Pew Research Center

2018 FCSM

Discussion: Nonprobability Sampling and Estimation

Courtney Kennedy

Director of Survey Research

Common Themes

- An interesting cross-section of studies evaluating nonprobability samples in a range of contexts
- As a field we are making strides in bringing theory to bear on these data and talking more precisely about our approaches
- But we are not finding much compelling support for the key assumptions underpinning some approaches

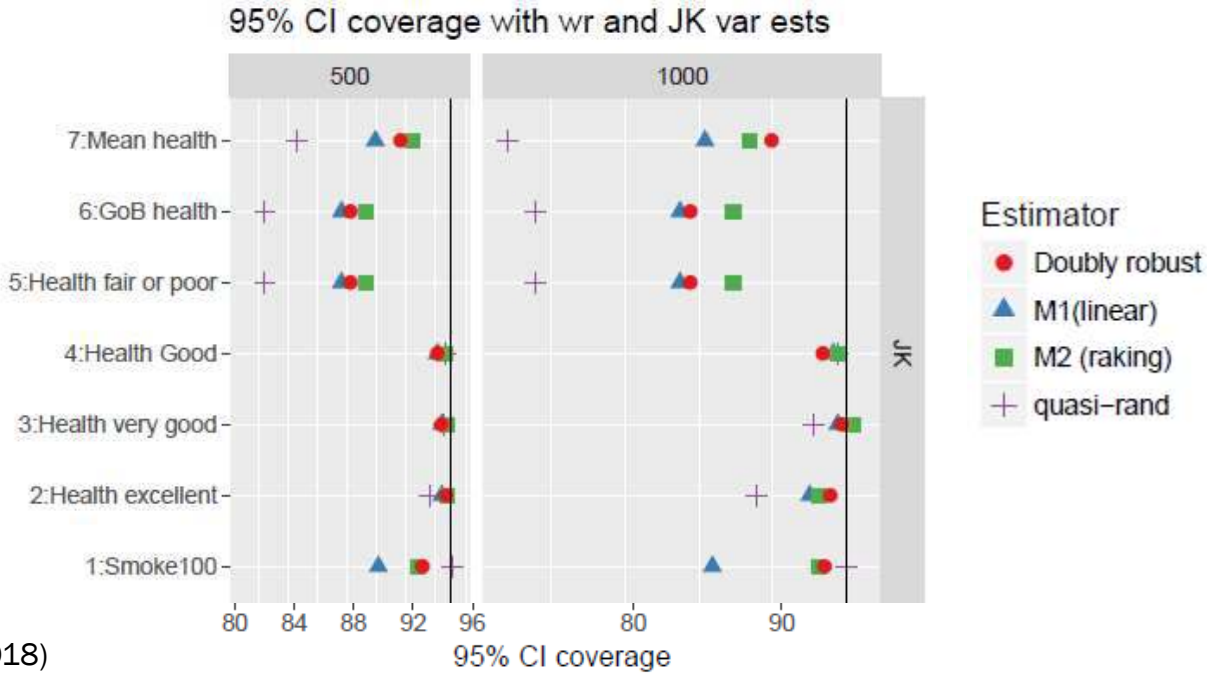
Valliant “Alternatives for Estimation from Nonprobability Samples”

- Provides a lucid framework for thinking about estimation with nonprobability data
- Terminology and definitions in this paper (and recent works by others) seem to be converging, which is a very positive development for survey researchers
  - “quasi-randomization”
  - “superpopulation modeling”
  - “doubly robust estimation”

- We’re starting to talk about nonprobability surveys using more precise and more standardized language

Valliant “Alternatives for Estimation from Nonprobability Samples”

