Case Study in the Use of a Probability Panel: NCHS' Research and Development Survey (RANDS) Program

Paul Scanlon, PhD

National Center for Health Statistics

Division of Research and Methodology

Federal Committee on Statistical Methodology
2022 Research and Policy

October 25, 2022

Presentation Overview

- Introduction to RANDS
- Methodological Research Using RANDS
- Special Case: Data Dissemination during COVID-19

RANDS

- NCHS' Division of Research and Methodology started the Research and Development Survey (RANDS), an ongoing set of surveys from commercial panel surveys, in 2015
- Established in response to the growing interest in alternative survey modes and the maturation of relatively inexpensive, probability-sampled commercial survey panels
- Three major goals have driven RANDS throughout the life of the program:
  1. Research into Panel Properties
  2. Supplement and Develop Methods for NCHS' Question/Survey Evaluation Work
  3. (Limited) Production of (Experimental) Estimates

Completed RANDS

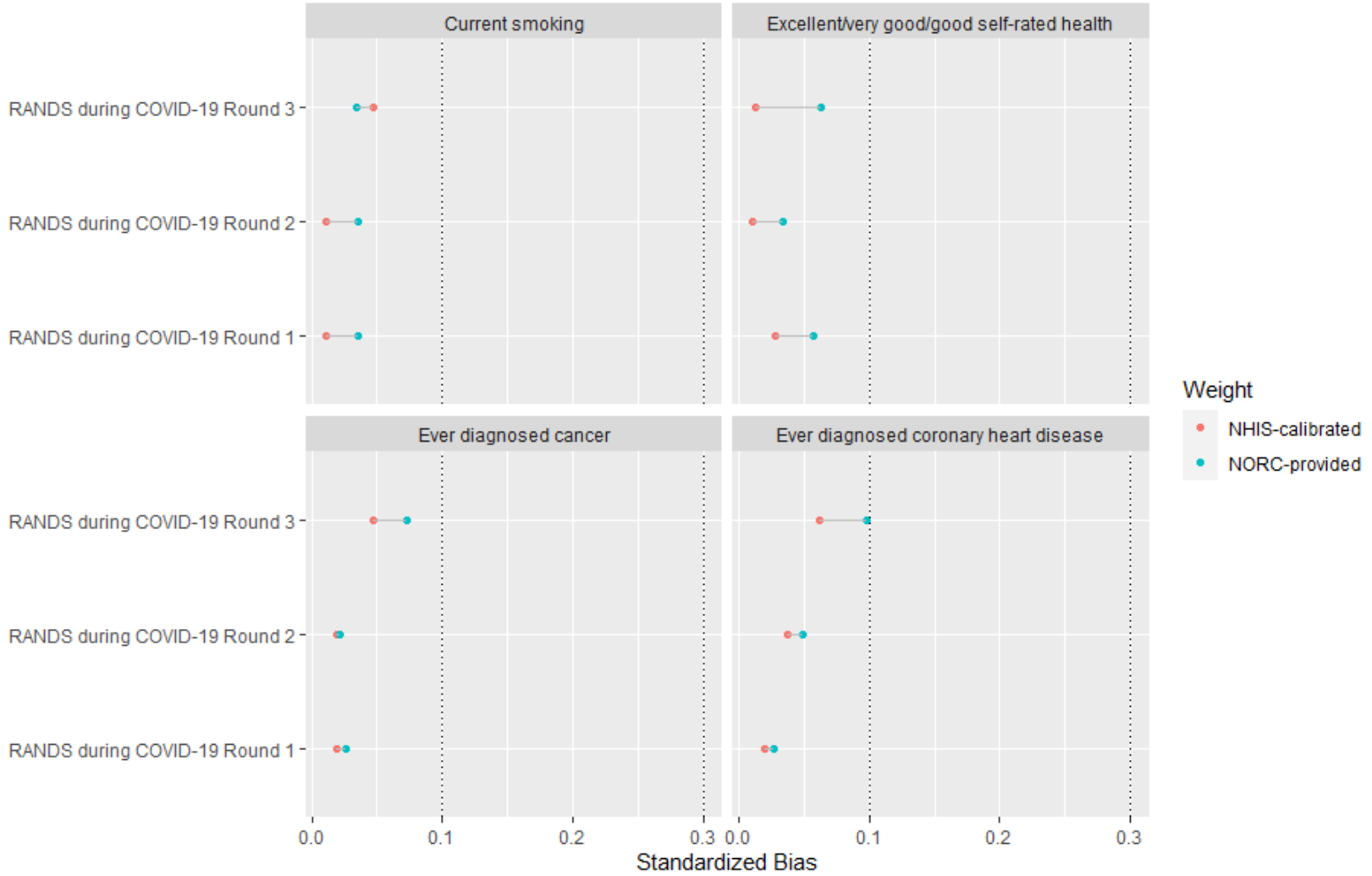
Survey	RANDS 1	RANDS 2	RANDS 3	RANDS 4	RANDS 5	RANDS 6
Year	2015	2016	2019	2020	2022	2022
Panel	Gallup	Gallup	AmeriSpeak	AmeriSpeak	AmeriSpeak	AmeriSpeak
Mode	Web	Web	Web	Web and phone	Web and phone	Web and phone
Survey Focus	Health conditions and behaviors	Health conditions and behaviors	Disabilities and opioids	Disabilities and opioids	Intimate partner violence	NSFG items, gender identity, discrimination
Completed Responses	2,304	2,480	2,646	3,442	6,896	2,312
Completion Rate	23.5%	30.1%	62.2%	70.0%	72.8%	78.3%
Cumulative Response Rate	NA	NA	18.1%	14.0%	11.1%	12.9%

