

Keep the Baby, Throw Out the Bath Water

The Promises of Real-Time Data Quality Evaluations

Steven Snell, PhD & Carol Sue Haney

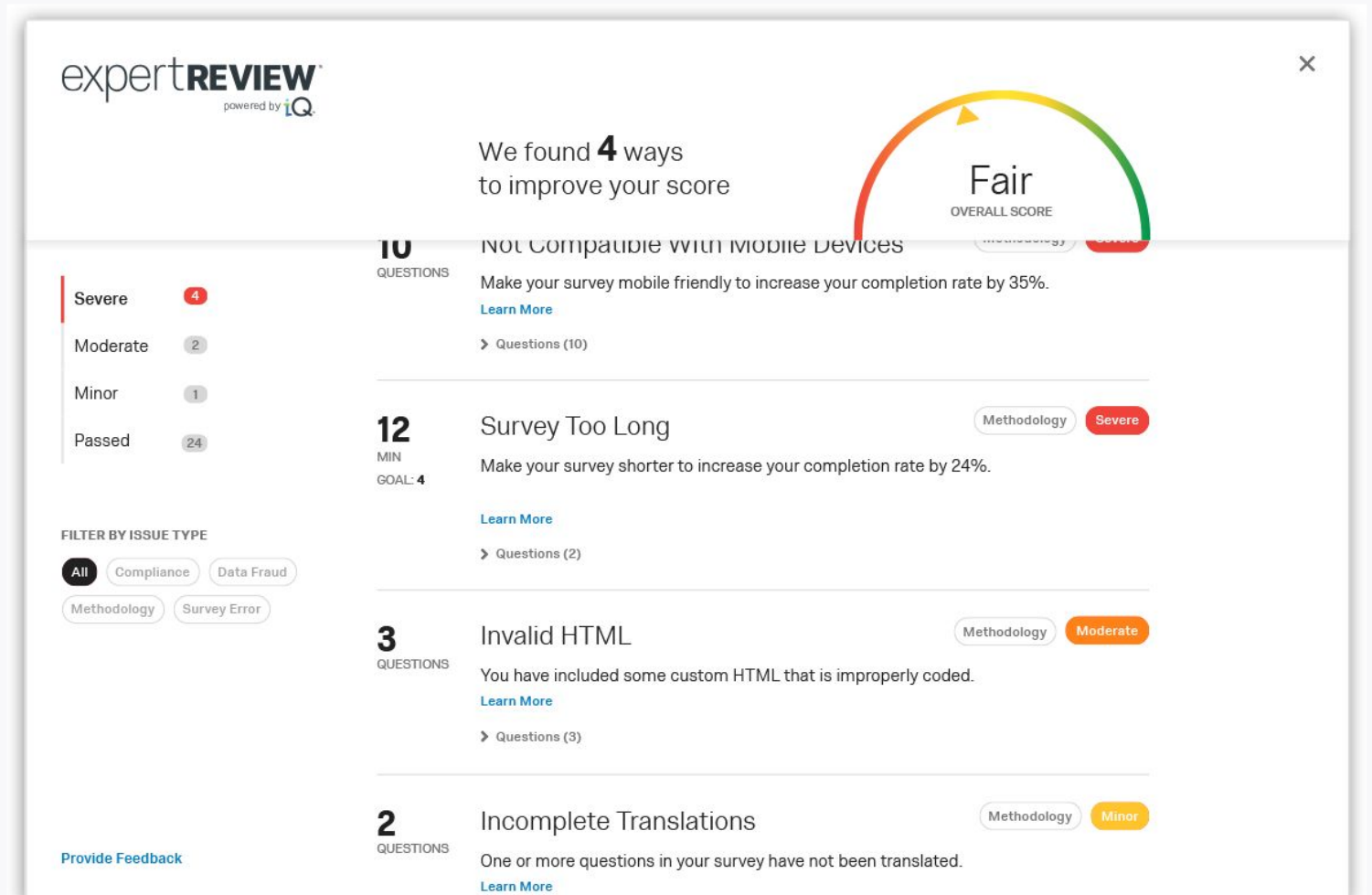
ITSEW, June 2019

qualtrics^{XM}

Motivation — *a Total Survey Error approach*

Can we do more to protect survey novices from themselves? A TSE approach

- Sampling
- Coverage
- Nonresponse
- Measurement
- Postsurvey



Motivation — *applications*

