Distribution Agreement

The text below should be reproduced exactly as written, on the Distribution Agreement page. Sign the page on the signature line and type your name under the signature line. Write the date on the date line.

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Trisha Deshmukh Date

Quantifying reporting bias in social contact data collected from corporate and nursing home employees in the United States

By

Trisha Deshmukh

Master of Public Health

Epidemiology

_____

Ben Lopman, PhD

Committee Chair

_____ [Member's signature]

Moses Chapa Kiti, PhD

Committee Member

Quantifying reporting bias in social contact data collected from corporate and nursing home employees in the United States

By

Trisha Deshmukh

B.S., Virginia Tech, 2020

Thesis Committee Chair: Ben Lopman, PhD

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2023

Abstract

Quantifying reporting bias in social contact data collected from corporate and nursing home employees in the United States

By Trisha Deshmukh

Relevance
There is a current need to obtain direct estimates of contact behavior involved in the transmission of respiratory pathogens. This can be affected by biased estimates of self-assessed behavior or reporting bias. This includes the over and under-reporting of social contacts.

Objective
This study aimed to measure the degree of bias in reporting of contacts from COVID-19 social contact surveys by participant demographic characteristics.

Design
The data used for this study was obtained from the two cross-sectional studies, "Comprehensively Profiling Social Mixing Patterns in Workplace Settings to Model Pandemic COVID-19 and Influenza Transmission and Control" (Corporate Mix (CM)) and "Comprehensively Profiling Social Mixing Patterns in Nursing Homes to Model COVID-19 Transmission" (Nursing Home Mix (NHM)). Data was used from Rounds 2 through 4 of data collection, from November 2020 – December 2021. A logistic regression model was used to examine the relationship between reporting type and participant demographic characteristics from the Corporate Mix study.

Findings
Participants in the Corporate Mix study were more likely to overreport, while those in the Nursing Home Mix study were more likely to have no difference in reporting. The demographic factors – sex, age, race/ethnicity, and education were included in the final model as explanatory variables. We determined that there was a higher likelihood of overreporting among female, younger (20-29 year old), White, Non-Hispanic, and higher educated (Bachelors or higher) participants across all Rounds 2-4 of the Corporate Mix study. An important limitation to this study was that as researchers, we are unable to determine the "true" number of social contacts that a participant has. For this reason, the categorization of over and under-reporting of contacts is arbitrary. Despite this, the results from this study allow us to gain a better understanding of how to design questionnaires to potentially reduce reporting bias in different demographic populations.

Quantifying reporting bias in social contact data collected from corporate and nursing home employees in the United States

By

Trisha Deshmukh

B.S., Virginia Tech, 2020

Thesis Committee Chair: Ben Lopman, PhD

Field Advisor: Moses Chapa Kiti, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Epidemiology

2023

Acknowledgements

I would like to express my deepest gratitude to the mentors who have guided and supported me throughout this research project. I could not have undertaken this journey without my thesis field advisor, Dr. Moses C. Kiti who generously provided knowledge, expertise, and feedback and was patient with me during every step of the process. Additionally, I would like to thank my thesis faculty advisor, Dr. Ben Lopman, for providing valuable guidance and feedback regarding the study analysis plan and composition of this manuscript.

I also want to thank Carol Liu, PhD candidate, for her kind support and technical expertise regarding statistical analysis.

Additionally, this endeavor would not have been possible without the numerous study staff, research assistants, and study participants who tirelessly worked to support the studies that supplied my analytic datasets.

Finally, I would be remiss in not mentioning my family members, classmates, and friends that have supported me as I undertook the journey of finishing my degree and completing this manuscript. Your support means the world to me.

Thank you,

Trisha Deshmukh

Table of Contents

Introduction

Background

Social contact patterns are important for understanding the transmission of airborne and direct-contact infectious diseases such as COVID-19, influenza, and respiratory syncytial virus (RSV). Empirical data can be collected on who mixed with whom, which can be used to model how a disease spreads in a population, who is most at-risk, and the impact of interventions. Models use social contact data to provide information on complex networks that influence infectious disease dynamics[7].