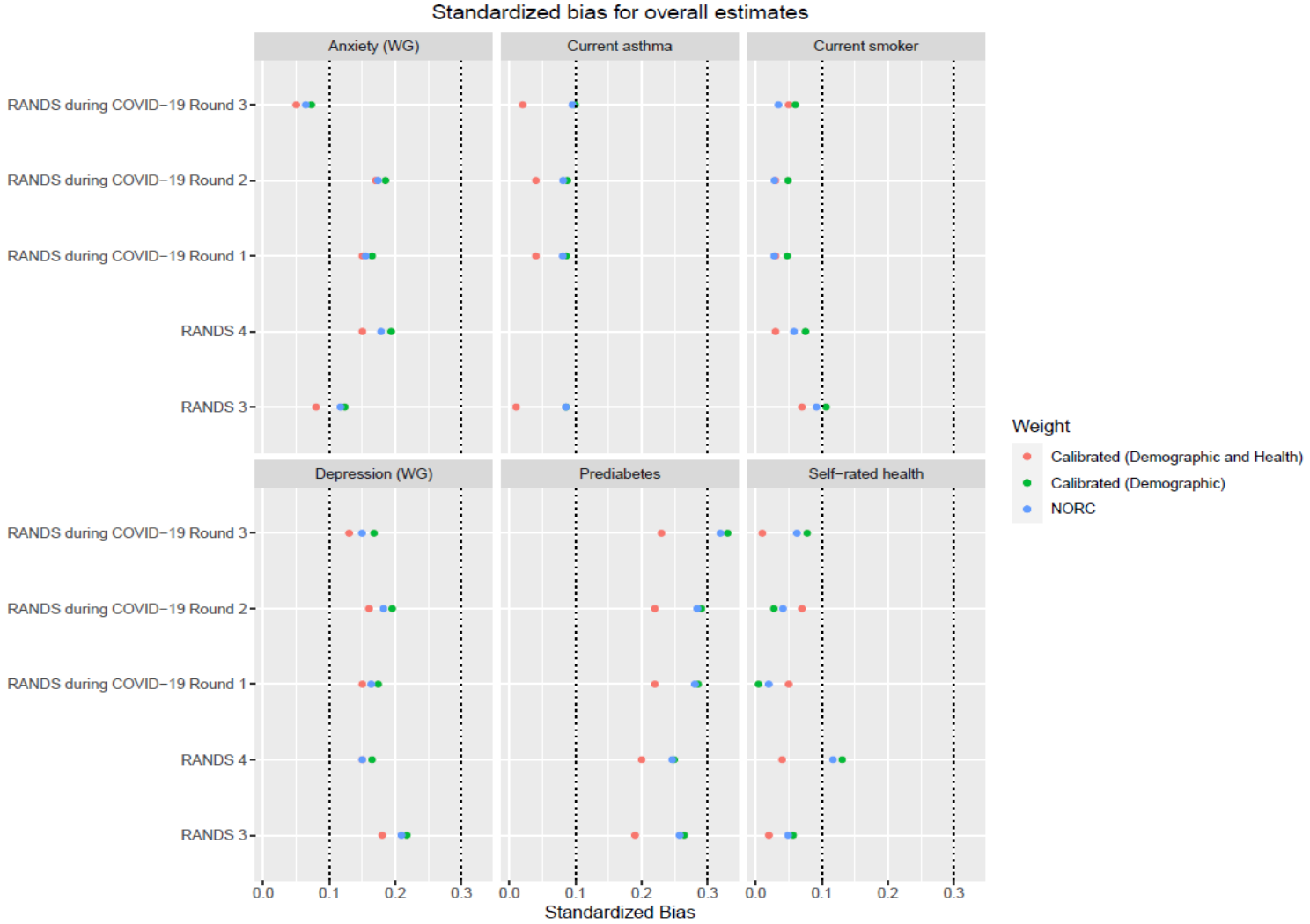
Completed RANDS during COVID-19

Survey	Round 1	Round 2	Round 3
Collection Dates	June 9, 2020–July 6, 2020	August 3, 2020–August 20, 2020	May 17, 2021–June 30, 2021
Completed Surveys	6,800	5,981	5,458
Completion Rate	78.5%	69.1%	69.5%
Cumulative Response Rate	23.0%	20.3%	11.8%

Research into Panel Properties

- Mainly concerned with the *statistical* properties of commercial survey panels
  - Focus is on the "Errors of representation" side of TSE
- Coverage error
  - Impact of including/excluding phone respondents
  - Impact on including hard-to-reach pop—using non-response follow-up flag as proxy
- Non-response error
  - Differential item non-response by mode/panel type/intersectional group
  - Differential response quality (incl item non-response) in open-ended responses
- Adjustment Error
  - Big question: Can we make panel data "look" like data from NCHS' traditional household surveys?
  - Evaluated various approaches, including propensity score and calibrated weighting


Current smoking
Excellent/very good/good self-rated health
Ever diagnosed cancer
Ever diagnosed coronary heart disease
RANDS during COVID-19 Round 3
RANDS during COVID-19 Round 2
RANDS during COVID-19 Round 1
0.0
0.1
0.2
0.3
Standardized Bias
Weight
NHIS-calibrated
NORC-provided


Standardized bias for overall estimates
Anxiety (WG)
Current asthma
Current smoker
Depression (WG)
Prediabetes
Self−rated health
RANDS during COVID−19 Round 3
RANDS during COVID−19 Round 2
RANDS during COVID−19 Round 1
RANDS 4
RANDS 3
0.0
0.1
0.2
0.3
Standardized Bias
Weight
Calibrated (Demographic and Health)
Calibrated (Demographic)
NORC

Question Evaluation Research

- Focus on the “Errors of Measurement” side of TSE
- Both research and “production”
  - Methodological research—focused on developing/refining mixed method approaches to uncovering measurement error
    - “Web probing”—structured, embedded, cognitive probes
    - Experimental design
    - Natural language processing for open-text probe data
  - “Production”—actively using these mixed methods
    - When done after cognitive interviews: expanding/quantifying findings from interviews
    - When done before/alongside cognitive interviews: producing primary qualitative information

In the last two months, has this provider offered you an appointment with a doctor, nurse, or other health professional by video or by phone?