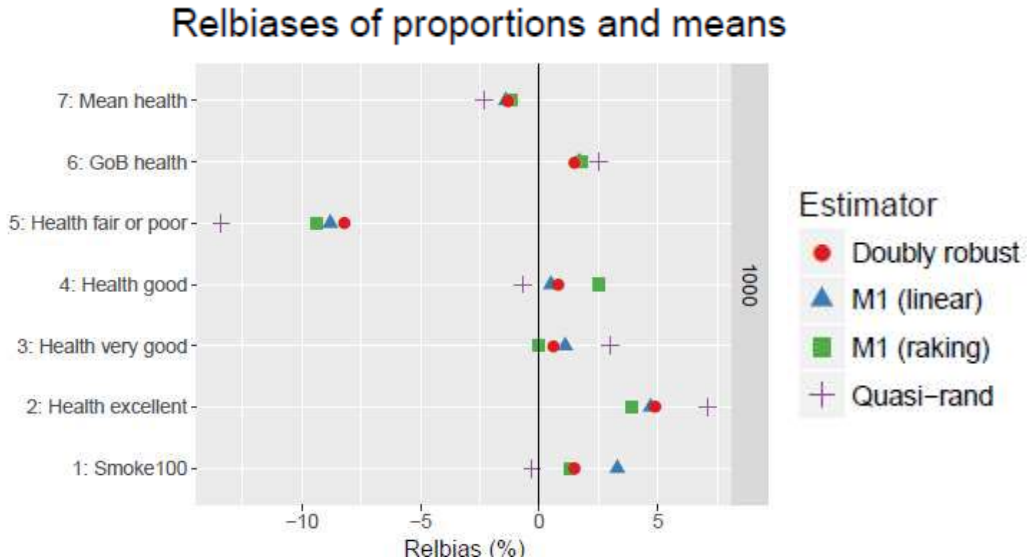
- Analysis of the coverage of 95% confidence intervals was particularly insightful



Source: Valliant (2018)

Valliant “Alternatives for Estimation from Nonprobability Samples”

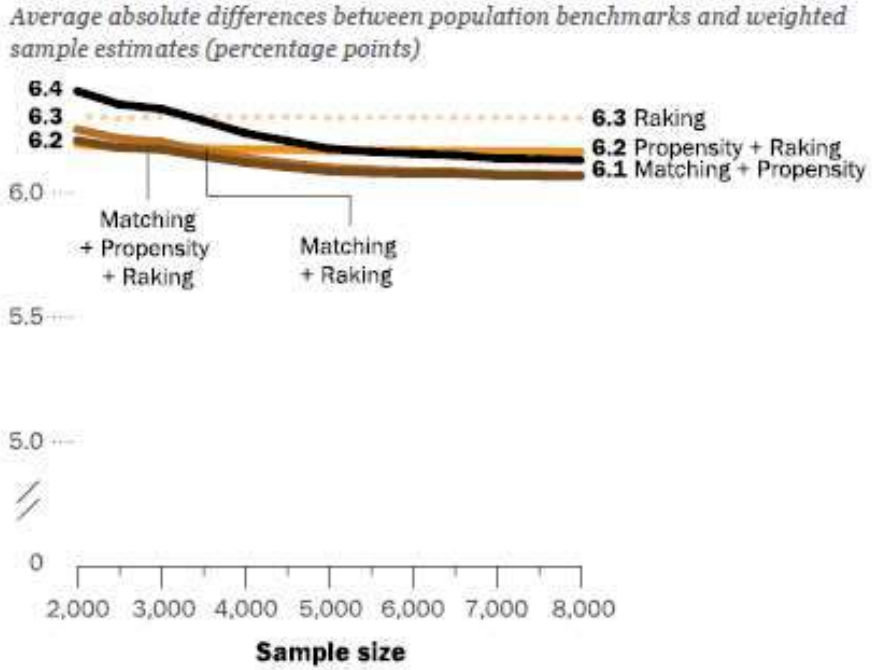
- Doubly robust estimation performs only slightly better than raking (average abs. relbiases of 2.7 and 2.9, respectively)



Source: Valliant (2018)

Valliant “Alternatives for Estimation from Nonprobability Samples”

- Bias reduction seems more about selecting the right adjustment variables than using an elaborate model



Source: Mercer et al. (2018)