Data on social contact patterns are collected through social contact surveys. The data provides realism to mathematical models they hope to inform and evaluate relevant policy questions.[7] The COVID-19 pandemic has resulted in a myriad of studies to be published that have focused on collecting empirical data on social contact patterns that can help inform mathematical models. Additionally, studies and meta-analyses such as Dodd et. al., have helped to inform future surveys for collecting social contact data relevant for the spread of respiratory pathogens.[4]

There is a need to obtain direct estimates of contact behavior involved in the transmission of respiratory pathogens. This can be difficult as respondents may be subject to inherent biases. A common bias observed in surveys is recall bias. A respondent may forget a contact entirely, especially if the reporting time window is large. While a respondent remembers in their reporting of contacts that an interaction occurred, they may not be able to remember every detail, such as the total duration of their interaction. These details are important as they reflect some aspects of behavior associated with patterns of infection. Respondents who provide self-reported information may also be subjected to reporting bias. It occurs when individuals provide biased estimates of self-assessed behavior.[16] There are many reasons why this may occur, ranging from misunderstanding the question asked to social-desirability bias. Social-desirability bias is where the individual wants to be viewed positively by the researchers and may not

provide accurate responses about their behavior, such as social distancing or mask wearing. It is also influenced by their perceived risk of infection and risk of infecting others.[15] Personal relevance of the survey topic and perceived benefit of participation have also been shown to influence reporting. Reporting also may differ by participant demographic factors such as sex, age, and race and ethnicity. This is an important topic of research as the increased burden of COVID-19 has disproportionately affected older, lower-income, and minority populations.[5] In social contact surveys, reporting bias has the potential to result in the over-reporting or under-reporting of contacts. This will provide a "biased" or inaccurate estimate of how many susceptible individuals may be contracting the illness, therefore influencing disease forecasting and the design of control measures. There has, however, not been an effort in establishing a measure to quantify reporting bias in social contact surveys, which can be used to better understand participant behavior when reporting.

Instead, there have been a few studies conducted to evaluate the effect of different types of reporting bias, such as leading question and social-desirability bias on estimated compliance with COVID-19 guidelines. A study was conducted in Ireland that estimated compliance with COVID-19 guidelines based on the way survey questions were framed.[18] It was identified that estimated compliance depends strongly on how the question is asked. The study also highlighted that social contact surveys tend to pose questions in a way that result in estimates of higher compliance than in other approaches to measure widespread compliance.

Objective

Our study aimed to measure the degree of bias in reporting of contacts from COVID-19 social contact surveys by participant demographic characteristics.

Methods

The data used for our analysis were obtained from two cross-sectional studies, namely "Comprehensively Profiling Social Mixing Patterns in Workplace Settings to Model Pandemic COVID-19 and Influenza Transmission and Control" (Corporate Mix (CM)) and "Comprehensively Profiling Social Mixing Patterns in Nursing Homes to Model COVID-19 Transmission" (Nursing Home Mix (NHM)).

Ethical approval for human subject research was given by Yale University (IRB #2000026906) and Emory University (IRB #00001344), respectively. All participants signed an electronic informed consent form. Participants were provided compensation of a $40 gift card upon completion and submission of the questionnaire. All data were de-identified before analysis.

Corporate Mix Study

The study population consisted of employees from five companies primarily based in Atlanta, Georgia. These companies include professionals that fall under job categories in the education, management, business and financial operations, computer and mathematics, and life, physical, and social sciences sectors as defined by the US Bureau of Labor Statistics (2021)[19].

Five rounds of data collection were conducted between April 2020 – March 23, 2022. This analysis includes data from Rounds 2 (November 2020 – January 2021, n_2=343), 3 (June – August 2021, n_3=376), and 4 (November – December 2021, n_4=433). Round 1 data were not used due to inconsistencies in the coding of data while de-identified data for Round 5 were not available at the beginning of this analysis.

Participants reported, in an online diary, their social contacts over two consecutive workdays. Access to the diary was provided through a personalized email survey link. A contact was defined as either proximate (no conversation and no physical contact but within 6 feet of another individual for 20 seconds or more), conversational (a two-way conversation with three or more words exchanged in the physical presence of another person) or physical (directly touching someone, either skin-to-skin or over the clothes).[7] Encounters with contacts were recorded in the diaries, broken down by individual encounters over only one or both days and total time spent with that person. Participants were requested to first give the total number of people they intended to report as contacts (C_t). Then for each unique contact, participants were asked to report the age (20-29, 30-39, 40-49, 50-59, 60+ years), sex (female or male), race/ethnicity (White, non-Hispanic, Black, non-Hispanic, Asian, non-Hispanic, Mixed, non-Hispanic, Other, non-Hispanic, or Hispanic), and education level (Bachelors or higher, Less than Bachelors) of each contact. We refer to the sum of contacts actually completed in the diary as reported contacts (C_r).