- Yes
- No
- Don't know

PREVIOUS CONTINUE

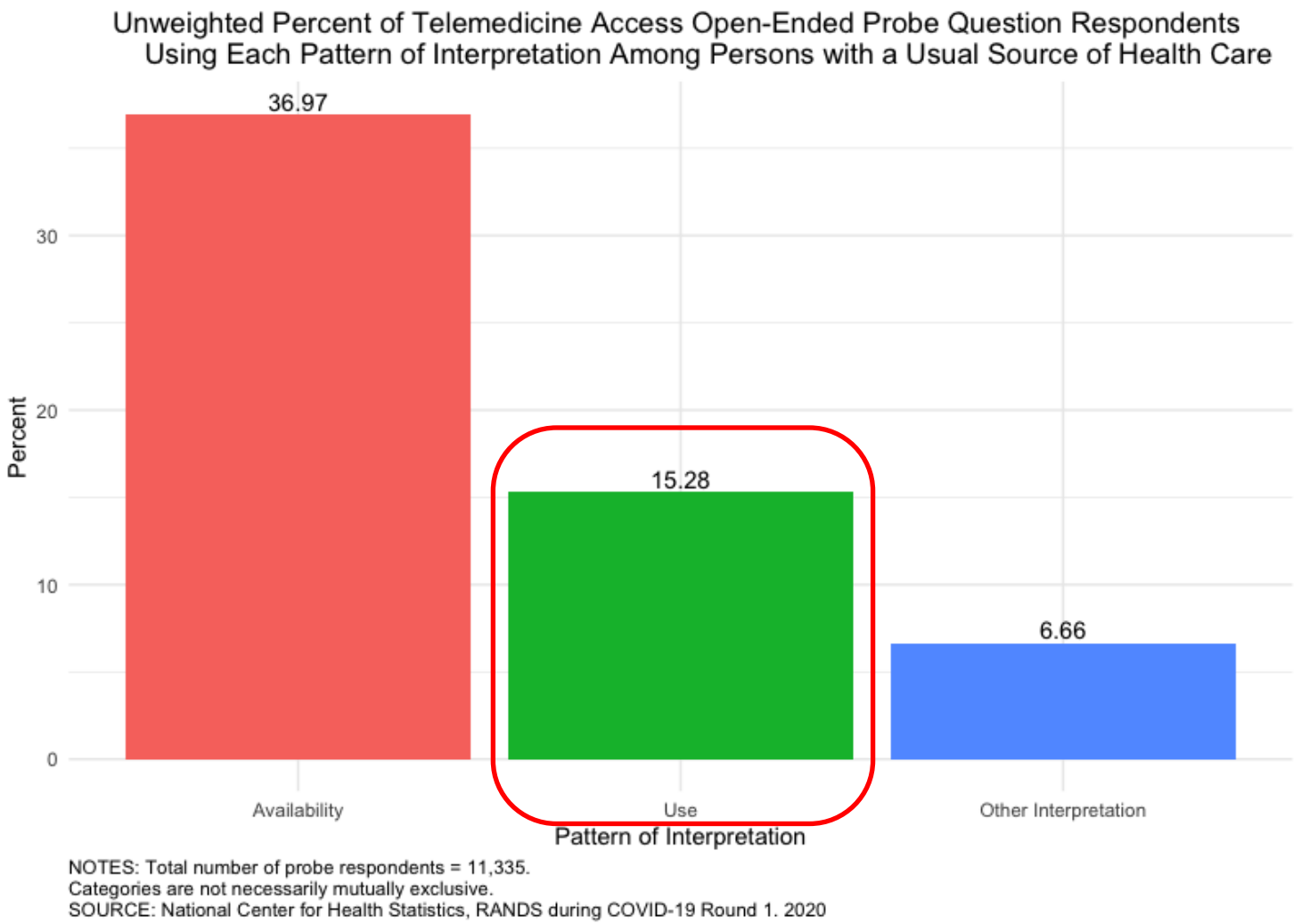
How do you know whether your provider offers telemedicine, or not?

PREVIOUS CONTINUE

n=10,855 non-blank responses…

Analysis of Open-Ended Responses

- Hand coding!
  - Two coding schemes:
    1. Construct captured: telemedicine access or telemedicine use
    2. Potential cases of response error
- Experimenting/Implementing some Machine Learning approaches
  - Data cleaning (i.e., removing PII)
  - Error detection
  - Machine coding



NOTES: Total number of probe respondents = 11,335.
Categories are not necessarily mutually exclusive.
SOURCE: National Center for Health Statistics, RANDS during COVID-19 Round 1. 2020

How do you know whether your provider offers telemedicine or not?

Select all that apply.

