Assessing Differences of Mean Estimates from Longitudinal Surveys with an Application to the Research and Development Survey (RANDS)

*Rong Wei, Yulei He, Van Parsons, Paul Scanlon

Division of Research and Methodology
National Center for Health Statistics, CDC

2021 FCSM Research & Policy Conference
November 3, 2021

*Disclaimer: The findings and conclusions in this study are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, the Centers for Disease Control and Prevention.

Motivation

- The National Center for Health Statistics' (NCHS) Research and Development Survey (RANDS) is a series of surveys conducted using commercial probability web survey panels (*https://www.cdc.gov/nchs/rands/index.htm*).

- During the COVID-19 pandemic, special surveys (RANDS during COVID-19) were launched to collect information related to the pandemic (*https://www.cdc.gov/nchs/covid19/rands/telemedicine.htm#limitations*).

- In the summer of 2020, RANDS during COVID-19 had two surveys based on a longitudinal design
  - 6,800 respondents to Round 1 (6/9-7/6) and 5,981 respondents to Round 2 (8/3-8/20). Of these respondents, 5,452 respondents participated in both rounds (80.2% of respondents to Round 1, 91.2% of respondents to Round 2).

- The goal is to estimate the difference of estimates between two rounds while appropriately accounting for the longitudinal design.

Main Statistical Issues

- In longitudinal surveys, observations are collected repeatedly from the same set of subjects so individual structural correlations are induced by these paired samples

- In complex probability surveys, subjects within the same sampling units (e.g., primary sampling units or PSUs) may be subject to an intra-class cluster correlation.

- Survey nonresponses in both rounds can result in partially overlapping samples in terms of a scenario described by Derrick *et al*, 2017.

- Estimating the differences between two rounds of RANDS during COVID-19 needs to account for these issues.

Two Testing Methods Are Developed in This Study

1. Two Extensions of Derrick's t-test:

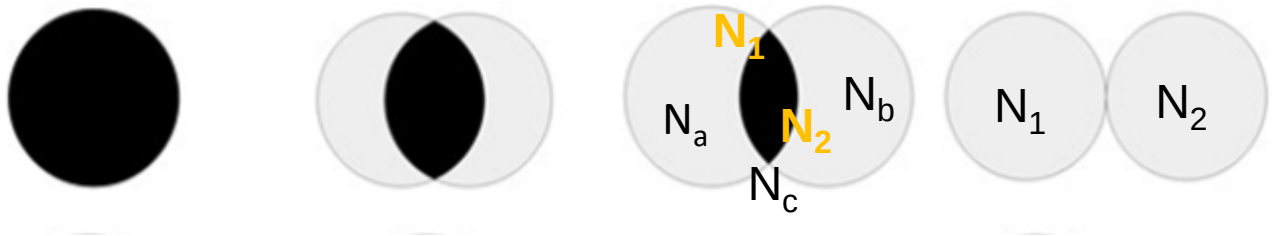
   Extension I – a modified Derrick's t-test assuming equal variance from the two populations by incorporating a complex design structure;

   Extension II – a proposed t-test considering different complex survey design effects for longitudinal samples.

2. Regression method:

   SAS/SURVEYREG procedure using joined longitudinal sampling cluster design information

Method One, t-tests – Notations in Equations



N_a = number of observations exclusive to Sample 1

N_b = number of observations exclusive to Sample 2

N_c = number of overlapping samples

N_1 = total number of observations in Sample 1 i.e., $N_1 = N_a + N_c$

N_2 = total number of observations in Sample 2 i.e., $N_2 = N_b + N_c$

n_i = effective sample size corresponding to N_i above ($n_i = N_i/deff_i$)