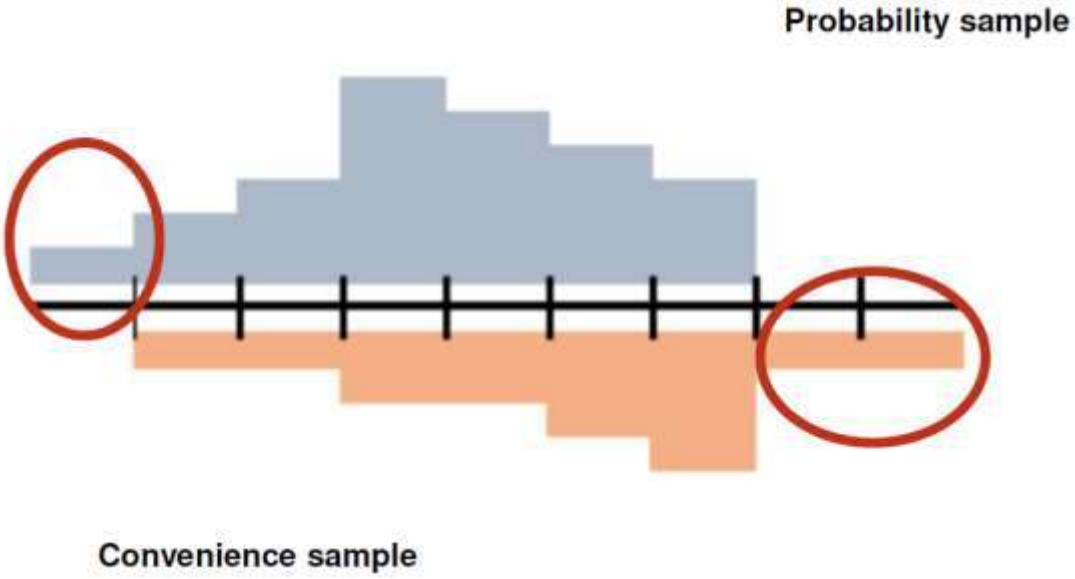
Valliant “Alternatives for Estimation from Nonprobability Samples”

- The main limitation of the study seems to be the data: 2003 MI BRFSS
- Advantage is that full sample estimates can serve as benchmarks
- Disadvantages are several
  - BRFSS, while useful and important, is not necessarily a “gold-standard” data source
  - The non-internet population has changed greatly in 15 years
  - BRFSS data does not capture the selection biases of online opt-in sources

- The theory presented in this paper is very strong, the data less so

Dever: Combining Probability and Nonprobability Samples

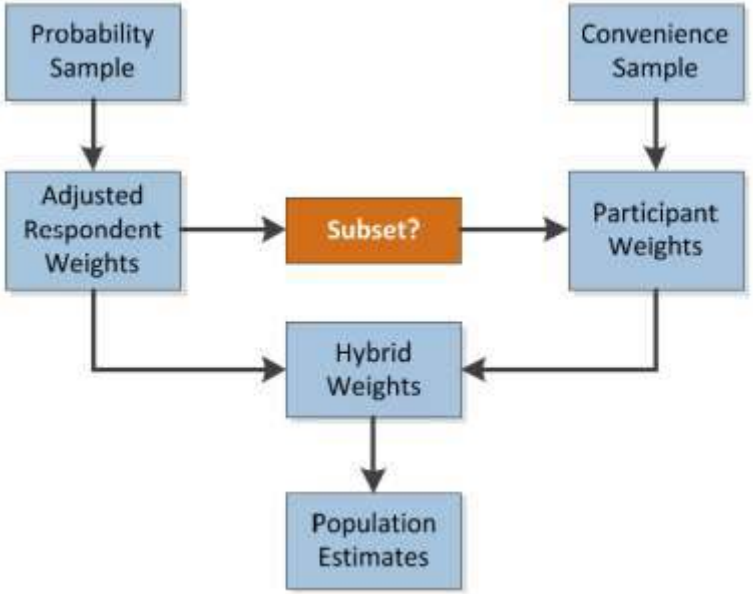
- Nice graphic!



Source: Dever (2018)

Dever: Combining Probability and Nonprobability Samples

- The idea of including only matched ABS cases in the propensity model is intriguing. Interested in more explanation on that point.



Source: Dever (2018)

Dever: Combining Probability and Nonprobability Samples

- If I'm reading this correctly, Facebook users like their weed
- Other findings seem inconclusive. Having just two outcomes is a challenge.
- Why not include benchmark questions from NHIS or NHANES?

	Difference	
Social Media Sample Only	Ever Used	Support Medical Use
No weights	13.4	8.7
Convenience Weights:		
Full ABS sample	-1.2	6.4
92% matched ABS subset	-1.3	6.4
66% matched ABS subset	0.2	7.6

Source: Dever (2018)

Dever: Combining Probability and Nonprobability Samples

- It was not clear (from the slides) how rigorous the ABS survey was
  - Sampling frame?
  - Was oversampling used to address known demographic skews in mail surveys?
  - Incentives?
  - Number of mailings?
  - Response rate?
- Similar to the Valliant paper, the “gold standard” survey seems limited in potentially important ways
- This paper has some interesting ideas and appears to be one component of a larger nonprobability research agenda

Lee “Respondent Driven Sampling”