Researchers urgently looking to improve rigor around online panels

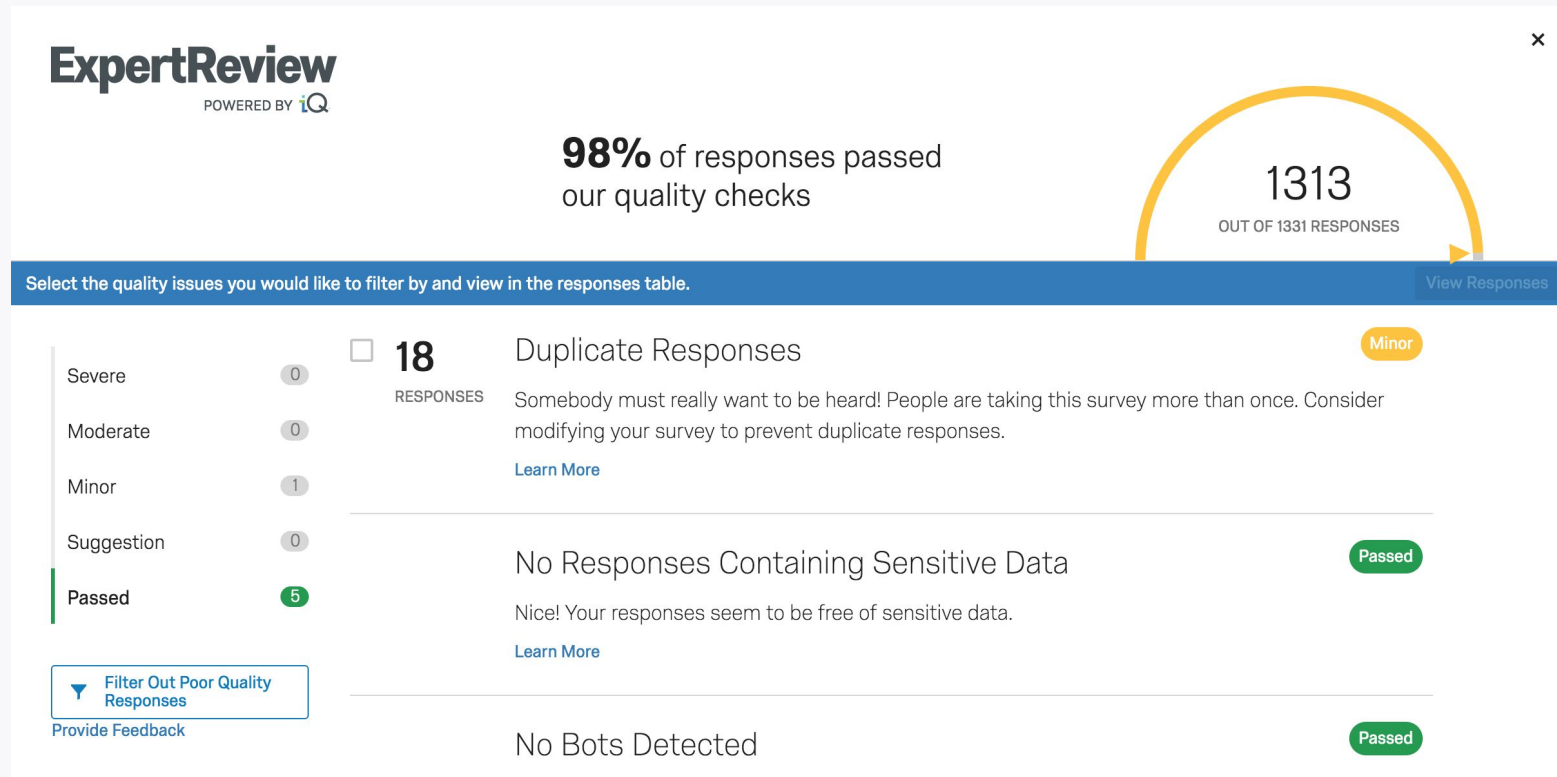
- Replication crisis in the social sciences
- Concerns about non-probability v. probability sample
- Heterogeneity across online panels

*Can we “make do” and improve insights from online panel
by thoughtfully cleaning data?*

Motivation — *after the design phase*

What can we do to help users working with low quality data?