- [ ] The provider told you in an email, phone call, or mailing.
- [ ] Had a previous telemedicine appointment
- [ ] Checked provider's website or social media pages
- [ ] Told by a family member
- [ ] Do not know whether the provider offers this
- [ ] Some other place, please specify:

PREVIOUS CONTINUE

Unweighted Percentages (With Standard Errors) Of Eligible RANDS During COVID-19 Round 2 Respondents Using The "Do Not Know" Close-ended Probe Answer Category Only, By Response To The Telemedicine Availability Question

Answer to Telemedicine Availability Question	Eligible Sample	Percent Responding with Only the "Do not know" Close-Ended Probe Answer Category	Standard Error
Yes	2119	6.4	0.5
No	2784	54.1	1.0
Don't Know	219	55.4	3.3

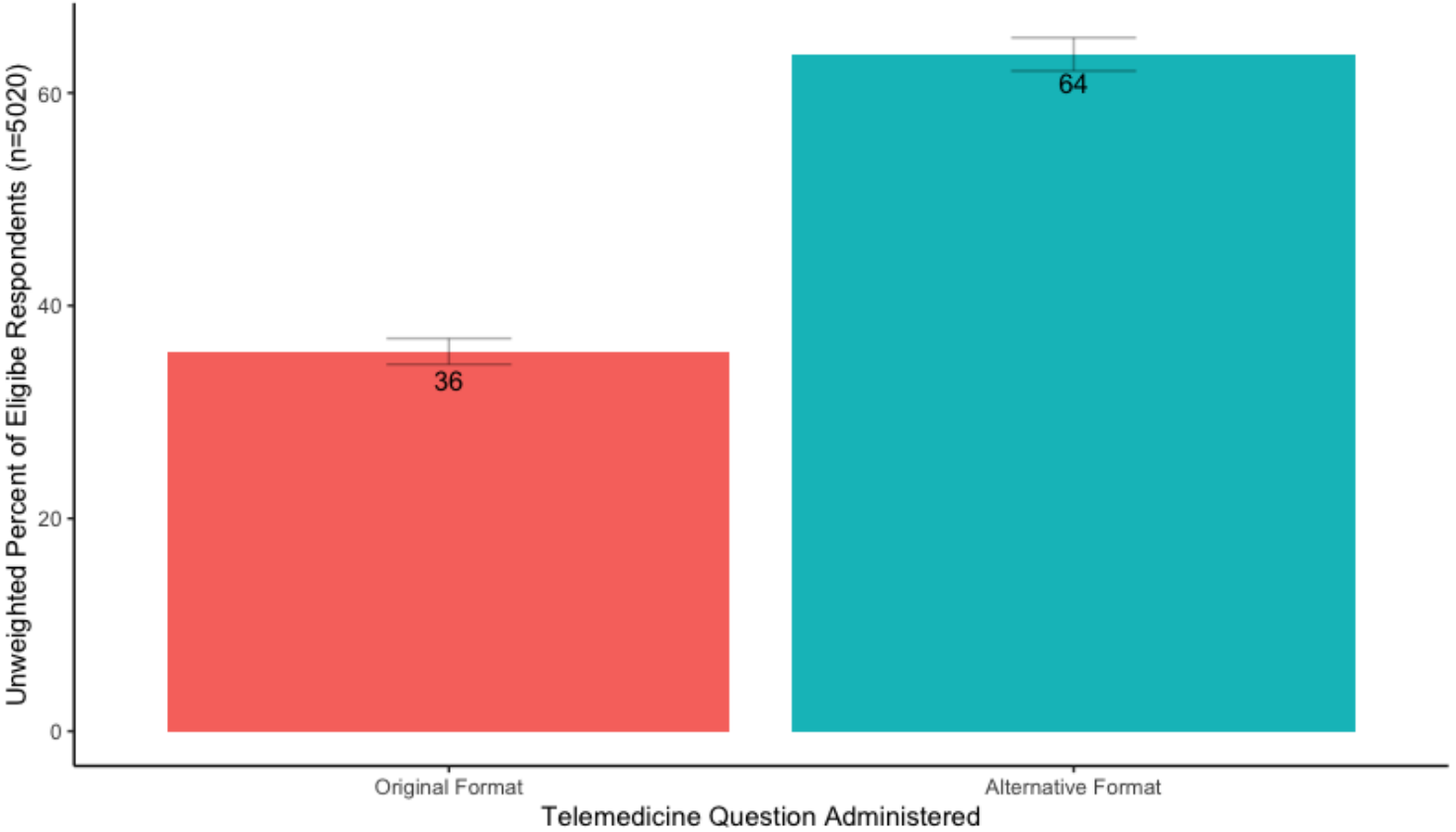
NOTE: Eligible respondents were those respondents assigned to the close-ended probe condition and indicating that they had a usual place of health care, total n = 5,129. There were n=7 instances of item non-response within this eligible group of respondents to the telemedicine availability question; none of those respondents answered the follow-up probe question.
SOURCE: National Center for Health Statistics, RANDS during COVID-19 Round 2. 2020

