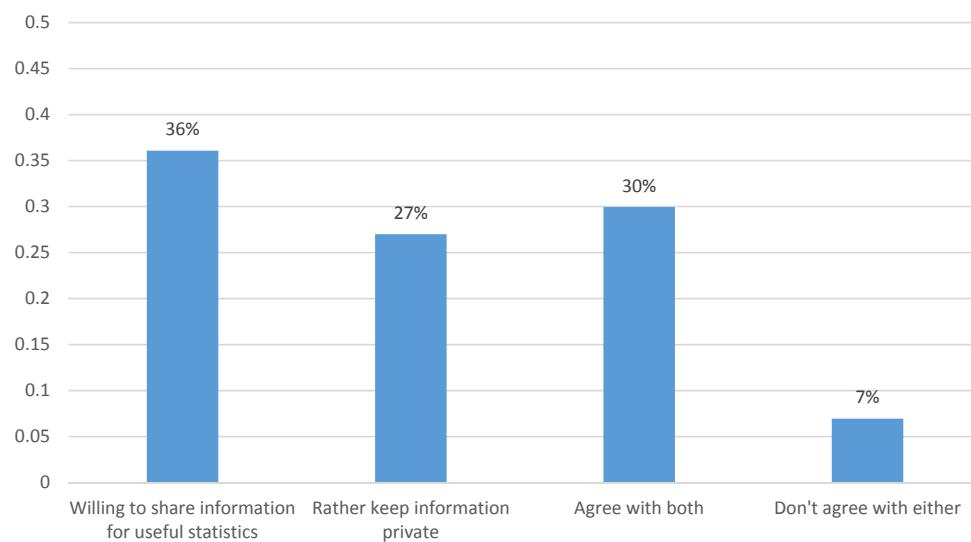


Perceived difficulty of the reidentification question by comprehension



- A. I am willing to share information about me and my household with some government agencies (like the Census Bureau) so the government can produce more useful data and statistics, even if it means having less control over that information.
- B. I would rather keep information about me and my household private even if it means the data and statistics produced by the government are less useful.
- C. I agree equally with both
- D. I don't agree with either

Willingness to share information so government can produce useful statistics



Discussion

- The term re-identification is a problem
 - Respondents seemed to understand the behavior and definition but the term was confusing
 - Examples helped
- Methodological implications
 - Web probing confirmed in-person findings
 - Provided data on prevalence of misinterpretations
 - Open and closed comprehension probes performed similarly
 - Perceived difficulty probe not very useful
- Next steps
 - Working on finalizing instrument for larger study