- Duplicates
- Bots
- Speeders
- Straightliners
- Low quality open-ends



Approach — *overview*

- Attitudinal survey—gauging health, political/policy attitudes
 - More than 19,000 responses
 - Administered across four web-based panels
 - Six quarterly waves
 - Median LOI ~7.5 minutes
 - Coarsened panel-based quotas to mirror overall US demographics
- Passively flagged “low quality response” (see next slide)
- Familiar framework—benchmarking against “the truth”

▪

Approach — *measuring quality*

Current inclusions

- Duplicates (repeat IP address in same wave)
- Speeders (complete survey in less than one-half median response time)
- Straight-liners (give the same response to all items in a grid)
- Low quality open-ends (received more than one OE flag)

Future candidates (not in this iteration)

- Item non-response (including DK and N/A)
- Inconsistent responses



Descriptives on data quality

- Data quality is similar across panel providers
- Flags are extremely weakly correlated with one another

| | Sample A | Sample B | Sample C | Sample D | Total |
|----------------|----------|----------|----------|----------|-------|
| Duplicates | 1.5% | 3.1% | 2.8% | 3.3% | 2.6% |
| Speeders | 1.3% | 0.8% | 1.4% | 1.6% | 1.2% |
| Straightliners | 15.3% | 16.3% | 15.3% | 18.5% | 16.2% |
| Low quality OE | 3.0% | 2.3% | 1.8% | 2.3% | 2.4% |

| | Duplicates | Speeders | Straightliners | Low quality OE |
|----------------|------------|----------|----------------|----------------|
| Duplicates | 1.00 | | | |
| Speeders | 0.02 | 1.00 | | |
| Straightliners | 0.01 | 0.01 | 1.00 | |
| Low quality OE | 0.01 | 0.10 | 0.07 | 1.00 |

Benchmarks

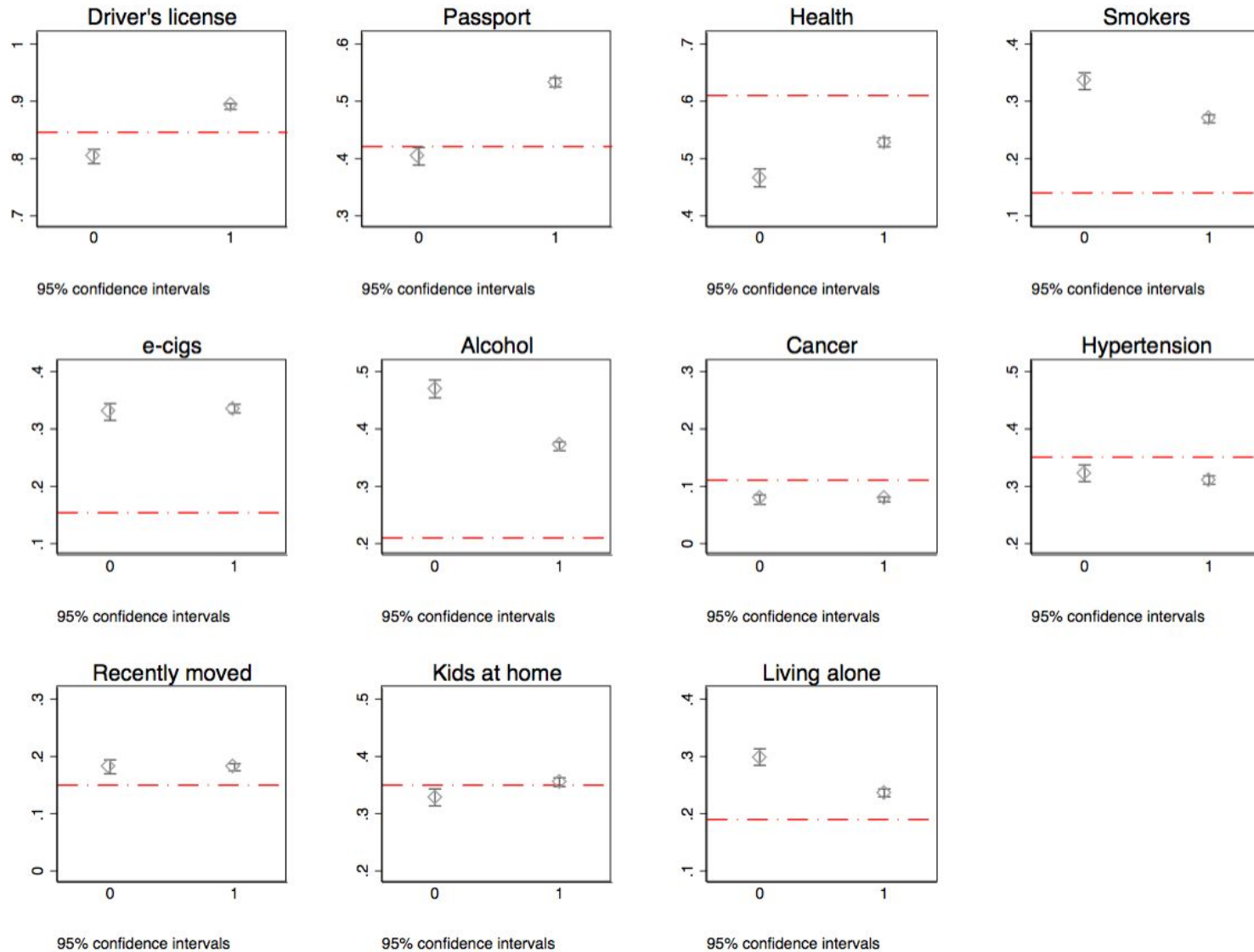
- Do you personally have a current driver's license? **(85%)**
- Do you personally own a valid U.S. passport? **(42%)**
- Would you say your health in general is excellent, very good, good, fair, or poor? **(61% excellent/very good health)**
- Have you EVER been told by a doctor or other health professional that you had any of the following medical conditions? [Hypertension **35%**, Cancer or a malignancy of any kind **12%**]
- In ANY ONE YEAR, have you had at least 12 drinks of any type of alcoholic beverage? **(21% lifetime abstainers)**
- Have you smoked at least 100 cigarettes in your ENTIRE LIFE? Do you now smoke cigarettes... [Every day, Some days, Not at all] **(14% current smokers)**
- Have you ever used an e-cigarette, even one time? **(15%)**
- How many adults, ages 18 and older, including yourself, live in your household? **(19% one-adult households)**
- And how many children younger than 18 years of age live in your household? **(35% with one or more children)**
- How long have you lived at this address? **(15% one year or less)**