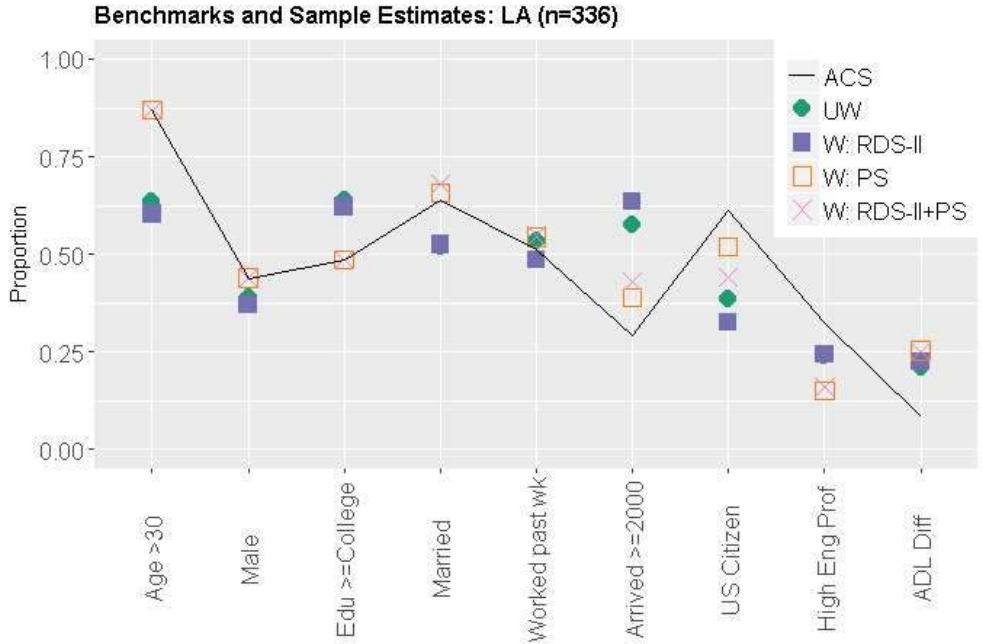
- I applaud Lee’s forthrightness about the challenges with the approach and with the data collected

“Inferences unclear and limited” (slide 27) Amen!

- This paper joins others (e.g., Giles and Handcock 2010) in seriously calling into question the claims made about RDS estimates
- One cannot review these results and put stock in the idea that the estimators are “asymptotically unbiased,” at least in this application

Lee “Respondent Driven Sampling”

- After adjustment, fairly large biases persisted on disabilities, English proficiency, years in US, and citizenship



Source: Lee (2018)

Lee “Respondent Driven Sampling”

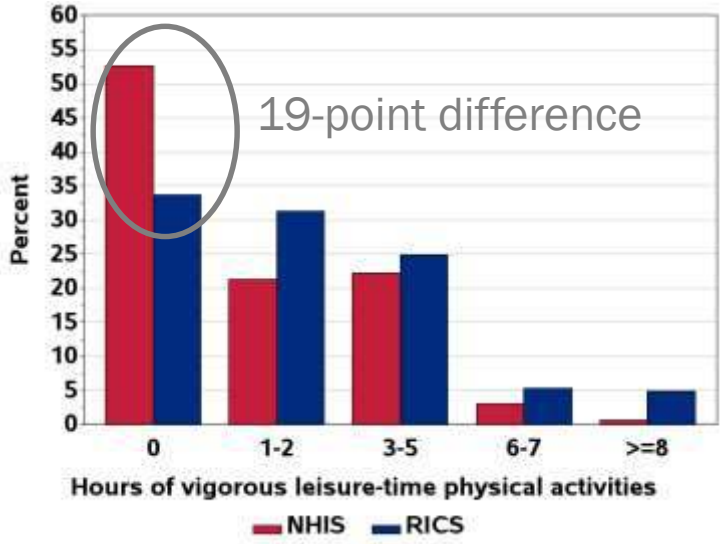
- In this study, the ability to calibrate to the ACS allowed researchers to address 20-point biases on age and education
- What about RDS studies where population benchmarks are not available?
- Might surveys with 20 point biases do more harm than good?
- This study does not answer that question, but it alerts us to the imperative to ask it.

Levine and Krotki “Redirected Inbound Call Sampling”

- A number of different analyses mentioned:
  - Item nonresponse with continuous questions
  - Order effects
  - Prefer Not to Answer prompt
  - Softball questions
  - No Barge In feature
- “31% are night owls” result is curious. What is that parameter in a typical online opt-in survey (e.g., Dever’s social media sample)?

