USDA

United States Department of Agriculture

Transparent Reporting for Integrated Data Quality: Assessing the User's Perspective

Mark Prell
Economic Research Service

Panel on Transparency and Reproducibility of Federal Statistics, CNSTAT
May 21, 2019

The findings and conclusions in this presentation are those of the author and should not be construed to represent any official USDA or U.S. Government determination or policy

Outline

- Background and Design of the Transparent Reporting Project
- Conducting the project
- Results of the project: data quality
- Results of the project: documentation (transparent reporting)
- Challenges of improving documentation
- Conclusions

Transparent Reporting: Meaning and Significance

- *Transparent reporting* = agency documentation is clear
- A “statistical product” can be an estimate (GDP, Consumer Price Index) or a micro-level dataset
- *Transparent reporting* enables users to determine data quality (“fitness-for-use”) in their own particular applications

FCSM Transparent Reporting Project

- Chris Chapman (co-chair), NCES
- Mark Prell (co-chair), ERS
- Samson Adeshiyan, NCSES
- Dennis Fixler, BEA
- Tom Garin, NCVAS
- Lisa Mirel, NCHS
- Polly Phipps, BLS
- John Eltinge (co-chair), FCSM
- Shelly Wilkie-Martinez (co-chair), FCSM

Context for Transparent Reporting Project: workshops and presentations

- FCSM Working Group /WSS sponsored series of workshops
  - Completed: input; process; output; metadata; geospatial
  - Next workshop: sensitivity analysis (June 19, 2019)
- Two panels at 2018 FCSM Research and Policy Conference

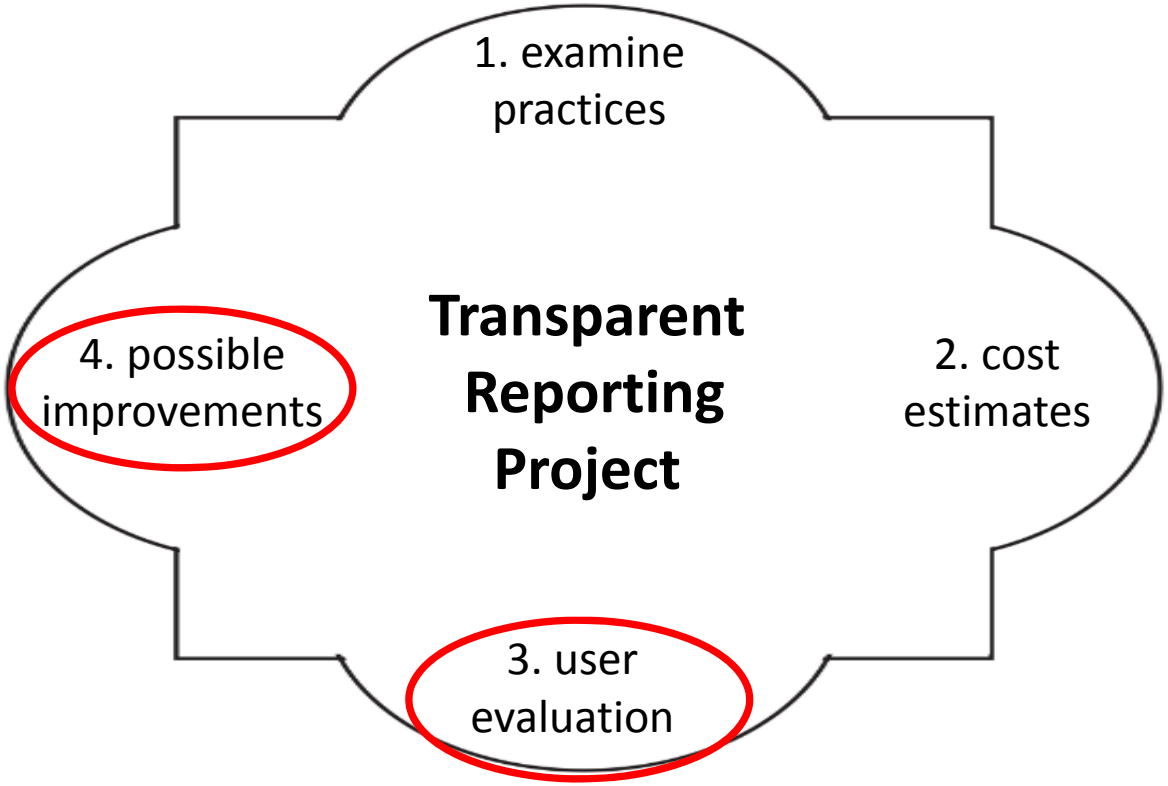
Context for Transparent Reporting Project: publications

- Czajka, John and Mathew Stange. (2018) *Transparency in the Reporting of Quality for Integrated Data: A Review of International Guidelines and Standards*. Mathematica Policy Research
- Brown Alexandra, Andrew Caporaso, Katharine Abraham and Frauke Kreuter. (2018) *Findings from the Integrated Data Workshops hosted by the Federal Committee on Statistical Methodology and Washington Statistical Society*
- FCSM Working Group on Transparent Quality Reporting in the Integration of Multiple Data Sources. (2018) *Transparent Quality Reporting in the Integration of Multiple Data Sources: A Progress Report, 2017-2018*.