Results — *on average, how did we fare?*

Better on four benchmarks, worse on two, and no difference on five!

| | Truth | Estimate | Difference | Cleaned estimate | Cleaned difference |
|---------------------------|-------|----------|------------|------------------|--------------------|
| Passport (yes) | 42% | 50.6% | 8.6% | 52.6% | 10.6% |
| Driver's license (yes) | 85% | 87.3% | 2.3% | 89.0% | 4.0% |
| Health (excellent/v.good) | 61% | 51.6% | -9.4% | 52.3% | -8.7% |
| Hypertension (yes) | 35% | 31.4% | -3.6% | 31.5% | -3.5% |
| Cancer (yes) | 12% | 7.7% | -4.3% | 7.7% | -4.3% |
| Smoker (yes) | 14% | 28.3% | 14.3% | 27.4% | 13.4% |
| E-cig (ever) | 15% | 33.4% | 18.4% | 33.7% | 18.7% |
| Drinker (never) | 21% | 39.1% | 18.1% | 37.0% | 16.0% |
| One adult (yes) | 19% | 24.9% | 5.9% | 23.7% | 4.7% |
| Children (yes) | 35% | 35.0% | 0.0% | 35.4% | 0.4% |
| Mover (yes) | 15% | 18.1% | 3.1% | 18.0% | 3.0% |