These are potential false negative responses!

Split Question Experiment in Round 3

- In the last two months, has this provider offered you an appointment with a doctor, nurse, or other health professional by video or by phone? [Yes/No]
- Does this provider offer telephone or video appointments, so that you don't need to physically visit their office or facility? [Yes/No]

Unweighted Percent of Respondents Indicating They Have Access to Telemedicine, by Telemedicine Availability Question Format Among Respondents Who Have a Usual Place of Health Care



NOTES: Assignment to experimental condition: Original format n = 2,518, Alternative format n = 2,502.
F = 357.9, nDF = 1, dDF = 5457, p-value <0.001
SOURCE: National Center for Health Statistics, RANDS during COVID-19 Round 3. 2021

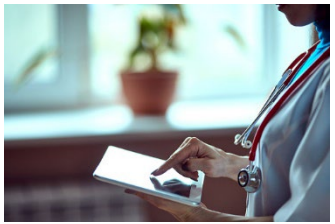
Production of Public Estimates

- Shifted program away from a strictly methodological one in response to the pandemic
  - A number of NCHS' traditional surveys were impacted by stay-at-home orders and need to limit in-person interviews
- Worked with stakeholders to develop a short list of constructs that:
  1. Were not being collected by other federal surveys at the time
  2. Would provide useful information on the public health response and healthcare access
- Conducted and released findings from three special "RANDS during COVID-19" rounds

Published Estimates

Loss of work due to illness
in the previous week because either they or someone in their family was sick with COVID-19

Telemedicine access and use
Access and use in the past 2 months, access prior to the coronavirus pandemic



Reduced access to care
Inability to receive medical care for any reason or due to the coronavirus pandemic

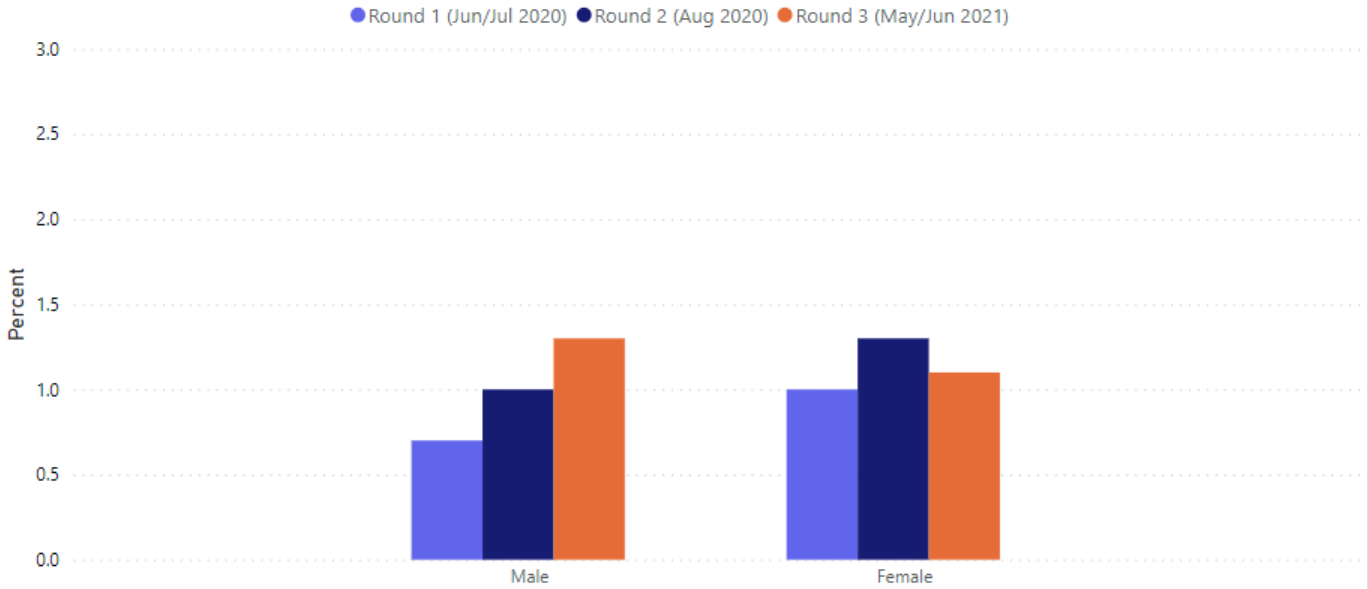
https://www.cdc.gov/nchs/covid19/rands.htm

