An Analysis of Nonresponse Bias in the World Trade Center Health Registry

Joe Murphy[1], Robert Brackbill[2,3], James H. Sapp II[3], Lisa Thalji[1], and Paul Pulliam[1]
RTI International (trade name of Research Triangle Institute)[1]
New York City Department of Health and Mental Hygiene[2]
Agency for Toxic Substances and Disease Registry[3]

Keywords: non-response, bias, raking, World Trade Center

1. Introduction

The World Trade Center Health Registry is a database for tracking persons who were exposed to the WTC disaster on September 11, 2001 (9/11). The study is a joint effort of the New York City Department of Health and Mental Hygiene (NYCDOHMH) and the Agency for Toxic Substances and Disease Registry (ATSDR). Baseline Registry building and data collection activities were conducted by RTI International. The baseline enrollment phase was completed in November 2004 with 71,437 persons enrolling and completing a thirty minute interview over the telephone or in person. The WTC Health Registry is the largest exposure registry in the United States and members of the Registry will be followed for up to twenty years.

The purpose of the Registry is to evaluate potential short and long term physical and mental health effects of the exposure to the disaster. To enroll, potential registrants were asked to report demographic information; their location on 9/11; what they saw; their exposure to dust, smoke, and debris; the amount of time before returning to work or home; their physical and mental health before and after 9/11; and contact information to assist with future follow-up.

Exposed groups were broadly defined based on proximity to the WTC disaster and its aftermath. The Registry includes persons who were downtown (South of Chambers Street in Manhattan) on the morning of September 11, 2001 and who may have been present during the collapse of the two towers and the subsequent dust/debris cloud; rescue, recovery, and clean-up workers and volunteers who worked on the pile or its vicinity in the days and weeks following the disaster; residents who lived in the surrounding area around the WTC disaster site (South of Canal Street in Manhattan); and school children and staff in schools in downtown Manhattan (South of Canal Street).
The broadly defined exposure groups were separated into high priority and low priority exposure populations. High priority exposed persons are defined as those who had relatively high levels of exposure and a greater chance of being located; this group is referred to as Group 1. Group 2 includes persons who are have less acute exposures than those in Group 1, such as persons who were on the street south of Chambers on September 11, 2001 but not in one of the 35 damaged or destroyed buildings or 3 structures nearest to the WTC site. All workers and volunteers and students and school staff are in Group 1. Residents who lived South of Chambers Street (closer to the WTC site) are Group 1, while residents between Canal and Chambers are Group 2. People who were in any one of the damaged or destroyed buildings prior to or at the time of the attack were designated as Group 1 and other occupants or passersby south of Chambers Street on 9/11 are in Group 2. Figure 1 presents a map of the approximately one square-mile area for reference.

Figure 1. Map of Lower Manhattan

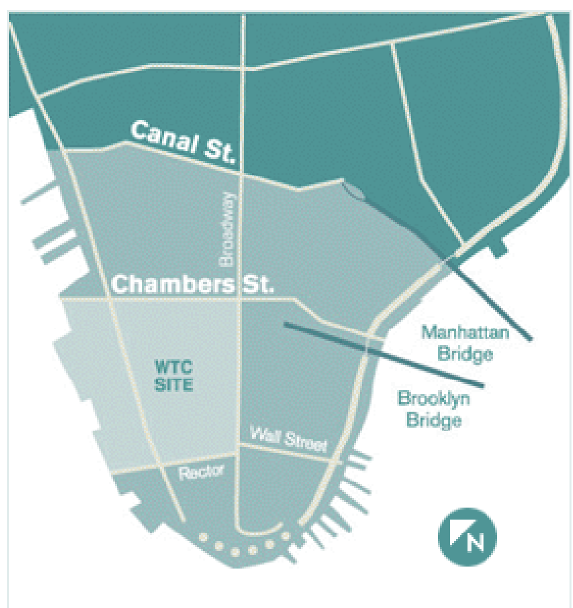
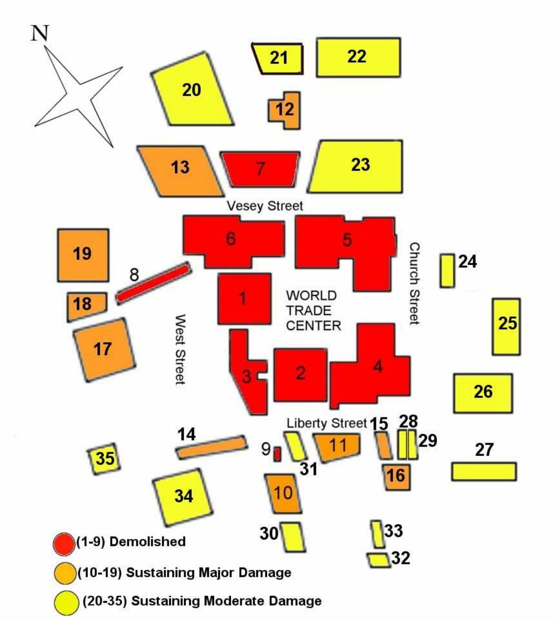