Results — *how weird are the low quality respondents?*

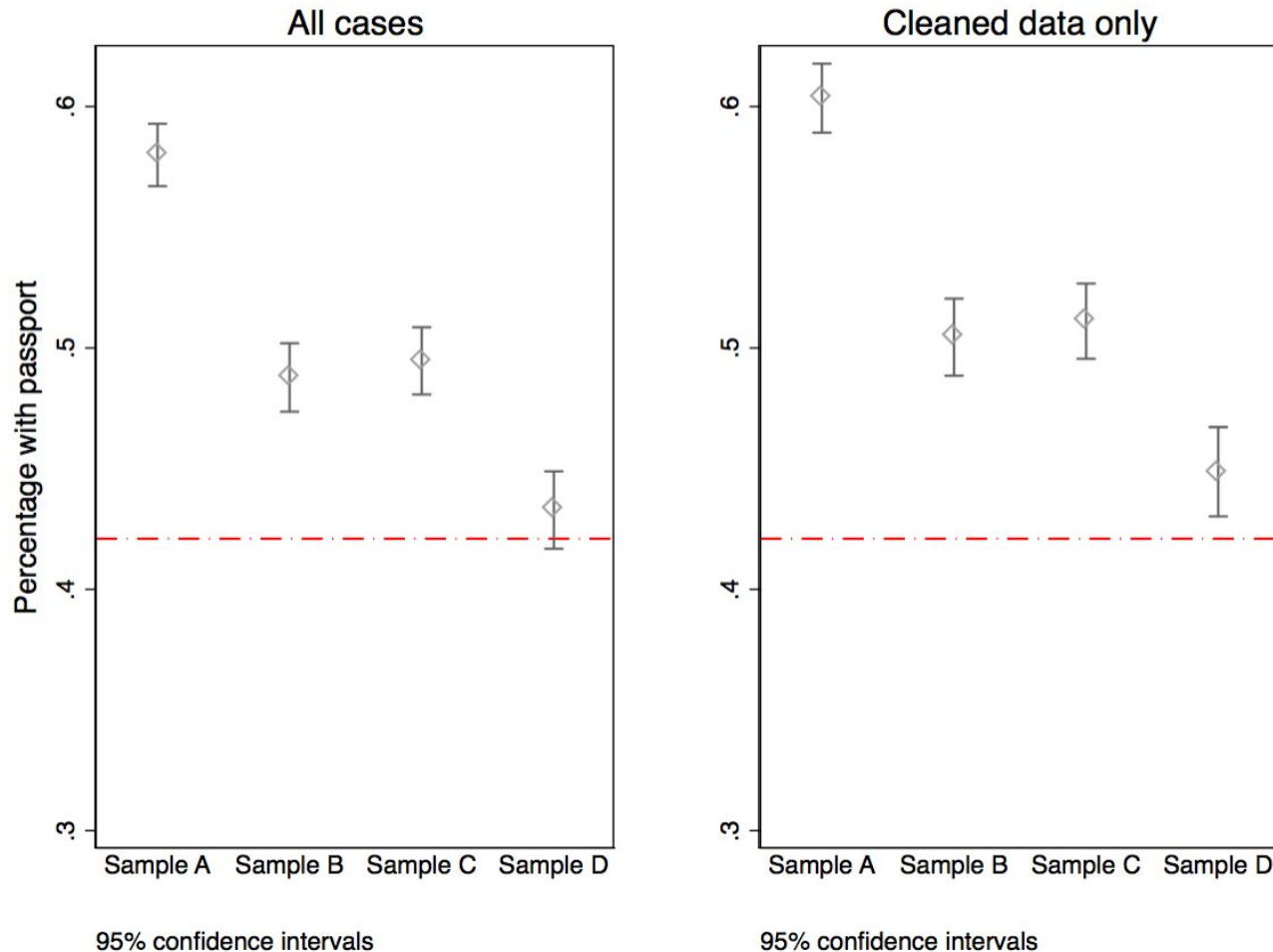


In several instances, low quality respondents give answers that are not statistically different from other respondents.

Pattern is relatively inconsistent.