Analyses were performed using R v.4.1.2. All code and data were made available on GitHub.

The outcome variable in this analysis was reporting type – underreported, no difference in contacts, and overreported. Reporting type was classified based on the difference between total contacts (C_t) and reported contacts (R_t). While this terminology is used for this study, there is no "gold standard" or correct number of contacts. However, for these classifications, we assume reported contacts as the standard.

$$\boldsymbol{Overreported}\ (\boldsymbol{C\_o}): total\ contacts(C\_t) > reported\ contacts\ (C\_r)$$

$$\boldsymbol{No\ difference\ in\ contacts}\ (\boldsymbol{C\_n}): total\ contacts(C\_t) = reported\ contacts\ (C\_r)$$

$$\boldsymbol{Underreported}\ (\boldsymbol{C\_u}): total\ contacts(C\_t) < reported\ contacts\ (C\_r)$$

We used a logistic regression model to analyze reporting type for the Corporate Mix study and dichotomized to only include those who `overreported` and had `no difference in reporting`. Those who `underreported` made up 1% of total participants and were not included in the analysis. A causal framework was constructed and the forementioned socio-demographic factors - sex, age, race/ethnicity, and education were selected as explanatory variables. Logistic regression models were created to perform multivariate assessments across each of the study rounds using the explanatory variables.
A similar analysis was not conducted on results from the Nursing Mix Home study due to low statistical power.

$$\boldsymbol{logit}\left(\frac{\boldsymbol{p}}{\boldsymbol{1-p}}\right) = \ \mathbf{ln(odds\ of\ OVERREPORTING)}$$

$$= \boldsymbol{\alpha} + \boldsymbol{\beta_1}\,\mathbf{RE} + \boldsymbol{\beta_2}\,\mathbf{EDU} + \boldsymbol{\gamma_1}\,\mathbf{SEX} + \boldsymbol{\gamma_2}\,\mathbf{AGE} + \boldsymbol{\delta_1}\,\mathbf{RE\ x\ AGE} + \boldsymbol{\delta_2}\,\mathbf{EDU\ x\ AGE}$$

As similarly done in previous studies, a confounding assessment was conducted for age and sex. The variable age was also tested for interaction with the other predictors for difference in reporting between younger and older adults.[4, 6-7]

Nursing Home Mix Study

Similar methods for data collection and analysis were used in this study, however, there were a few key differences. The study population was long-term care facility (LTCF) staff in four nursing homes in the US. A total of 157 LTCF staff completed both days of the contact diary and were included in this analysis. Data was collected from December 2020 to June 2021.

Results

Overall, 1,152 participants from the Corporate Mix study and 157 participants from the Nursing Mix study were included in this analysis. In the Corporate Mix study, the number of participants between each round increased slightly, with 343 in round 2, 376 in round 3, and 433 in round 4. Of the total 1,152 participants from the Corporate Mix study, most participants were in the 20-29 (31%) and 30-39 (29%) age groups. In the Nursing Mix study, most participants were comprised of 40-49 (52%) and 30-39 (40%) year olds. Most participants in the Corporate Mix study self-identified as female (65%) compared to male (35%), while in the Nursing Mix study, there was an equal percentage of males and females (50%). Participants in the Corporate Mix study identified as White, non-Hispanic (59%), followed by Asian, non-Hispanic (20%), Black, Non-Hispanic (9%), Hispanic (7%), Other, Non-Hispanic (2%), and Mixed, Non-Hispanic (2%). In the Nursing Mix study, the largest percentage of participants self-identified as White, Non-Hispanic (78%), followed by Black, non-Hispanic (13%), and Hispanic (9%). One participant identified as Asian, non-Hispanic (0.6%). Finally, in both the Corporate and Nursing mix studies, most participants reported a bachelor's degree or higher (87% and 82%). A summary of these results can be found in Table 1.