Figure 2 presents depicts the buildings sustaining moderate or major damage and those that were destroyed in and around the WTC site.

Figure 2. Damaged or Destroyed Buildings



The study was designed such that all exposed persons (estimated at more than 360,000) were eligible to register. This design was undertaken in lieu of a smaller sample survey because the Registry itself will serve as a sampling frame for future smaller studies.

Within Group 1, the sample was primarily composed of records obtained from over 200 list sources (resident databases, lists obtained from businesses in and around the WTC site, rescue/recovery organizations, etc.). Eligible Group 1 registrants could also self-identify through the study web site or toll free telephone number. Self-identification was the main enrollment method for Group 2.

Because of the study design, a degree of unequal representation of eligible persons was expected. Eligible persons had unequal likelihoods of being included on various lists, volunteering to complete the interview, and responding if contacted. The purpose of this paper is to analyze the degree to which health estimates from the Registry may be biased because of self-selection and nonresponse. The research questions we address are:

1) Is nonresponse bias sufficient to alter results differentially among sample types?

2) Can nonresponse adjustment weights be computed simply to address this and provide indication of the direction and degree of the bias by sample type?

Our focus is on illustratrating the value of our approach, the importance of its concepts to those who may use the Registry data, and its application to other similar studies.

When we discuss bias, we mean the differences between those who enrolled in the Registry and those who did not. These differences need to be considered in relation to exposure and health outcomes. Responders or those with the highest propensity to respond may be different from nonresponders in terms of exposure or outcome differences. For example, people who believe they experienced a dangerous exposure may over report health symptoms. This is important to consider since estimates produced from the Registry will come to represent the entire population at risk.

2. Methods

To analyze nonresponse, we focus on three types of data:

1) Process Measures. These are data available for both responders and nonresponders in the Registry database. They can be used to calibrate respondent data to match sample marginals. The process measures we analyze are whether the sample member was included on a list or self-identified; the number of calls required to finalize the case; the proportion of calls in which a human contact was made; and whether the sample member ever refused to be interviewed.

2) Population Measures. These are data available for individual responders and at the aggregate level for the entire true eligible population. The sample data can be adjusted to match the population totals on different dimensions. For workers and volunteers we have responder and population totals for the number from the New York Fire Department (FDNY), the New York Police Department (NYPD), the Department of Sanitation, and other organizations. For residents, we have responder and Census data on age, gender, race/ethnicity, and ZIP code. For students and school staff, we have responder, National Center for Education Statistics, and Bureau of Day Care data on public school, private school, and preschool/daycare enrollment. For building occupants we have counts of whether responders and members of the true eligible population were in either of the two WTC towers at the time of impact.

3) Outcomes. These data are only available for the responders, but we use the auxiliary process and population measures to adjust the outcomes to fit the population at risk. In this paper, we look at whether responders reported a new or worse cough since 9/11; new or worse breathing problems since 9/11; and new or worse depression since 9/11.

To assess whether nonresponse bias is present and the degree to which it may affect the outcome estimates, we employed a technique called raking ratio estimation (Kalton, 1983). This method adjusts data so their marginal totals match specified control totals on a specified set of variables (Battaglia, et al., 2004). This is definitely not the only approach that could be taken, but was chosen for its ease in implementation and interpretation. We conducted the raking in two stages, producing two sets of weights that when applied to the unadjusted outcome measures produced estimates adjusted to match the characteristics of the population. The first stage adjusted the data for the responders to match the marginal control totals for the entire sample. The second stage adjusted the sample totals to match the marginal control totals for the entire true eligible population. For a discussion of other appropriate adjustment methods see, for example, Creel, 2005.