P_i = weighted proportion estimate for sample$_i$, i=1,2

$deff_i$ = design effect due to survey stratum, PSU, sample weight, i=1,2

S_1^2 = variance of all observations in Sample 1

S_2^2 = variance of all observations in Sample 2

r_c = Pearson's correlation coefficient for the paired/overlapping observations

t-tests for Two Sample Means

1. For completely paired samples:

$$t = (\hat{p}_D - \hat{\mu}_D) / \sqrt{\frac{S_D^2}{N_D}}$$

D=pairwise difference

2. For completely independent samples: *Welch's t-test*:

$$t = (\hat{p}_1 - \hat{p}_2) / \sqrt{\frac{N_2 S_1^2 + N_1 S_2^2}{N_1 N_2}}$$

3. For partially paired/overlapped two samples: *Derrick's t-test*:

Derrick *et al* (2017) proposed testing methods to incorporate correlation between two partially overlapping samples (next slide).

Note: the corresponding degrees of freedom for *t* are not shown here.

Method One – Adjusted t-test Sampling Errors: Extensions of Derrick's t-test

- For the numerator of the t-test, we use the difference of weighted proportions.
- We modified testing errors by incorporating survey design structures.

$$S_{Derrick\_t} = \sqrt{\frac{(N_1-1)S_1^2+(N_2-1)S_2^2}{(N_1-1)+(N_2-1)}} \bullet \sqrt{\frac{1}{N_1}+\frac{1}{N_2}-2r_c\left(\frac{N_c}{N_1N_2}\right)}$$

$$S_{Ext.I\_t} = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1-1)+(n_2-1)}} \bullet \sqrt{\frac{1}{n_1}+\frac{1}{n_2}-2r_c\left(\frac{N_c}{N_1N_2}\right)}$$

Note: $n_i = N_i/deff_i$; $S_i^2 = p_i(1-p_i)$

Adjusted t-test sampling errors – Based on estimated design effects (Extension II)

This study derived a t-testing error approximation by considering Design Effects from the two survey components and the Correlation for overlapping sampled individuals

$$S^2_{Ext.II\_t} = \widehat{S^2}_{(\hat{p}_1 - \hat{p}_2)} \approx \left(\widehat{S^2}(\hat{p}_1) + \widehat{S^2}(\hat{p}_2)\right) \cdot (1 - f)$$

$$f = 2\frac{N_c\sqrt{\hat{deff}_1 \cdot \hat{deff}_2}}{N_2\hat{deff}_1 + N_1\hat{deff}_2} \cdot \hat{r}_c$$

Method Two – Regression Based on “Survey Round”.

Typically, limited design information leads to variance estimators not explicitly using correlations, but the mathematical expectations frequently will show low-order biases. The regression method concatenates the two samples with design-consistent strata/PSUs labels.

```
PROC Surveyreg;
Strata Strata_ID;
Cluster PSU_ID;
Class  survey_round;
Model outcomes=survey_round;
Weight samplewt;
Run;
```

Case Study: Four Binary Health Variables

1. Self-reported health status (Poor or Fair *vs*. Good+)

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

2. No health insurance coverage (Yes *vs*. No)

    -Are you covered by any kind of health insurance or some other kind of health care plan?

3. Anxiety outcomes (Yes *vs*. No based on a cut-off point of the count of problems)

    -Over the last two weeks, how often have you been bothered by the following problems?

        A. Feeling nervous, anxious, or on edge

        B. Not being able to stop or control worrying

4. Depression outcomes (Yes *vs*. No based on a cut-off point of the count of problems)

    -Over the last two weeks, how often have you been bothered by the following problems?

        A. Little interest or pleasure in doing things

        B. Feeling down, depressed, or hopeless

Results: Percent Estimates

	Outcomes			
Survey	Fair/poor health	No health insurance	Anxiety	Depression
Round 1	13.63%	9.74%	18.43%	16.89%
Round 2	13.04%	10.67%	18.15%	16.96%
Difference (R2 - R1)	-0.59%	0.93%	-0.28%	0.07%