Context for Transparent Reporting Project ongoing work

- Transparent Reporting Project (the “case studies” project)
- “Expanding Data Quality Practices to Non-Survey and Integrated Data”

Four Goals


1. examine practices
Transparent Reporting Project
2. cost estimates
3. user evaluation
4. possible improvements

7 case studies in 4 categories of data

I. **National income accounts**

1. Consumption component of Gross Domestic Product (BEA)

II. Integration of administrative and survey data

2. National Postsecondary Student Aid Study (NCES)
3. Linkage between National Hospital Care Survey and National Death Index (NCHS)
4. Veteran Population Model 2016 (NCVAS)

III. Proprietary data

5. Telecommunications component of Consumer Price Index (BLS)
6. National Household Food Acquisition and Purchase Survey (ERS)

IV. Integration of multiple surveys

7. Scientists and Engineers Statistical Data System (NCSES)

Customer Survey: Two Pioneering Features

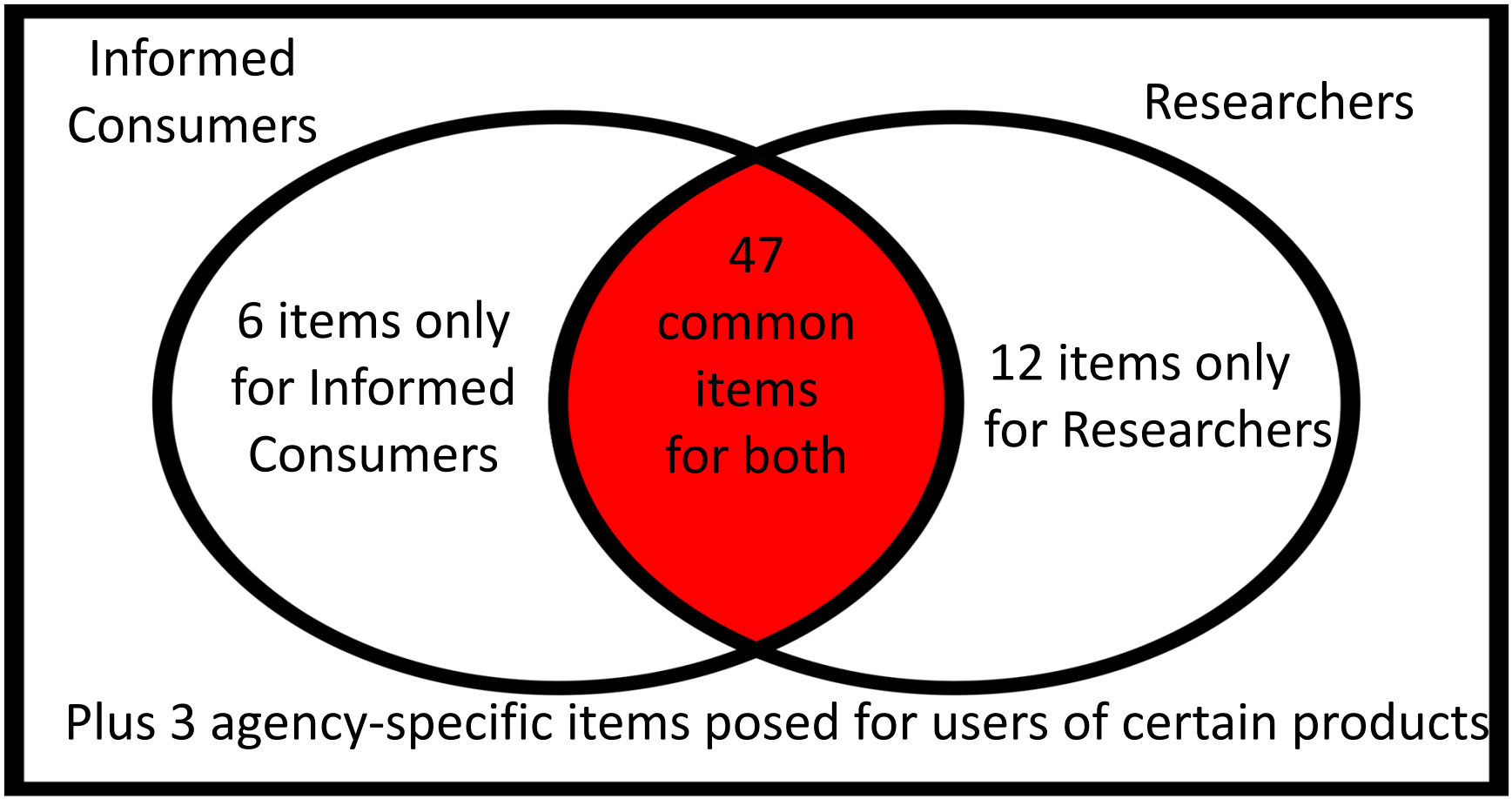
- First known survey to collect user assessments of transparent reporting using common items across statistical products.
- Two instruments for two groups of users:
  - Informed consumers
  - Researchers

Two Groups of Users

- Informed Consumers
  - use results based on the integrated data product
  - need information on how the product was developed to interpret the information it provides
- Researchers
  - directly use raw data in the integrated data product
  - need detailed technical information for research

Two instruments in the customer survey

(65 distinct items + 3 agency-specific items)



Outline

- Design of the Transparent Reporting Project
- Conducting the project
- Results of the project: data quality
- Results of the project: documentation (transparent reporting)
- Challenges of improving documentation
- Conclusions

Data Collection

- July-Sept: developed survey items and web-based interface
- Oct: received OMB approval; sent advance e-mails to users;
- Nov 1 – Dec 11: survey open
- Dec 21: received raw data
- Feb: cleaned data
- March: statistical analysis

Outline

- Design of the Transparent Reporting Project
- Conducting the project
- Results of the project: data quality
- Results of the project: documentation (transparent reporting)
- Challenges of improving documentation
- Conclusions