Results — *digging deeper (passport)*

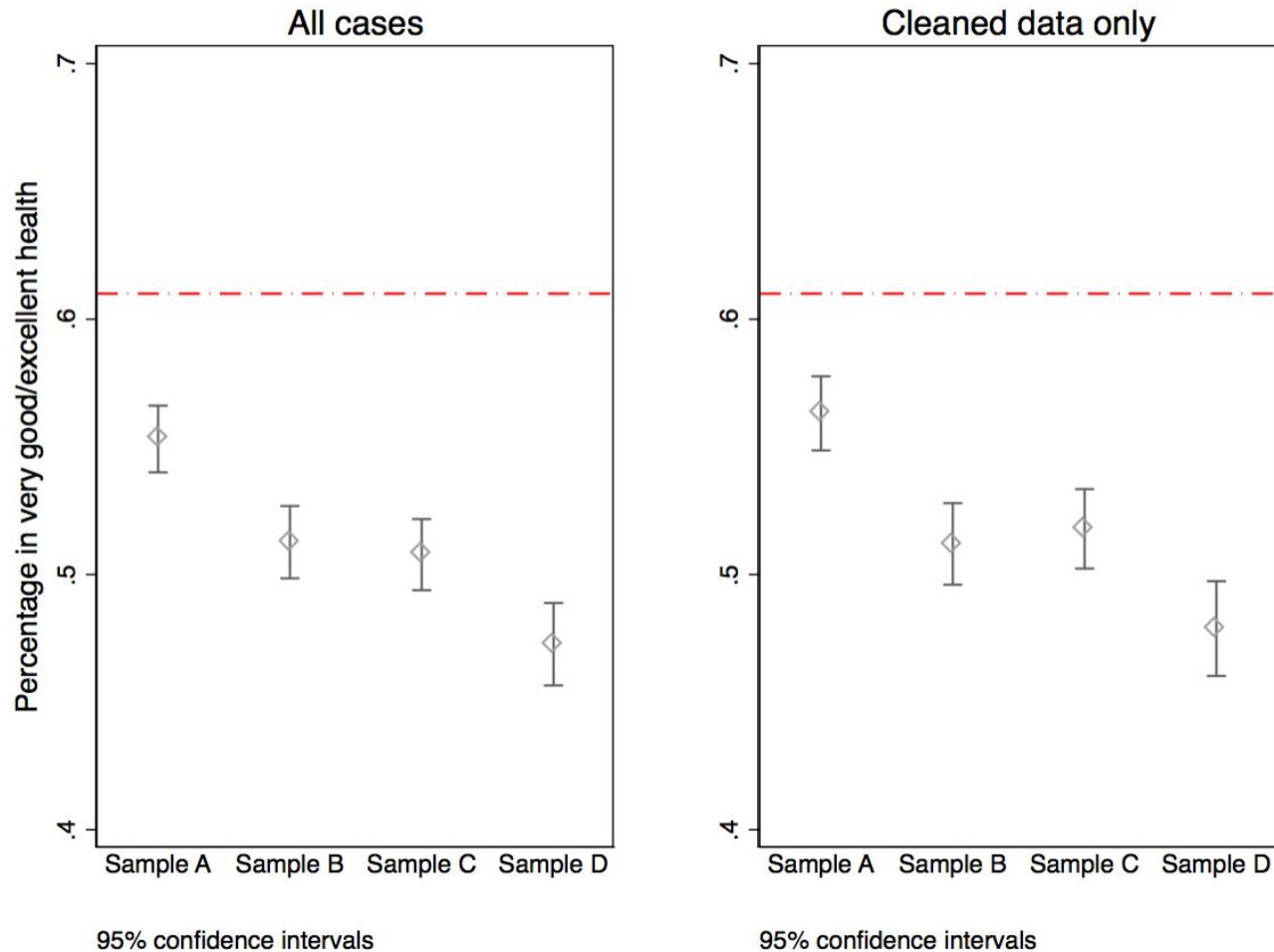
Do you personally own a valid U.S. passport?



| | Coefficient | SE |
|----------------|-------------|------|
| Duplicates | 0.03 | 0.02 |
| Speeders | 0.13 | 0.03 |
| Straightliners | -0.18 | 0.01 |
| Low quality OE | 0.01 | 0.02 |
| Constant | 0.53 | 0.00 |
| Truth | 0.42 | |