Data source: RANDS during COVID-19, Round 1 and Round 2, 2020

Results: *p-values* from Five t-Tests for a Difference of Proportions

	p-value (two-tailed)			
	Fair/poor health	No health insurance	Anxiety	Depression
Welch's t	0.514	0.354	0.772	0.936
Derrick's t(SRS)	0.128	0.006	0.566	0.846
Ext_I_Derrick's t	0.348	0.122	0.716	0.916
Ext_II_Derrick's t	0.353	0.112	0.711	0.916
Regression	0.298	0.144	0.647	0.939

Data source: RANDS during COVID-19, Round 1 and Round 2, 2020

Discussion - Five Tests Results

- Welch's *t-test* doesn't account for two samples having correlation due to a longitudinal design; the *p-values* are relatively larger than others;
- Derrick's original *t-test* accounts for overlapping samples but doesn't account for complex design with sampling clusters; it gives small *p-values;*
- Both extensions of Derrick's *t-test* developed in this study give very close p-values. In general, the *p*-values are smaller than the ones in Welch's *t-test* due to accounting for correlation between samples, but greater than the ones in Derrick's original *t-test* due to the consideration of survey design features;
- The survey regression method gives *p-values* that differ from ones in extensions of Derrick's *t-test*; it falls between Welch's and Derrick's original *t-tests* because it accounts for the survey design features as well as correlation due to shared clusters; it does not explicitly account for *Pearson correlation* due to overlapping individuals.

Results (details not shown) from a Simulation

A simulation was carried out to explore impacts of parameters on the p-values of *t*, and it was found that smaller *p-values* could be due to:

- Larger effective sample size;
- Greater overlapping sample proportion;
- Greater weighted Pearson correlation r_c ;
- Larger difference of two sample means;
- Smaller variances within samples.

Additional Examples from Round 3 data

- A third survey (Round 3) of RANDS during COVID-19 was launched to collect similar information related to the pandemic as in Round 1 and Round 2.

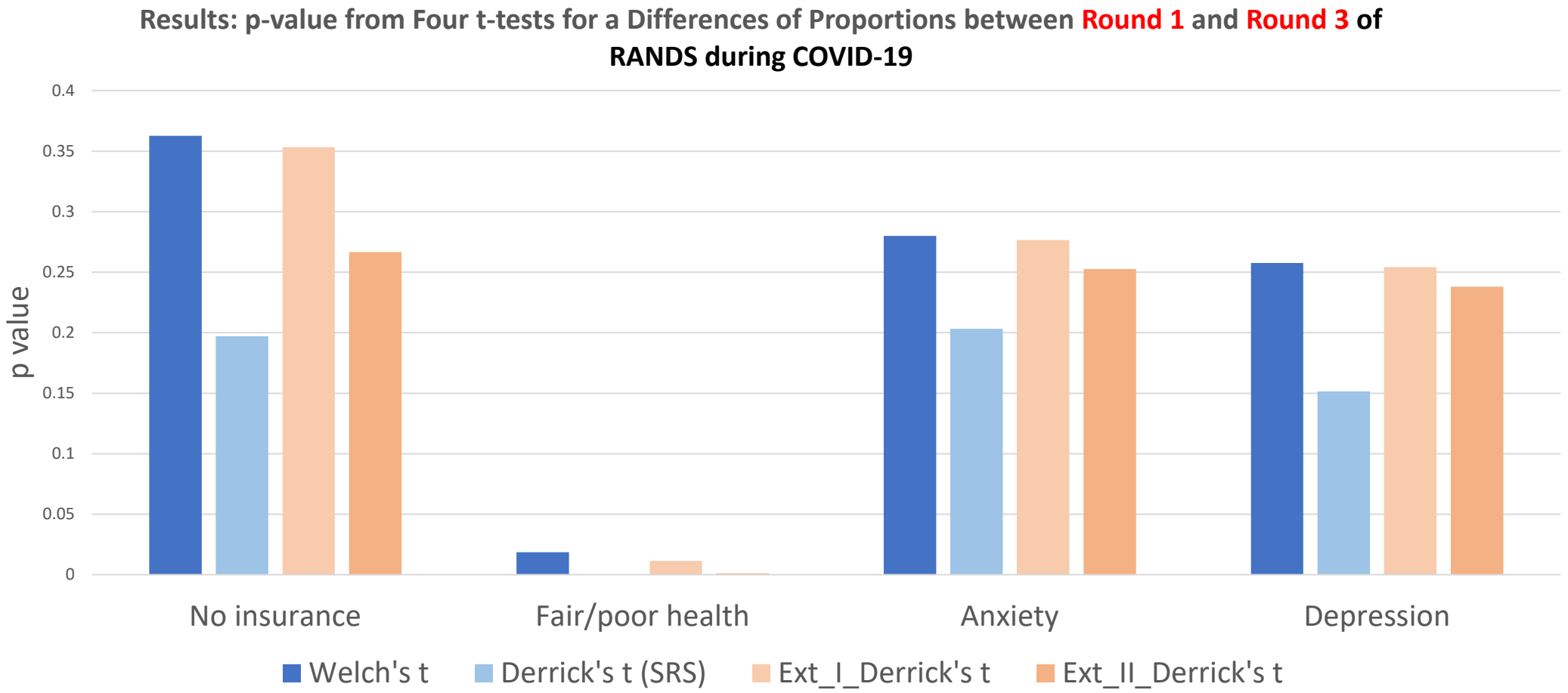
- No individual longitudinal samples for Round 3, but it had 70 out of 159 of PSU clusters sharing the first two rounds.