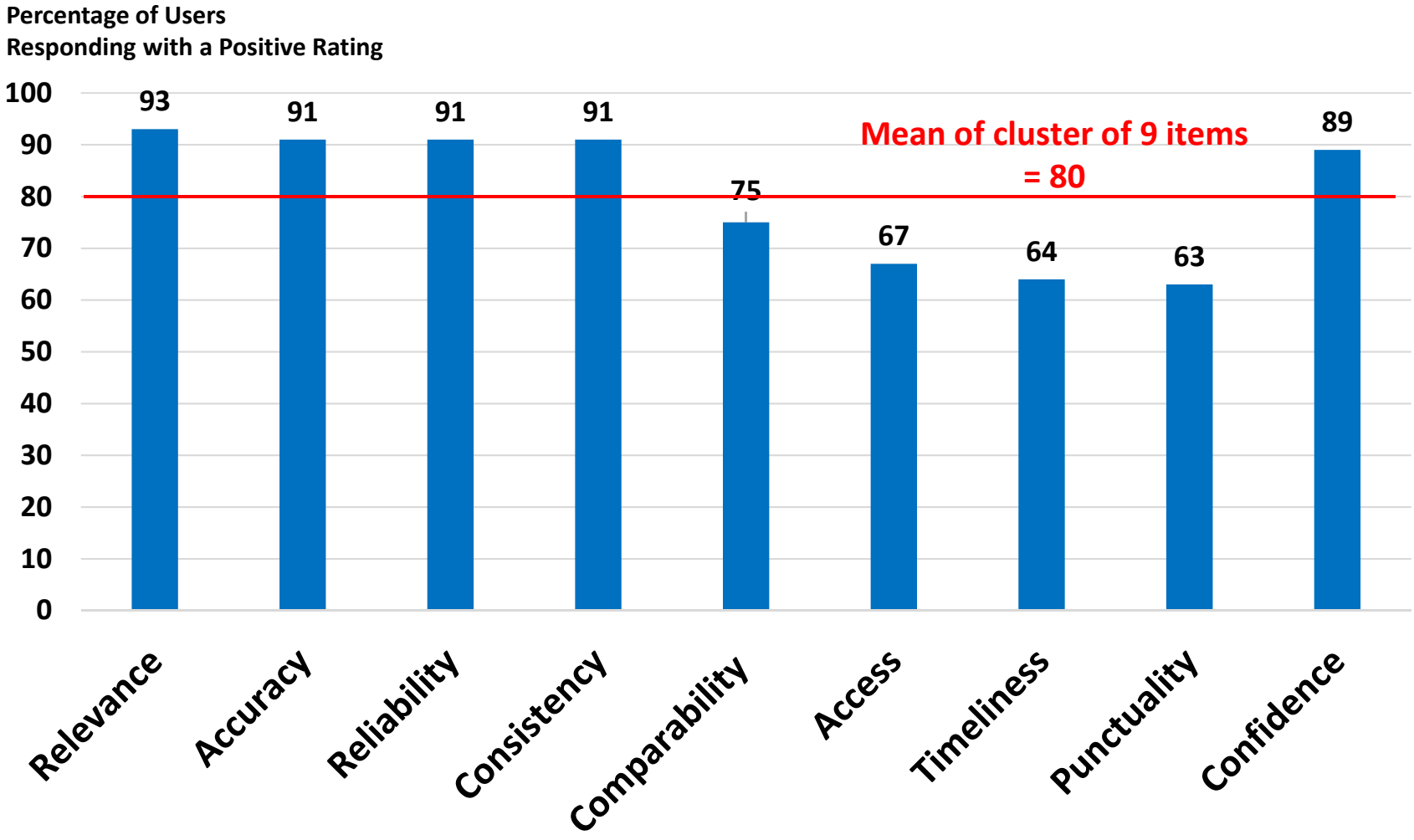
Interpreting Results

- Each agency selected known stakeholders (105 contacted)
- Survey was purposive rather than probabilistic
- Results are:
  - ➢ direct measures of responses of participating users
  - ➢ illustrative, suggestive
  - ➢ not statistically generalizable to all users

Response Rates

- Response Rates:
  - overall 46/105 (43.8 percent)
  - researchers (51.3 percent)
  - informed consumers (25.8 percent)