The number of participants and percentages who fall under each category of reporting type were compared across Rounds 2-4 and Overall in the Corporate Mix study and only Overall in the Nursing Mix study. Overall, in the Corporate Mix study, the highest percentage of participants overreported (69%), followed by no difference in reporting (30%). In the Nursing Home Mix study, there was equal but low percentages of participants who underreported (9%) and overreported (8%). The majority of participants had no difference in reporting (83%). There was overall a greater percentage of participants who overreported in the Corporate Mix study and who had no difference in reporting in the Nursing Home Mix study (Table 2). Additionally, there was a substantially higher number of participants who overreported in Round 2 (83%) compared to Rounds 3 (62%) and 4 (64%) of the Corporate Mix study.

The outcome variable (reporting type) was compared by study round and demographic characteristics (Table 3) in the Corporate Mix study and just by demographic characteristics in the Nursing Home Mix study (Table 4). Starting with comparing the unstratified results by age, there was low underreporting in all age groups (1%). No difference in reporting was greatest in 30–39-year-olds (32%), followed by 20–29-year-olds (31%) and 40–49-year-olds (30%). When comparing counts, overreporting was highest in 20–29-year-olds (n = 239), however, percentagewise overreporting was highest among 50+ year olds (74%). This was followed by 20-29- and 40–49-year-olds (68%) and 30–39-year-olds (66%). There was low underreporting in both males and females (1%). The highest counts of participants in both sexes overreported, with 70% females and 66% males overreporting. Among the different participant racial/ethnic groups, the highest counts observed were among those who overreported, with 72% being White, Non-Hispanic, 70% Black, Non-Hispanic, and 68% Mixed, Non-Hispanic. Finally, among the two education levels, there were higher counts of overreporting, with 68% having a Bachelors or higher and 72% having less than a Bachelors. In the Nursing Mix Home study, across all demographic factors, most participants had no difference in reporting. The highest percentages observed were in 30–39-year-olds (87%) and 40–49-year-olds (85%). These two age groups also made up 92% of the study participants. No difference in reporting varied by sex with 88% being male and 77% being female. White, Non-Hispanic (80%) had the highest percentage of participants who reported no difference in reporting compared to their counterparts.

Model results

The demographic factors – sex, age, race/ethnicity, and education were included in the final model as explanatory variables. We determined that there was a higher likelihood of overreporting among female, younger (20–29-year-old), White, Non-Hispanic, and higher educated (Bachelors or higher) participants

across all Rounds 2-4 of the Corporate Mix study. The odds ratios and confidence intervals used to make these determinations are provided in Table 5.

Discussion

Key Findings

Several meaningful associations were observed between participant demographic characteristics in the Corporate and Nursing Home Mix studies and reporting type. Overreporting was the most common reporting type among CM participants, across all study rounds and participant demographic characteristics. In contrast, NHM participants were more likely to report no difference in contacts. Participants who were female, 20-29 years old, White, Non-Hispanic, or had a Bachelors or higher had a higher likelihood of overreporting compared to other participants in the CM study. These results remained consistent across all rounds for participant sex, age, and education level. This was also observed amongst White, Non-Hispanic participants.

Interpretations and Implications

Based on how the questionnaire was structured, some participants reported more contacts than they intended or said they would report. In this case, we considered that they overreported contacts. If the participants under-reported, it was because they did not finish providing information for all their contacts. This is likely because they did not have time to finish the survey, were tired, or lost interest. There may be multiple explanations of why a participant over-reported contact. First, they may have needed to be probed further. By completing the first section of questions in full, this may have helped them remember more contacts, thus reducing recall bias, especially for those who did not use the memory aid provided to record contacts over the two days of data collection.

As identified through a national probability survey of US households (the COVIDVu Study) [15], younger participants (aged 25-34) reported higher numbers of daily contacts and increased their contact rates the most from baseline to follow-up. This is consistent with data collected in other studies that support that younger participants are less likely to comply with mitigation behaviors[15] and therefore, accumulate more social contacts.[1, 9-10] Not as many studies have investigated how social contact reporting varies by sex/gender and education level. However previous studies have identified trends in who is more likely to report information in surveys, in general. More educated and affluent people are likely to report information in surveys. Additionally, women are more likely to report than men. However, it is important to note the existence of gender bias in online survey response behavior, where observed differences in reporting are a product of how females and males undergo social exchange (Smith, G., 2008).