Loss of work due to illness with coronavirus by age, race and Hispanic origin, sex, education, urbanization, and chronic conditions

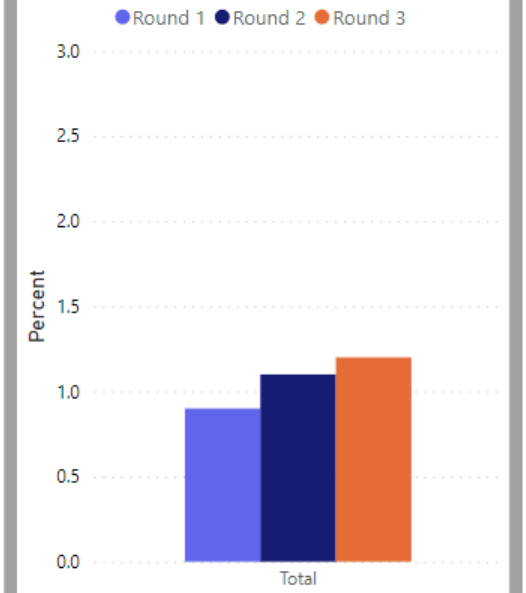
Select Group

Sex

Percentage unable to work in the prior week due to personal or family member illness with coronavirus



Total



NOTES: Due to rounding, the percent (0.02) and standard error (0.02) for the 65 years and over age group for Round 1 are reported as 0.0. See Technical Notes for more information about the content and design of the survey and for more information on the statistical tests used to compare the experimental estimates between rounds. Significance is indicated in the data table only. Estimates that do not meet NCHS standards of reliability are not shown in the bar charts.
SOURCE: National Center for Health Statistics, Research and Development Survey, RANDS during COVID-19, 2020–2021.

Bar Chart | Data Table

Loss of work due to illness with coronavirus by age, race and Hispanic origin, sex, education, urbanization, and chronic conditions

Percentage unable to work in the prior week due to personal or family member illness with coronavirus

Time Period	Round 1 (Jun/Jul 2020)			Round 2 (Aug 2020)			Round 3 (May/Jun 2021)		
Group	Sample Size	Percent	Standard Error	Sample Size	Percent	Standard Error	Sample Size	Percent	Standard Error
Total									
Total	6,794	0.9	0.2	5,974	1.1	0.2	5,450	1.2	0.3
Age group									
18-44 years	2,606	1.5	0.3	2,214	1.9	0.4	1,975	1.9	0.5
45-64 years	2,386	0.5	0.1	2,083	0.6	0.2	1,735	0.8	0.3
65 years and over	1,802	0.0	0.0	1,677	0.1	0.1	1,740	0.4	0.2
Race/Hispanic origin									
White non-Hispanic	4,514	0.5	0.2	4,074	0.3	0.1	3,476	0.4	0.1
Black non-Hispanic	810	2.5	0.6	691	2.6	1.1	762	*	*
Other non-Hispanic	528	0.5	0.3	461	2.1	1.0	393	2.0	1.0
Hispanic	942	1.3	0.6	748	2.8	0.8	819	2.5	1.0
Sex									
Male	2,968	0.7	0.2	2,589	1.0	0.3	2,464	1.3	0.3
Female	3,826	1.0	0.2	3,385	1.3	0.2	2,986	1.1	0.3
Education									
High school graduate or less	1,295	0.8	0.2	1,102	1.2	0.5	1,277	1.7	0.5
Some college	2,552	1.6	0.4	2,226	1.6	0.4	2,579	1.3	0.4
Bachelor's degree or above	2,947	0.2	0.1	2,646	0.6	0.2	1,594	0.4	0.2
Urbanization									
Metropolitan	6,056	0.8	0.1	5,317	1.2	0.2	4,617	1.3	0.3