User evaluation of data quality and overall user confidence in the data



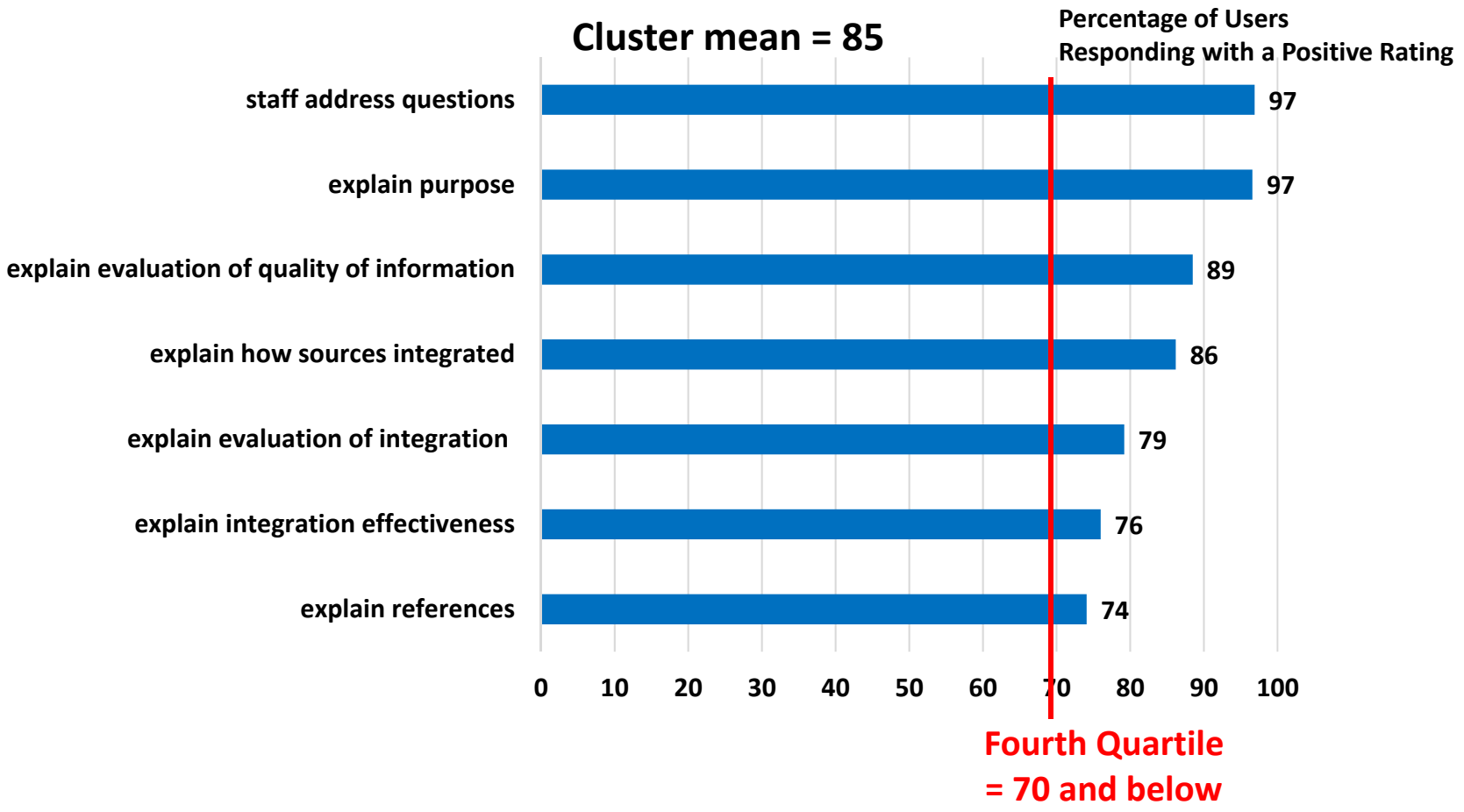
Outline

- Design of the Transparent Reporting Project
- Conducting the project
- Results of the project: data quality
- Results of the project: documentation (transparent reporting)
- Challenges of improving documentation
- Conclusions

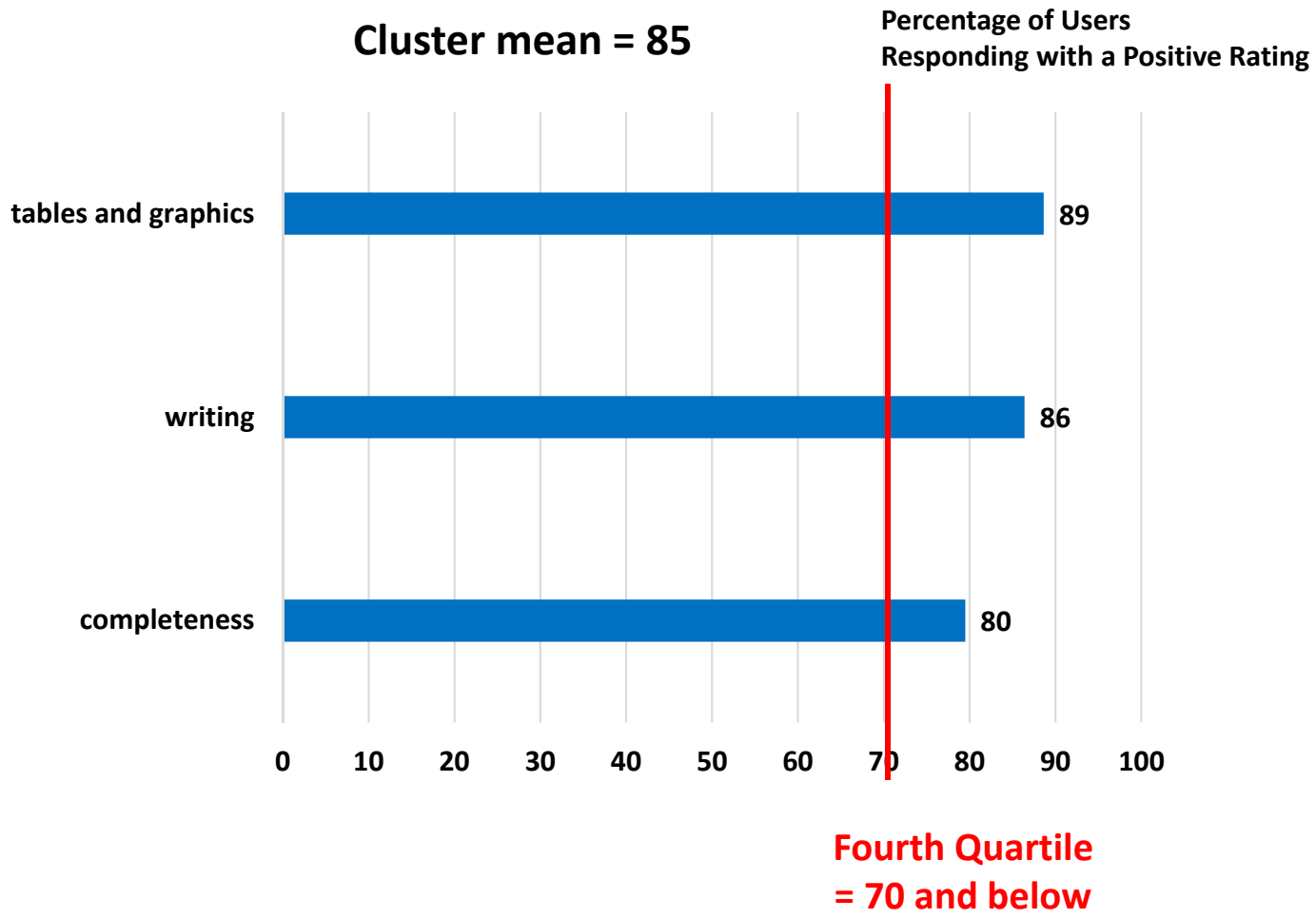
35 Items on Documentation: The Deep Dive