Limitations

There are a few limitations to take into consideration. We are unable to determine the "true" number of social contacts that a participant has. There is no "gold standard" or correct number of contacts. For this reason, the categorization of over and under-reporting of contacts is arbitrary. Additionally, the cross-sectional design allowed for testing of associations between predictor and outcome variables at a specific point in time, but longitudinal predictions cannot be made. It is also important to consider that respondents tended to be highly educated and majority white individuals working for private companies and nursing homes primarily based in Georgia. Since the overall US workforce is diverse and is made up of a myriad of occupations, the study samples from the Corporate Mix and Nursing Home Mix studies cannot be used to represent all occupations, departments, and individuals of the US workforce. Despite these limitations, our findings that certain demographic groups report more contacts were still similar to findings in previous studies.

Recommendations

Further studies and initiatives need to be taken to better understand why certain demographic groups, such as men and individuals with lower education are less likely to report more contacts. Answering this question informs researchers how populations respond differently to surveys compared to their counterparts. While the reason is currently unknown, survey design can be improved to make groups less likely to report feel comfortable and inclined to share their social contacts. There are some strategies that can be incorporated for participants to recall more contacts, which appeared to help in this study design. One would be to ask participants to report their total contacts and give them another opportunity in the survey to report additional contacts they may have forgotten. Additionally, provide participants with a memory aid that they can print and fill out as they come across other individuals during their data collection period. Finally, allow participants to provide de-identified information about their social contacts, to alleviate worries about providing identifying information.

Conclusion

By establishing a quantitative measure for reporting bias in contact surveys, we found that reporting varies primarily based on participant demographic characteristics in the Corporate Mix study. Being female, 20-29 years old, White or Asian, Non-Hispanic race and ethnicity, and having a Bachelors degree or higher are predictors of overreporting contacts.

References

(1) Coroiu, A., Moran, C., Campbell, T., & Geller, A. C. (2020). Barriers and facilitators of adherence to social distancing recommendations during COVID-19 among a large international sample of adults. *PLoS ONE*, *15*(10). https://doi.org/10.1371/JOURNAL.PONE.0239795

(2) Costantini, L., Nwafor, C., Lorenzi, S., Marrano, A., Ruffa, P., Moreno-Sanz, P., … Grando, M. S. (2008).

Does gender influence online survey participation?: A record-linkage analysis of university faculty online survey response behavior. *ERIC Document Reproduction Service No. ED 501717*, (22), p.43. https://doi.org/10.4/JQUERY-UI.MIN.JS

(3) Datta, S., Mercer, C. H., & Keeling, M. J. (2018). Capturing sexual contact patterns in modelling the spread of sexually transmitted infections: Evidence using Natsal-3. *PLoS ONE*, *13*(11). https://doi.org/10.1371/JOURNAL.PONE.0206501

(4) Dodd, P. J., Looker, C., Plumb, I. D., Bond, V., Schaap, A., Shanaube, K., … White, R. G. (2016). Editor's choice: Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. *American Journal of Epidemiology*, *183*(2), 156. https://doi.org/10.1093/AJE/KWV160

(5) Escobar, G. J., Adams, A. S., Liu, V. X., Soltesz, L., Chen, Y. F. I., Parodi, S. M., … Lee, C. (2021). Racial Disparities in COVID-19 Testing and Outcomes: Retrospective Cohort Study in an Integrated Health System. *Annals of Internal Medicine*, *174*(6), 786–793. https://doi.org/10.7326/M20-6979

(6) Fava, E. Del, Cimentada, J., Perrotta, D., Grow, A., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). The differential impact of physical distancing strategies on social contacts relevant for the spread of COVID-19: Evidence from a multi-country survey. *MedRxiv*, 2020.05.15.20102657. https://doi.org/10.1101/2020.05.15.20102657

(7) Hoang, T., Coletti, P., Melegaro, A., Wallinga, J., Grijalva, C. G., Edmunds, J. W., … Hens, N. (2019). A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. *Epidemiology*, *30*(5), 723–736. https://doi.org/10.1097/EDE.0000000000001047

(8) Horton, K. C., Hoey, A. L., Béraud, G., Corbett, E. L., & White, R. G. (2020). Systematic Review and Meta-Analysis of Sex Differences in Social Contact Patterns and Implications for Tuberculosis Transmission and Control - Volume 26, Number 5—May 2020 - Emerging Infectious Diseases journal - CDC. *Emerging Infectious Diseases*, *26*(5), 910–919. https://doi.org/10.3201/EID2605.190574