Results — *digging deeper (health evaluations)*

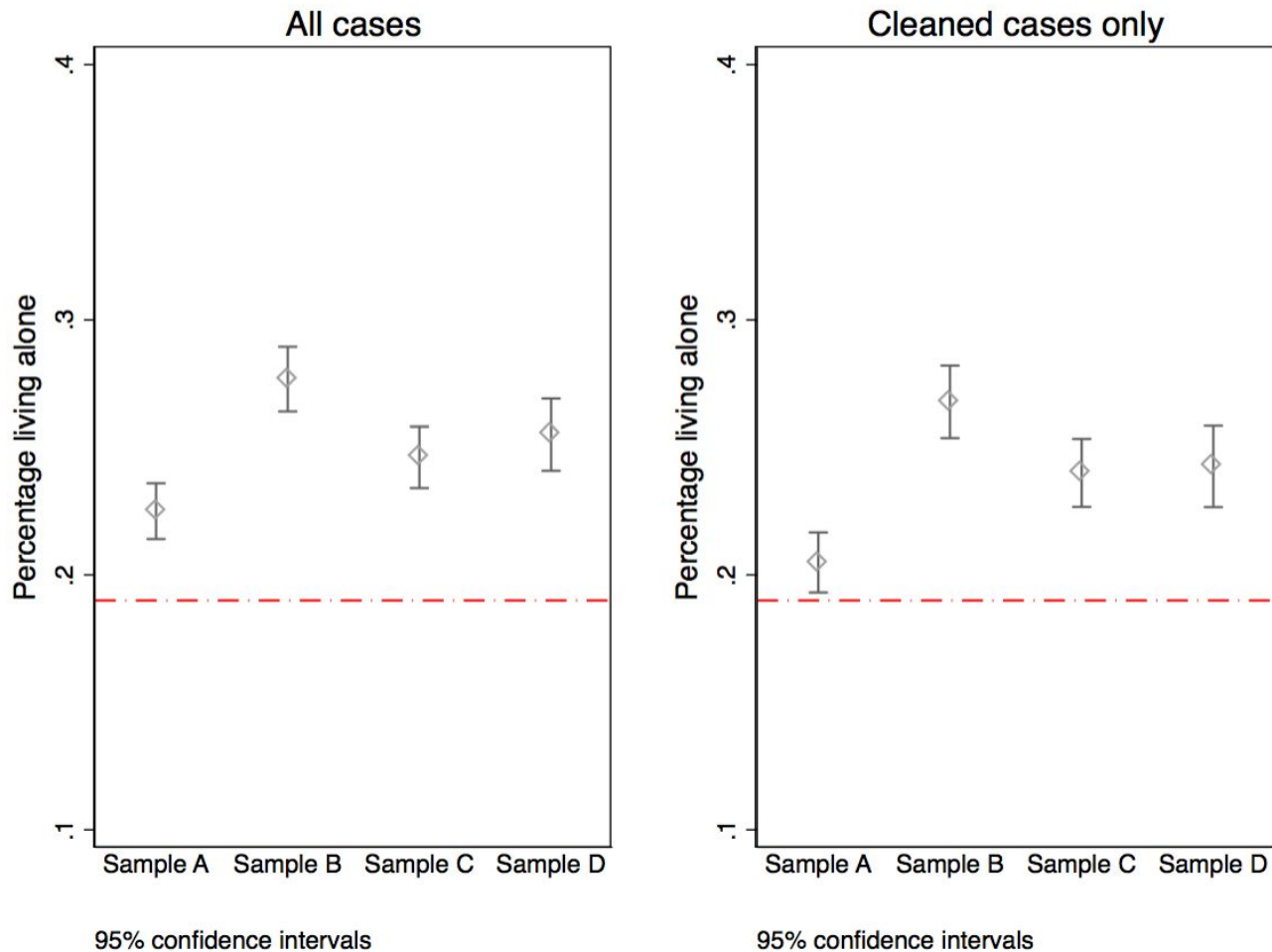
Would you say your health in general is excellent, very good, good, fair, or poor?



| | Coefficient | SE |
|----------------|-------------|------|
| Duplicates | 0.01 | 0.02 |
| Speeders | 0.15 | 0.03 |
| Straightliners | -0.10 | 0.01 |
| Low quality OE | 0.03 | 0.02 |
| Constant | 0.53 | 0.00 |
| Truth | 0.61 | |

Results — *digging deeper (adults in household)*

How many adults, ages 18 and older, including yourself, live in your household?



| | Coefficient | SE |
|----------------|-------------|------|
| Duplicates | 0.00 | 0.02 |
| Speeders | 0.11 | 0.03 |
| Straightliners | 0.06 | 0.01 |
| Low quality OE | 0.06 | 0.02 |
| Constant | 0.24 | 0.00 |
| Truth | 0.19 | |

Concluding thoughts

Findings

- Differences in data quality across vendors are generally small
- Data quality flags are not correlated with one another
- Aggressive data cleaning produced small and inconsistent improvements in survey accuracy, sometimes making estimates worse

Next steps

- Analysis by device and browser
- Interactive cleaning
- Real-time replacement
- Are panel providers on to us?

Thank you.

Questions

- Do we have the right inputs? What are we missing? If you were trying to automate the data quality assessment of thousands of surveys that you didn't write, what would you look for?
- How should we handle straightlining? People want to see it, but there are lots of instances where the behavior is completely appropriate; not at all indicative of inattention or low data quality.
- Should we expect for data quality flags to highly correlated with one another? Should we worry that they are not more strongly correlated?