- Highly detailed customer survey; some items have small # responses
- Average of 35 items' percentage of positive ratings: 80
- Large variation around mean: quartiles are roughly 10-point spreads
  - First quartile: 89-100 (90s)
  - Second quartile: 80-88 (80s)
  - Third quartile: 71-79 (70s)
  - Fourth quartile: 58-70 (60s)

Cluster 1 of 6: Agency Contacts

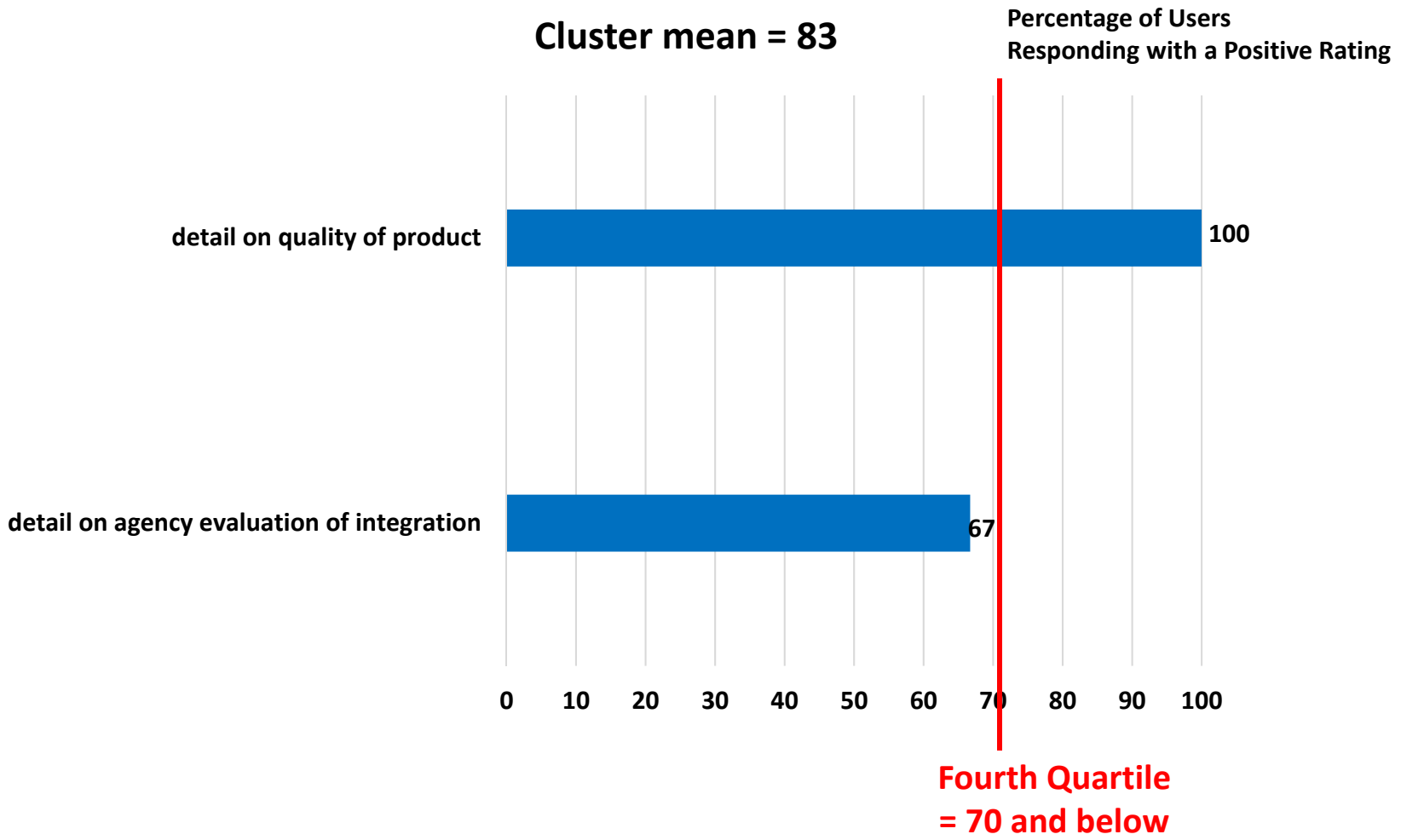


Cluster 2 of 6: Quality of Documentation

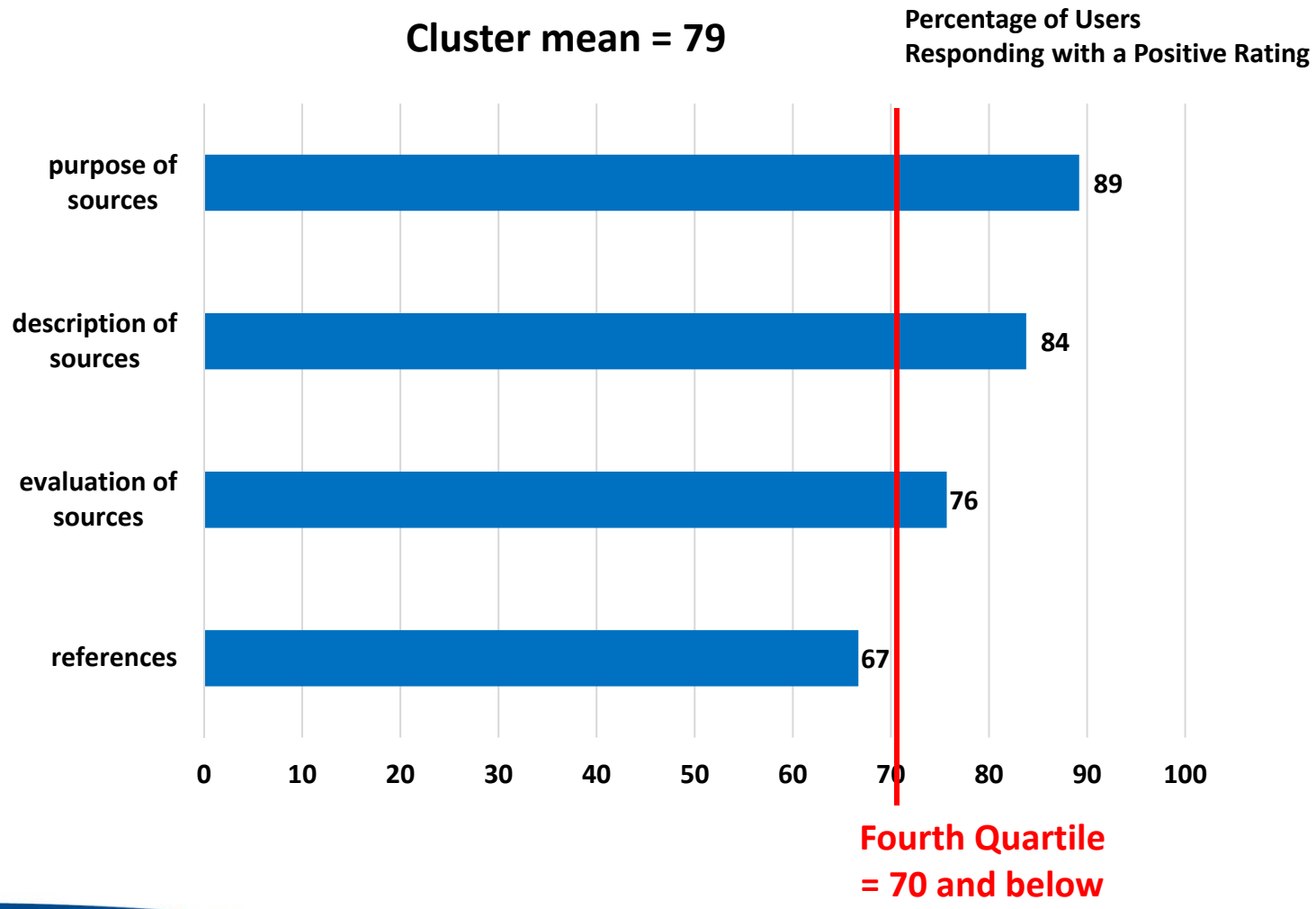


Fourth Quartile
= 70 and below

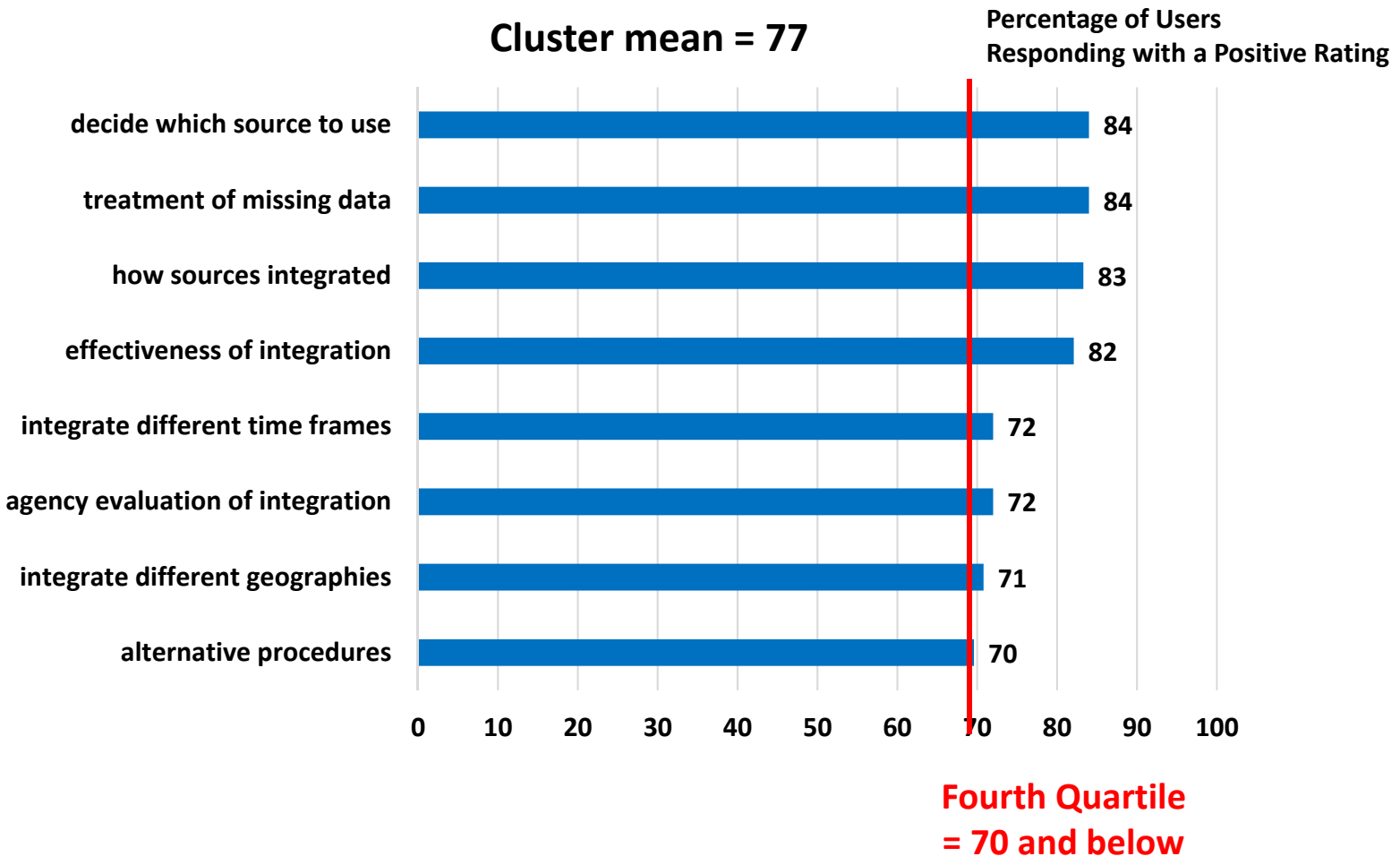
Cluster 3 of 6: Agency Evaluation of Quality