(9) Hutchins, H. J., Wolff, B., Leeb, R., Ko, J. Y., Odom, E., Willey, J., … Bitsko, R. H. (2020). COVID-19 Mitigation Behaviors by Age Group — United States, April–June 2020. *Morbidity and Mortality Weekly Report*, *69*(43), 1584. https://doi.org/10.15585/MMWR.MM6943E4

(10) Ingram, M., Zahabian, A., & Hur, C. (2021). Prediction of COVID-19 Social Distancing Adherence (SoDA) on the United States county-level. *Humanities and Social Sciences Communications 2021 8:1*, *8*(1), 1–7. https://doi.org/10.1057/s41599-021-00767-0

(11) Kiti, M. C., Aguolu, O. G., Zelaya, A., Chen, H. Y., Ahmed, N., Battross, J., … Lopman, B. A. (2022). Changing social contact patterns among US workers during the COVID-19 pandemic: April 2020 to December 2021. *MedRxiv*, 2022.12.19.22283700. https://doi.org/10.1101/2022.12.19.22283700

(12) McCreesh, N., Mohlamonyane, M., Edwards, A., Olivier, S., Dikgale, K., Dayi, N., … Middelkoop, K. (2022). Improving Estimates of Social Contact Patterns for Airborne Transmission of Respiratory Pathogens - Volume 28, Number 10—October 2022 - Emerging Infectious Diseases journal - CDC. *Emerging Infectious Diseases*, *28*(10), 2016–2026. https://doi.org/10.3201/EID2810.212567

(13) Mossong, J. L., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., … Riley, S. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine | Www*, *5*. Retrieved from http://epp.eurostat.ec.europa.eu/

(14) Mousa, A., Winskill, P., Watson, O. J., Ratmann, O., Monod, M., Ajelli, M., … Whittaker, C. (2021). Social contact patterns and implications for infectious disease transmission: A systematic review and

meta-analysis of contact surveys. *ELife*, *10*. https://doi.org/10.7554/ELIFE.70294

(15) Nelson, K. N., Siegler, A. J., Sullivan, P. S., Bradley, H., Hall, E., Luisi, N., … Lopman, B. A. (2022). Nationally representative social contact patterns among U.S. adults, August 2020-April 2021. *Epidemics*, *40*, 100605. https://doi.org/10.1016/J.EPIDEM.2022.100605

(16) Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural & Healthcare Research*, *2*(4), 320. https://doi.org/10.1504/IJBHR.2011.043414

(17) Sonnenberg, P., Clifton, S., Beddows, S., Field, N., Soldan, K., Tanton, C., … Johnson, A. M. (2013). Prevalence, risk factors, and uptake of interventions for sexually transmitted infections in Britain: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *Lancet (London, England)*, *382*(9907), 1795. https://doi.org/10.1016/S0140-6736(13)61947-9

(18) Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2021). It depends on how you ask: measuring bias in population surveys of compliance with COVID-19 public health guidance. *J Epidemiol Community Health*, *75*(4), 387–389. https://doi.org/10.1136/JECH-2020-215256

(19) U.S. Bureau of Labor Statistics. (n.d.). Retrieved March 9, 2023, from https://www.bls.gov/

(20) Wallinga, J., Teunis, P., & Kretzschmar, M. (2006). Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *American Journal of Epidemiology*, *164*(10), 936–944. https://doi.org/10.1093/AJE/KWJ317

(21) Yoon, P., Hall, J., Fuld, J., Mattocks, S. L., Lyons, B. C., Bhatkoti, R., … Pillai, S. K. (2021). Alternative Methods for Grouping Race and Ethnicity to Monitor COVID-19 Outcomes and Vaccination Coverage. *MMWR. Morbidity and Mortality Weekly Report*, *70*(32), 1075–1080. https://doi.org/10.15585/MMWR.MM7032A2

Tables and Figures

Table 1. Baseline characteristics of the study population, by study round - November 2020 – December 2021 - Corporate and Nursing Home Mix.