Levine and Krotki "Redirected Inbound Call Sampling"

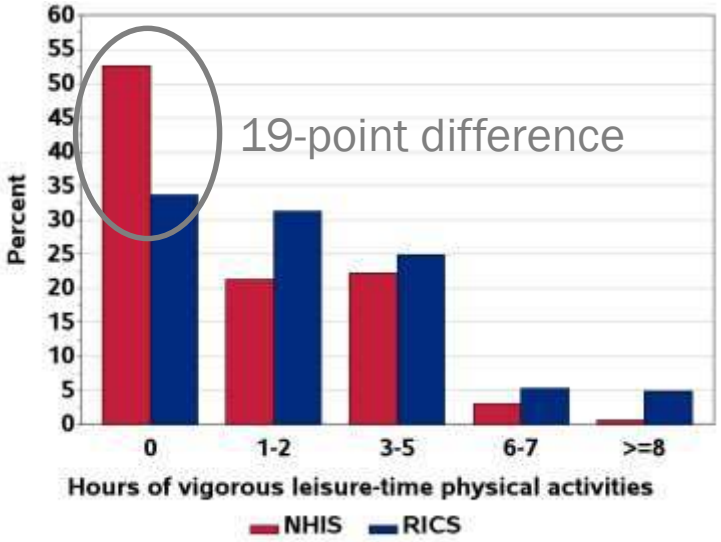
- Some large discrepancies on continuous outcomes



Source: Levine and Krotki (2018)

Levine and Krotki "Redirected Inbound Call Sampling"

- Some large discrepancies on continuous outcomes



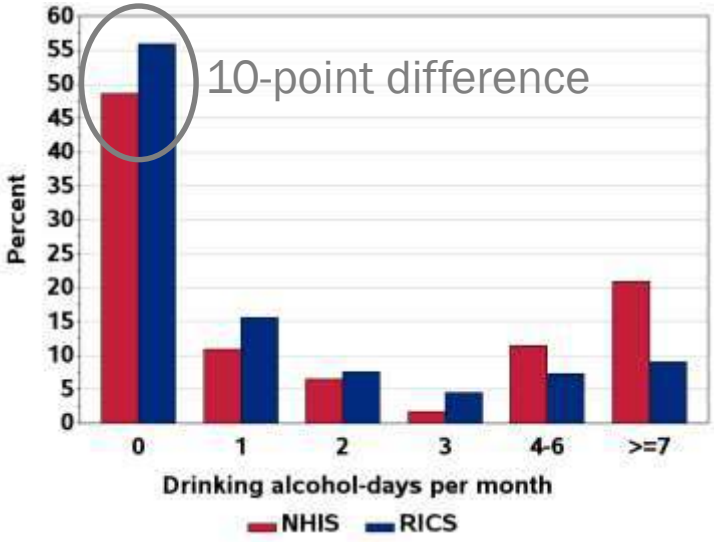
NHIS:
"How often do you do VIGOROUS leisure-time physical activities for AT LEAST 10 MINUTES that cause HEAVY sweating or LARGE increases in breathing or heart rate?"

IVR:
"How many times per week do you do VIGOROUS leisure-time physical activities for AT LEAST 10 MINUTES that cause HEAVY sweating or LARGE increases in breathing or heart rate?"

Source: Levine and Krotki (2018)

Levine and Krotki "Redirected Inbound Call Sampling"

- Some large discrepancies on continuous outcomes



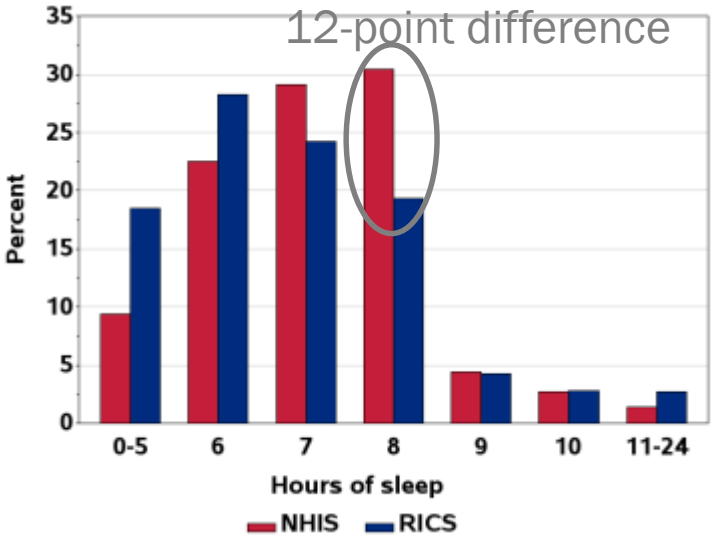
NHIS:
"In the PAST YEAR, how often did you drink any type of alcoholic beverage?"