- Sample intra-class correlation could be a dependency factor in these shared clusters.

- These intra-class correlations between Round 3 and the first two rounds were incorporated into the t-tests for differences.

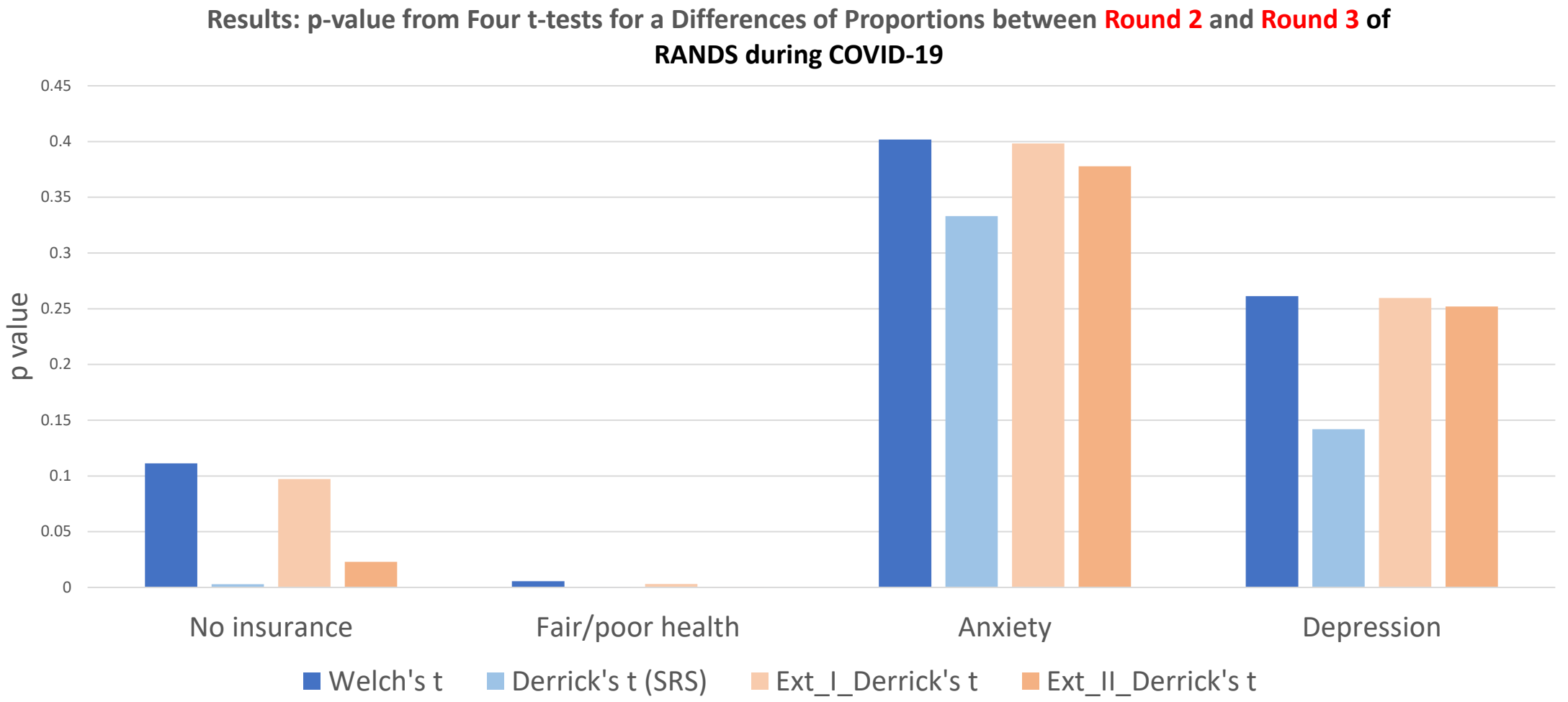
Compare p values of testing differences between Round1 and Round 3 with four t-tests.