To assure meaningful adjustments, we included variables correlated with nonresponse, controlling for other factors (Farooque, et al., 1999). For example, we ran logistic regression models for the first adjustment stage predicting response by sample members based on process measures. A separate model was run for each sample type and group (Group 1 workers and volunteers; Group 1 residents; Group 2 residents; Group 1 students and school staff; Group 1 building occupants; and Group 2 building occupants and passersby). This model takes the form:

$$\ln[(p/1-p)] = a + bS + bC + bH + bR$$

where:

ln=the natural logarithm, $\log_{exp}$ (exp=2.71828…)
p=probability of response
a=intercept
b=slope coefficient
S=self-identified (Yes, No)
C=number of calls (0-1, 2-6, 7-29, 30+)
H=percent of calls with human contact (<50, >=50)
R=ever refused (Yes, No)

In this model, all predictors turned out to be significantly correlated with response at $p<.001$, controlling for the other predictors. Self-identification had the strongest effect, with an average odds ratio (OR) of 16.7, suggesting that those who self-identified were much more likely to respond than those who were included on an obtained list. The number of calls made to a respondent was negatively correlated with response, meaning that the more calls it took to finalize a case, the less likely it was that a response would be obtained (average OR=0.62). Obtaining a high percentage of human contact among calls made to a case was positively correlated with response, meaning the greater the percentage of contact, the greater the likelihood of obtaining a response (average OR=2.9). The act of ever refusing to respond to the interview was negatively correlated with response (average OR=0.26).

Because all predictors in the model were significantly correlated with response, we included them in the first stage of adjustment. To complete the raking procedure, we used the IHB macro for SAS software developed by Izrael, et al. (2004). This macro allows the programmer to input the control totals, point SAS to the sample data set and run the program which filters through multiple iterations of the raking procedure to output weight values for every observation in the set.

After the first stage of the raking procedure, the process was repeated using the population measures listed above for all sample groups and types except Group 2 building occupants and passersby, for whom no population control totals were available.

3. Results

We completed the raking procedures and applied the first and second stage weights to the outcome measures listed in the Methods section. In general, the unadjusted estimates appear slightly inflated compared to the adjusted estimates. This suggests that those who completed the Registry interview were more likely to report having new or worse symptoms or conditions than those who did not since 9/11. One caveat is that we must assume that the relationship between the demographic or control variables and the outcome variables is constant between the responders, nonresponders, and entire true eligible population. Without a definitive external data source, however, this assumption cannot be validated.

We present the effect of the adjustments on our three outcome measures by sample type and group graphically below. We chose not to include the specific values of the data points because that is not the focus of this paper. Values of exact estimates will be reserved for a future study publication where they will be discussed fully. The focus here is to simply illustrate the value of our approach, the importance of its concepts to those who may use the Registry data, and its application to other similar studies.

Figure 3 presents the legend for Figures 4-6. The bars in blue show the unadjusted outcome estimates; red bars show the adjusted estimate after raking to the sample totals; the white bars show the final adjusted estimates after raking to the population totals.

Figure 3. Legend for Figures 4-6

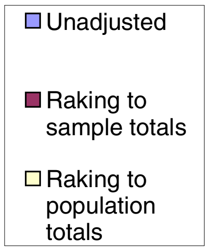
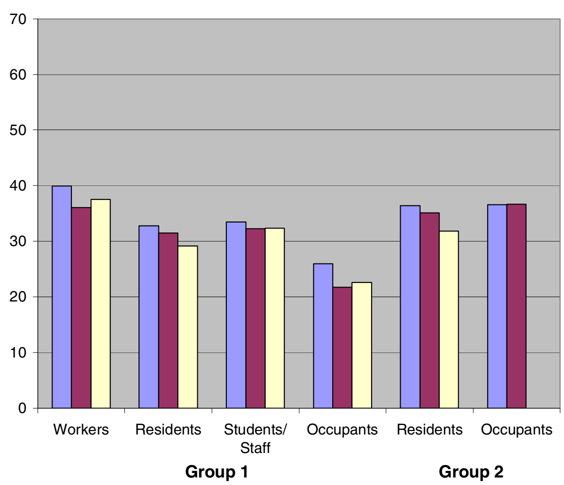