* Estimate does not meet NCHS standards of reliability.
† The difference from Round 1 is statistically significant at the 0.05 significance level.
‡ The difference from Round 2 is statistically significant at the 0.05 significance level.
NOTES: Due to rounding, the percent (0.02) and standard error (0.02) for the age group 65 years and over for Round 1 are reported as 0.0. See Technical Notes for more information about the content and design of the survey and for more information on the statistical tests used to compare the experimental estimates between rounds. Significance is indicated in the data table only. Estimates that do not meet NCHS standards of reliability are not shown in the bar charts.
SOURCE: National Center for Health Statistics, Research and Development Survey, RANDS during COVID-19, 2020–2021.

Bar Chart | Data Table

RANDS Data

- Tables/charts for select constructs available on NCHS' COVID-19 website:
  - https://www.cdc.gov/nchs/covid19/rands.htm
- Public use data files and documentation are available on the RANDS website
  - https://www.cdc.gov/nchs/rands/data.htm
- Research Data Center (RDC)
  - https://www.cdc.gov/rdc/b1datatype/rdcrands.htm

https://www.cdc.gov/nchs/rands
https://www.cdc.gov/nchs/covid19/rands.htm

Paul Scanlon
pscanlon@cdc.gov

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics, Centers for Disease Control and Prevention.

APPENDIX: Weighting and Calibration

Overview

- RANDS 5 is a probability-based survey that uses a complex sampling design
- NORC provided panel weights and survey design variables (strata and PSU) that account for the sample design of the AmeriSpeak Panel and selection into the RANDS study
- Combining RANDS with high quality reference data can improve estimates of health outcomes and reduce potential selection bias
- To produce the estimates for the public release, NCHS implemented an additional weighting step to calibrate the weights to the NHIS
- Weighted national and subgroup estimates in this presentation were produced using the calibrated weights

NORC-Provided Weights

AmeriSpeak panel weights account for ...

- the **sample design** of the AmeriSpeak Panel
- **differential non-response and under coverage of groups on the sampling frame** by incorporating nonresponse follow up
- **population characteristics** by raking to external population totals using information obtained from the U.S. Census Bureau Current Population Survey (CPS) and the NHIS

RANDS panel weights account for …

- **probability of selection** of sampled panel member into RANDS
- **nonresponse-adjustment** by raking the overall survey sampling weights to adult population totals from the CPS associated with the following socio-demographic characteristics: age, sex, education, race/Hispanic ethnicity and Census Division
- **extreme weights** by trimming weights and re-raking to the same population totals

NHIS-Calibrated Weights

- The NHIS is a population health survey that has been conducted annually by NCHS since 1957 (https://www.cdc.gov/nchs/nhis)
- NHIS collects information on a range of health topics, primarily using personal household interviews
- NHIS interviews are conducted for one adult and one child (if applicable) in each sampled household
- NHIS-calibrated weights were generated by calibrating NORC-provided RANDS panel weights to the NHIS sample adult weights

Raking Procedure

- Raking is an iterative method of adjusting the sample weights to reflect external population counts
- Variables with missing values in RANDS and in NHIS were included in the raking procedure
- Variables with missing values in one of the surveys were removed for the raking step with an adjustment to the weights for the non-missing values
- Calibrated weights were proportionally adjusted to sum to the total number of RANDS respondents for each round

Calibration Variables

Variables
Age
Sex
Race and Hispanic origin
Education
Income
Census region
Marital status
Diagnosed high cholesterol
Diagnosed asthma
Diagnosed hypertension
Diagnosed diabetes
Metropolitan area
Telephone status