		Estimated outcomes %			
Survey	n	No insurance	Fair/poor health	Anxiety	Depression
Round 1	6800	9.74	13.63	18.43	16.89
Round 3	5458	9.39	15.71	17.86	16.21
Difference		0.35	-2.08	0.57	0.68



Compare p values of testing differences between Round 2 and Round 3 with four t-tests.

		Estimated outcomes %			
Survey	n	No insurance	Fair/poor health	Anxiety	Depression
Round 2	5981	10.67	13.04	18.15	16.96
Round 3	5458	9.39	15.71	17.86	16.21
Difference		1.28	-2.68	0.28	0.75



Summary

- Samples from two surveys can be correlated across surveys due to sharing strata/PSUs, overlapping/ longitudinal samples, etc. These correlations should be incorporated to adjust the sampling error for proper statistical inferences on the estimated difference of means.

- Correlation of samples due to design factors including clustering tends to reduce effective sample sizes, thus increasing p-values, but correlated subject responses between two sampled populations tend to decrease sampling errors, therefore obtaining smaller p-values.

- Alternative statistical tests are available for the assessment of mean differences between two surveys with paired/partial overlapping samples. Test selection among candidates depends on underlying assumptions on design features and sample correlations.

- Besides both Extensions of the t-test, which account for both shared design variables and the individual correlation, a regression method was proposed to deal with correlation of two samples by sharing survey clusters even not account for the individual correlation. This method is conservative in that available design information often only allows the use of coarse design-based variance estimation methods. With existing software (SAS/SURVEYREG) the regression test can be easily applied.

References

- Derrick, B., Toher, D., & White, P. 2017. How to compare the means of two samples that include paired observations and independent observations: A companion to Derrick, Russ, Toher and White (2017). The Quantitative Methods for Psychology, 13(2), 120–126. doi:10.20982/tqmp.13.2.p120

- Derrick, B., Russ, B., Toher, D., & White, P. 2017. Test statistics for the comparison of means for two samples that include both paired and independent observations. Journal of Modern Applied Statistical Methods, 16(1), 137-157. doi: 10.22237/jmasm/1493597280

- Irimata, K., Parsons, V., He, Y. 2021. Evaluating differences in estimates of loss of work due to illness, telemedicine access and use, and reduced access to care between RANDS during COVID-19 rounds 1 and 2. Unpublished NCHS report, March 2021.

- Parsons, V. 2021. Research notes: Approximations for comparing round1 and round 2 data release proportions. Unpublished NCHS report, June 2021.

Thank you!

Rong Wei: rrw5@cdc.gov