		Corporate Mix	Rounds			Nursing Home Mix
		Two (N = 343)	Three (N = 376)	Four (N = 433)	Overall (N = 1,152)	Overall (N = 157)
Age Group (years)						
	20-29	87 (25%)	120 (32%)	145 (33%)	352 (31%)	5 (3%)
	30-39	109 (32%)	104 (32%)	124 (29%)	337 (29%)	63 (40%)
	40-49	86 (25%)	80 (21%)	94 (22%)	260 (23%)	82 (52%)
	50-59	39 (11%)	56 (15%)	57 (13%)	152 (13%)	7 (5%)
	60+	22 (6%)	16 (4%)	13 (3%)	51 (4%)	NA
Sex						
	Female	227 (66%)	248 (66%)	274 (63%)	749 (65%)	79 (50%)
	Male	115 (34%)	128 (34%)	159 (37%)	402 (35%)	78 (50%)

Race/Ethnicity					
White, Non-Hispanic	240 (70%)	226 (60%)	212 (49%)	678 (59%)	121 (78%)
Black, Non-Hispanic	30 (9%)	35 (9%)	43 (10%)	108 (9%)	20 (13%)
Asian, Non-Hispanic	37 (11%)	75 (20%)	121 (28%)	233 (20%)	1 (1%)
Mixed, Non-Hispanic	12 (3%)	13 (3%)	0 (0%)	25 (2%)	0 (0%)
Other, Non-Hispanic	4 (1%)	3 (1%)	20 (5%)	27 (2%)	0 (0%)
Hispanic	20 (6%)	24 (6%)	37 (9%)	81 (7%)	14 (9%)
Education					
Bachelors or higher	308 (90%)	325 (86%)	374 (86%)	1,007 (87%)	129 (82%)
Less than Bachelors	35 (10%)	51 (14%)	59 (14%)	145 (13%)	28 (18%)

Description: This shows the number of participants across rounds 2-4 of the Corporate Mix and the Nursing Home mix studies in five US companies and four nursing homes, November 2020 – December 2021.

Table 2. Comparison of reporting type (n (%) by study round and overall, November 2020 – December 2021 – Corporate and Nursing Home Mix.

	Underreported N(%)	**No difference in reporting** N(%)	**Overreported** N(%)
Corporate Mix			
Round 2 (N = 343)	0 (0%)	59 (17%)	284 (83%)
Round 3 (N = 376)	4 (1%)	140 (37%)	232 (62%)
Round 4 (N = 433)	8 (2%)	149 (34%)	276 (64%)
Overall (N = 1,152)	12 (1%)	348 (30%)	792 (69%)
Nursing Home Mix			
Overall (N=157)	14 (9%)	130 (83%)	13 (8%)

Description: This table compares the number of participants (and percentages) across Rounds 2-4 and overall for the Corporate Mix study and overall for the Nursing Home Mix study who fall under each category of reporting type.

Table 3. Comparison of reporting type (n (%)) by demographic characteristics and round, November 2020 – December 2021 – Corporate Mix study.

Corporate Mix Study	Round 2	Round 3	Round 4	Overall (N = 1152)