IVR:
"In the PAST YEAR, on average, how many days per month, did you drink any type of alcoholic beverage?"

Source: Levine and Krotki (2018)

Levine and Krotki "Redirected Inbound Call Sampling"

- Some large discrepancies on continuous outcomes



NHIS:
"On average, how many hours of sleep do you get in a 24-hour period?

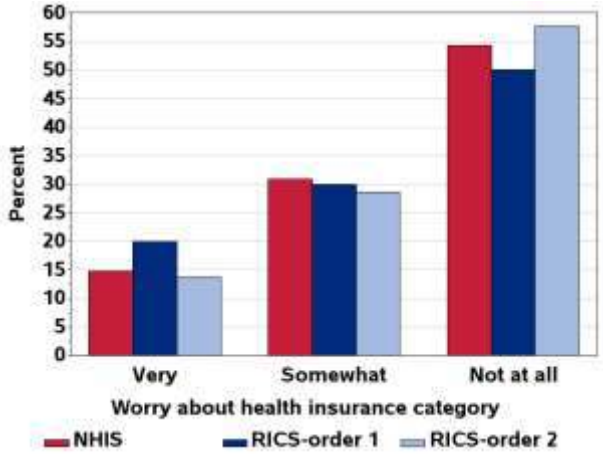
IVR:
"On average, how many FULL hours of sleep do you get in a 24-hour period?"

Source: Levine and Krotki (2018)

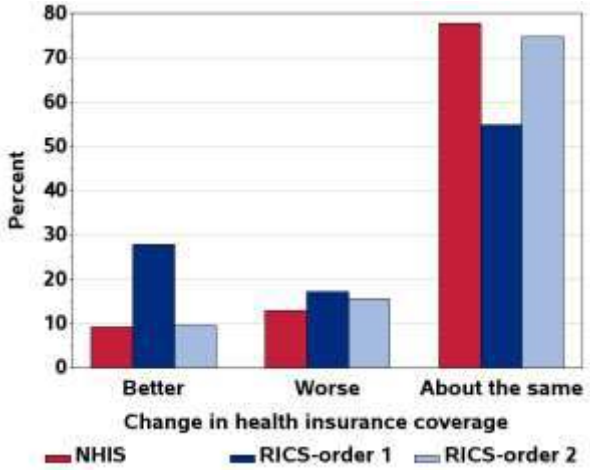
Levine and Krotki "Redirected Inbound Call Sampling"

- Concerns about order effects persist

6-points higher
when read first



18-points higher
when read first



Source: Levine and Krotki (2018)

Levine and Krotki "Redirected Inbound Call Sampling"

- Concerns about order effects persist
- Are we just measuring which answer choice was asked first?

How big a problem is the gap between the rich and poor in our country?	Redirected IVR Study			RDD CATI Study		
	"Very big..." read first	"Very big..." read last	Order effect	"Very big..." read first	"Very big..." read last	Order effect
Very big problem	59	40	19	51	51	0
Moderately big problem	27	26	1	26	29	-3
Small problem	9	20	-11	11	9	2
Not a problem at all	4	13	-9	8	9	-1
	100%	100%		100%	100%	
Interviews	(768)	(787)		(749)	(704)	

Source: Kennedy, Hatley, McGenney (2017)

Levine and Krotki “Redirected Inbound Call Sampling”

- Signs of good progress on “prefer not to answer” option, use of softball items, and “no barge in” feature. Hopefully, a focus of future presentations.
- Authors suggest this is a good fit for surveys of people fleeing a natural disaster or surveillance of gastroenteritis. But what about IRB concerns associated with redirecting telephone calls of people experiencing distress? Wouldn’t these people be urgently trying to reach doctors, caregivers, insurance companies?
- Dipko and Jodts (2017) shed some light on these concerns, but more research is needed to understand incidences of people in distress and people who do not understand that they’ve been redirected

Takeaways

- Each of these researchers deserves credit for testing the boundaries of what we can do with nonprobability surveys
- We should be encouraged by signs of a field coming to agreement on a theory for estimation with nonprobability survey data
- We should also be encouraged by the creativity, humility, and critical analytic skills displayed by these presenters
- We seem to be in a long process of taking the promise of nonprobability approaches, peeling away the hype, and determining what, if any, real value there is to extract.

Thank you