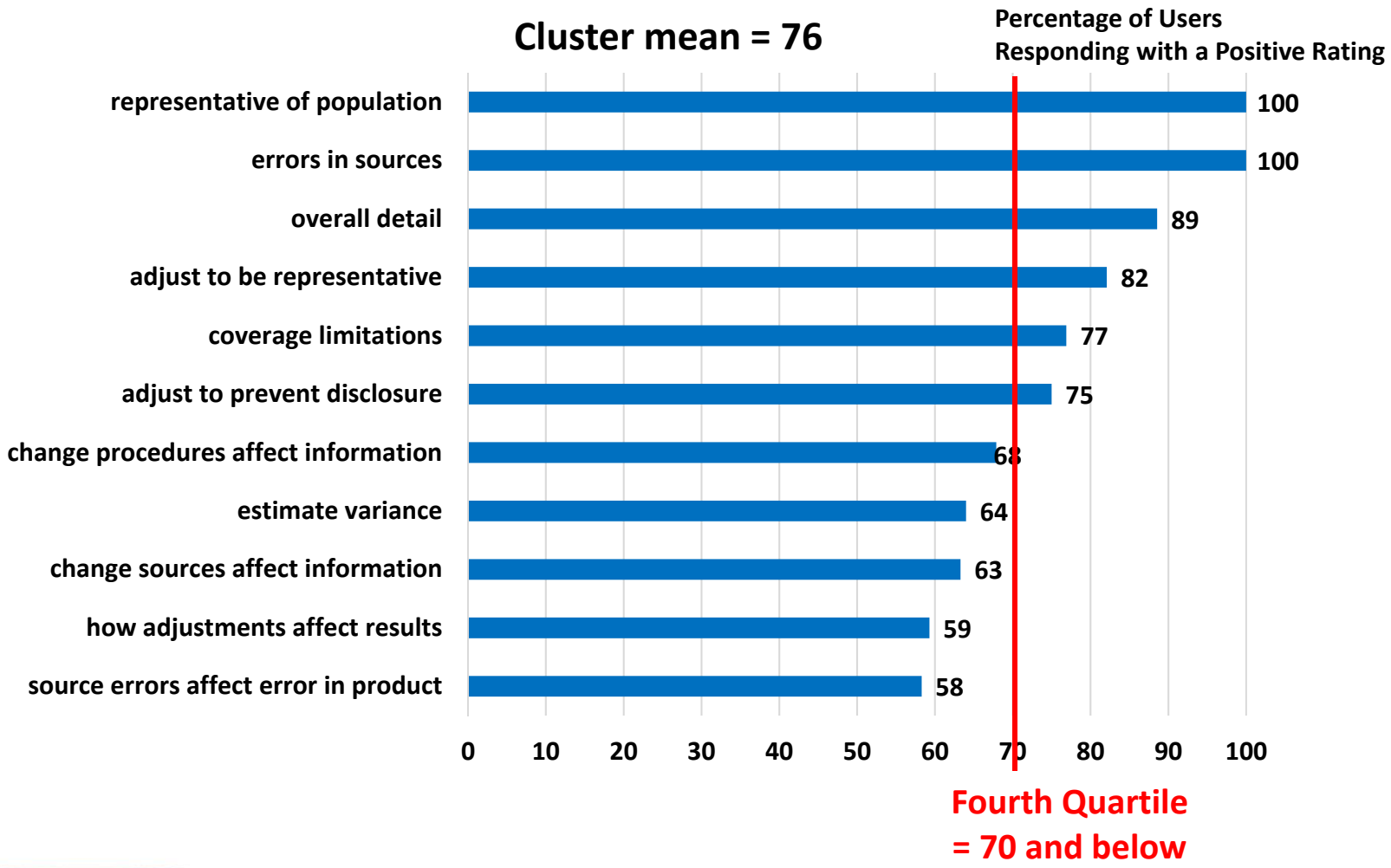
Cluster 4 of 6: Sources of Data



Cluster 5 of 6: Data Integration



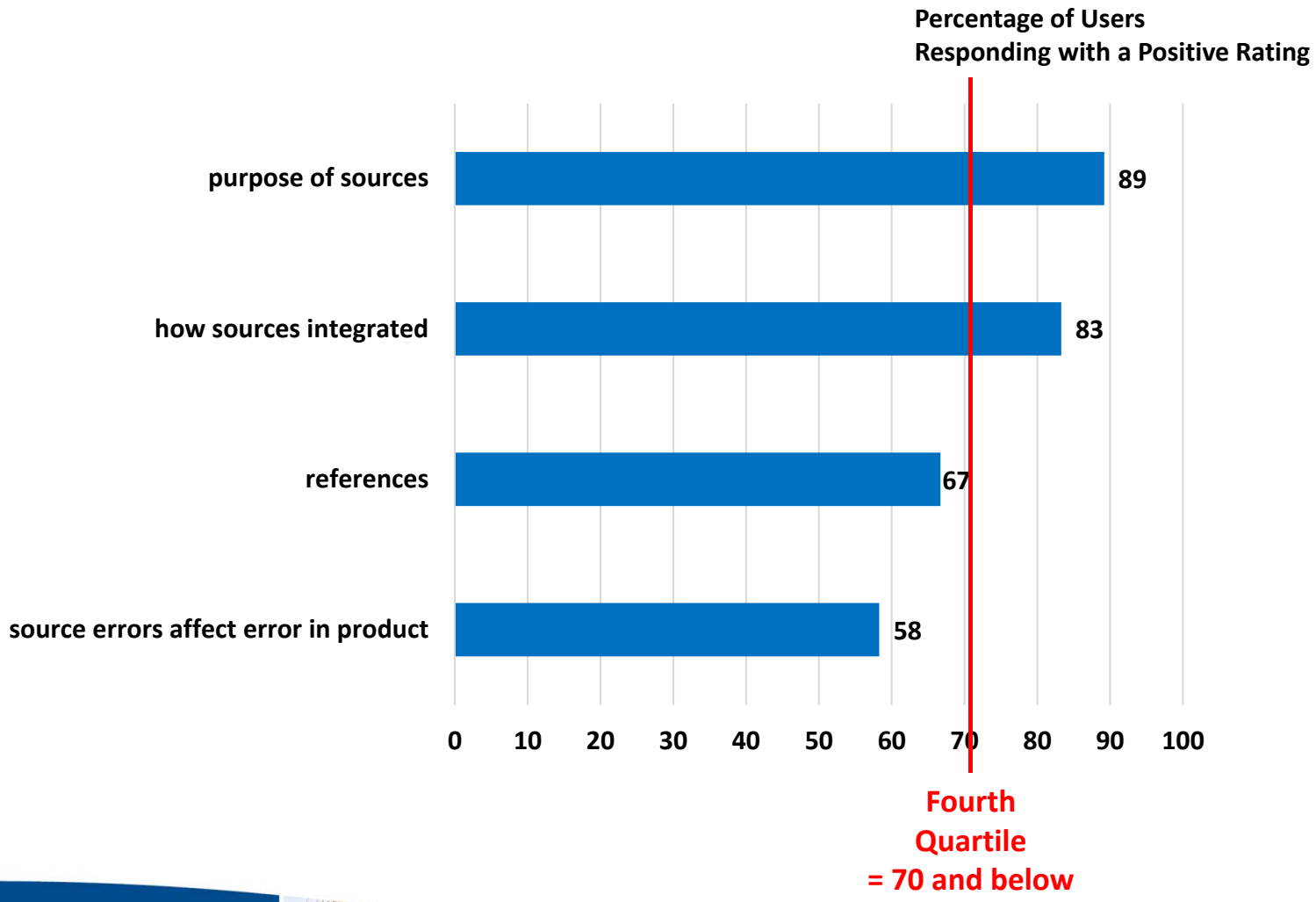
Cluster 6 of 6: How to Use Product



Outline

- Design of the Transparent Reporting Project
- Conducting the project
- Results of the project: data quality
- Results of the project: documentation (transparent reporting)
- Challenges of improving documentation
- Conclusions

Challenges of Improving Documentation



Dual Challenge: Obtaining information on benefits and costs of improving documentation

- A customer survey is useful, but provides only the benefit-side:
  - users express high or low satisfaction for various items
  - items with low percentages of positive ratings are ones where users would highly value more detail
- Agencies also need (internal) information about the cost-side:
  - how difficult/easy it is to improve an item
  - items that have a lower percentage are not necessarily the easiest or least costly to improve
  - possible opportunities to improve items that already have a high percentage

Outline

- Design of the Transparent Reporting Project
- Conducting the project
- Results of the project: data quality
- Results of the project: documentation (transparent reporting)
- Challenges of improving documentation
- Conclusions

Conclusions and Lessons: "In general . . .

- Transparent Reporting:
  - Agencies adopt different approaches to provide information
  - Distinguish documentation for estimate versus dataset
  - Agencies meeting many documentation needs of users
- Data Quality:
  - Users evaluations are high for several dimensions of data quality
  - Users are confident in the data and statistics

Conclusions and Lessons: Room for Improvement

- Users identified areas for improvement
  - Transparent Reporting: references; disclosure limitation methods
  - Data Quality: access; timeliness
- Customer survey gives information on what users value, which is one-side of the information an agency needs
- Improvements in documentation entail costs that need to be weighed against other applications of an agency budget

Thank you!

mprell@ers.usda.gov