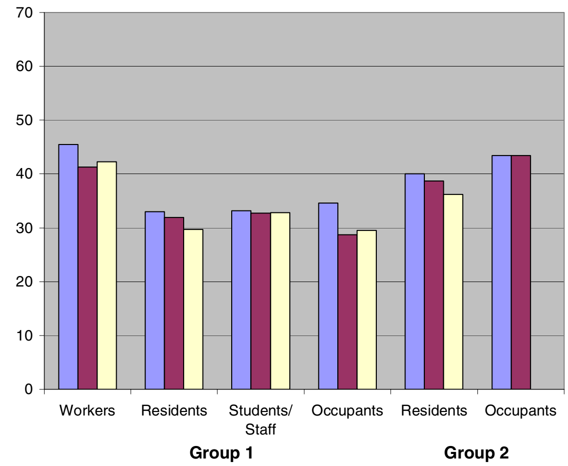
Figure 4 shows the adjusted percentage estimates for reporting a new or worse cough since 9/11. The figure shows that adjusting for nonrespone generally has a negative effect on the estimates meaning that we may have obtained responses disproportionately from those who were more likely to report a new or worse cough. The difference between unadjusted and adjusted measures was greatest for Group 2 residents, a group primarily composed of self-identifiers. Self-selection bias may have made the unadjusted estimate appear inflated.

Figure 4. Percent Reporting New/Worse Cough Since 9/11



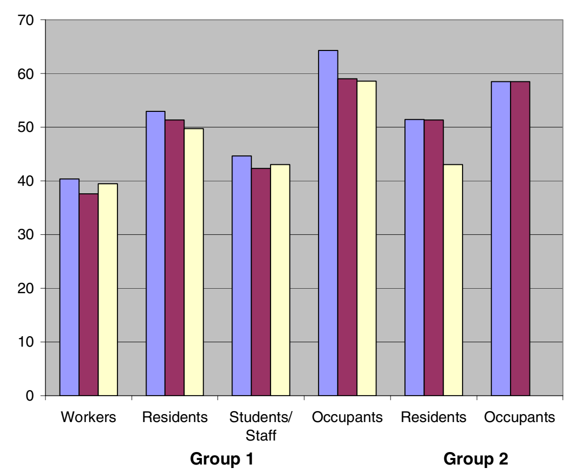
The estimates for new/worse breathing problems in Figure 5 show a similar pattern. Adjusted estimates are generally lower than the unadjusted estimates, suggesting that responders were more likely to have or report having new or worse breathing problems than nonresponders.

Figure 5. Percent Reporting New/Worse Breathing Problems Since 9/11



Finally, we analyzed the effect of nonresponse bias on new/worse depression since 9/11. As with the other measures, the adjustments deflated the unadjusted estimates, especially for the Group 2 residents.

Figure 6. Percent Reporting New/Worse Depression Since 9/11



4. Discussion

This paper aimed to address two research questions, the first being “Is nonresponse bias sufficient to alter results differentially among sample types?” Comparisons of the unadjusted and adjusted estimates show that the unadjusted estimates appear inflated, in general and that nonresponse bias may be an important factor to account for when analyzing these data. There appears to be more bias for sample types with high rates of self-selection (e.g. Group 2 residents) so nonresponse bias may be more of a problem for some sample groups compared to others.

Our second research question asked “Can nonresponse adjustment weights be computed simply to address this and provide indication of the direction and degree of the bias by sample type?” We believe we have demonstrated that these weights can be computed simply to provide a basic indication of whether nonresponse bias may be a problem and how it may be a problem. More time and resources devoted to the issue could find most ideal method for addressing nonresponse and coverage issues in this and other registries. Also, more direct measures from nonresponders could be extremely informative and provide a more accurate picture of the direction and degree of bias.

Analyses of WTC Health Registry should acknowledge that nonresponse bias may be present and generally inflates health outcomes estimates to a modest degree. The effect is not constant across sample types. Adjustment weights for nonresponse can be computed and may be important to analysis. This paper provides an example of a cursory analysis that suggests a more detailed investigation may be warranted. Information on the degree and direction of nonresponse bias can be obtained using methods like the ones used in this paper. They need not be extraordinarily complex at first, and are worth the effort, especially for surveys and registries allowing for self-identification.

References

Battaglia, M. et al. (2004). “Tips and Tricks for Raking Survey Data (a.k.a. Sample Balancing).” Abt Associates.

Creel, D. and M. Fahimi (2005). “Multidimensional Control Totals for Poststratified Weights.” Presented at the Joint Statistical Meetings, Minneapolis, MN.

Farooque, G. et al. (1999) “Selecting Variables for Poststratification and Raking.” Proceedings of the annual meeting of the American Statistical Association.

Izrael, D. et al. (2000). “A SAS Macro for Balancing a Weighted Sample.” Proceedings of the 25th Annual SAS Users Group International Conference.

Kalton, G. (1983). Compensating for Missing Survey Research Data. Research Report Series. Ann Arbor, MI: Institute for Social Research, University of Michigan.