	Underreported (n (%))	No difference in contacts (n (%))	Overreported (n (%))	Underreported (n (%))	No difference in contacts (n (%))	Overreported (n (%))	Underreported (n (%))	No difference in contacts (n (%))	Overreported (n (%))	Underreported (n (%))	No difference in contacts (n (%))	Overreported (n (%))
Age group												
20-29	0 (0%)	14 (16%)	73 (84%)	1 (1%)	49 (41%)	70 (58%)	2 (1%)	47 (32%)	96 (66%)	3 (1%)	110 (31%)	239 (68%)
30-39	0 (0%)	23 (21%)	86 (79%)	0 (0%)	43 (41%)	61 (59%)	5 (4%)	43 (35%)	76 (61%)	5 (1%)	109 (32%)	223 (66%)
40-49	0 (0%)	15 (17%)	71 (83%)	1 (25%)	27 (19%)	52 (22%)	1 (13%)	37 (25%)	56 (20%)	2 (1%)	79 (30%)	179 (68%)
50+	0 (0%)	7 (11%)	54 (87%)	2 (3%)	21 (29%)	49 (68%)	0 (0%)	22 (31%)	48 (68%)	2 (1%)	50 (25%)	151 (74%)
Sex												
Female	0 (0%)	37 (16%)	190 (84%)	3 (1%)	88 (35%)	157 (63%)	4 (1%)	91 (33%)	179 (65%)	7 (1%)	216 (29%)	526 (70%)
Male	0 (0%)	21 (18%)	94 (82%)	1 (1%)	52 (41%)	75 (59%)	4 (50%)	58 (39%)	97 (35%)	5 (1%)	131 (33%)	266 (66%)
Race/Ethnicity												
White, Non-Hispanic	0 (0%)	36 (15%)	204 (85%)	3 (1%)	82 (36%)	141 (62%)	2 (1%)	67 (32%)	143 (67%)	5 (1%)	185 (27%)	488 (72%)
Black, Non-Hispanic	0 (0%)	3 (10%)	27 (90%)	1 (3%)	10 (29%)	24 (69%)	0 (0%)	18 (42%)	25 (58%)	1 (1%)	31 (29%)	76 (70%)
Asian, Non-Hispanic	0 (0%)	11 (30%)	26 (70%)	0 (0%)	32 (43%)	43 (57%)	5 (4%)	45 (37%)	71 (59%)	5 (2%)	88 (38%)	140 (60%)
Mixed, Non-Hispanic	0 (0%)	4 (33%)	8 (67%)	0 (0%)	4 (31%)	9 (69%)	0 (0%)	4 (31%)	9 (69%)	0 (0%)	12 (32%)	26 (68%)
Other, Non-Hispanic	0 (0%)	1 (25%)	3 (75%)	0 (0%)	1 (33%)	2 (67%)	0 (0%)	3 (43%)	4 (57%)	0 (0%)	5 (36%)	9 (64%)
Hispanic	0 (0%)	4 (20%)	16 (80%)	0 (0%)	11 (46%)	13 (54%)	1 (3%)	12 (32%)	24 (65%)	1 (1%)	27 (33%)	53 (65%)
Education Level												
Bachelors or higher	0 (0%)	54 (18%)	254 (82%)	3 (1%)	122 (38%)	200 (62%)	8 (2%)	133 (36%)	233 (62%)	11 (1%)	309 (31%)	687 (68%)
Less than Bachelors	0 (0%)	5 (14%)	30 (86%)	1 (2%)	18 (35%)	32 (63%)	0 (0%)	16 (27%)	43 (73%)	1 (1%)	39 (27%)	105 (72%)

Description: This table compares participant reporting type by each demographic factor (age, sex, etc.) and across rounds 2-4 and overall. The values in the overall columns reflect the row totals for each level of each demographic factor.

Table 4. Comparison of reporting type (n (%)) by demographic characteristics, December 2020 – June 2021 – Nursing Home Mix study.

Nursing Home Mix (N = 157)		**Underreported (n (%))**	**No difference in contacts (n (%))**	**Overreported (n (%))**
Age group				
	20-29	2 (40%)	3 (60%)	0 (0%)
	30-39	3 (5%)	55 (87%)	5 (8%)
	40-49	5 (6%)	70 (85%)	7 (9%)
	50+	4 (57%)	2 (29%)	1 (14%)
Sex				
	Female	10 (13%)	61 (77%)	8 (10%)

Male	4 (5%)	69 (88%)	5 (6%)
Race/Ethnicity			
White, Non-Hispanic	7 (6%)	104 (86%)	10 (8%)
Black, Non-Hispanic	4 (20%)	14 (70%)	2 (10%)
Asian, Non-Hispanic	0 (0%)	1 (100%)	0 (0%)
Hispanic	2 (14%)	11 (79%)	1 (7%)
NA	1 (100%)	0 (0%)	0 (0%)
Education Level			
Bachelors or higher	6 (5%)	113 (88%)	10 (8%)
Less than Bachelors	8 (29%)	17 (61%)	3 (11%)

Description: This table compares participant reporting type (n (%)) by each demographic factor (age, sex, etc).

Table 5. Effect measure estimates (ORs) and 95% CIs for each strata - Corporate Mix.

Demographic		Estimate (ORs)	95% CI	
Sex				
	Male	0.83	0.56	1.10
	Female	2.44	2.28	2.60
Age (years)				
	20-29	2.17	1.94	2.40
	30-39	0.94	0.61	1.27
	40-49	1.04	0.68	1.40
	50+	1.39	0.99	1.79
Race and Ethnicity				
	White, Non-Hispanic	2.64	2.47	2.81
	Asian, Non-Hispanic	0.60	0.28	0.92
	Black, Non-Hispanic	0.93	0.47	1.39
	Mixed, Non-Hispanic	0.82	0.10	1.54
	Other, Non-Hispanic	0.68	-0.45	1.81
	Hispanic	0.74	0.24	1.24
Education Level				
	Bachelors or higher	2.45	2.25	2.65
	Less than Bachelors	1.14	0.86	1.42

Description: This table compares the odds ratio estimates (and CIs) among different levels of each predictor.