

January 2006

World Trade Center Health Registry Explanation and Calculation of Outcome Rates¹

Prepared for:

Robert Brackbill, Ph.D., M.P.H.

Mark Farfel, Sc.D.

Deborah Walker, Ph.D.

New York City Department of Health and Mental Hygiene

Sharon Campolucci

Jay Sapp, M.S.

Agency for Toxic Substances and Disease Registry

Prepared by:

Joe Murphy, M.A.

RTI International²

RTI Project Number 0208692

Contract No. 200-2002-M-00836



¹ The findings in this report have undergone external peer review as required by ATSDR policy and the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA).

² RTI International is a trade name of Research Triangle Institute.

**World Trade Center (WTC) Health Registry
Explanation and Calculation of Outcome Rates**

Table of Contents

Executive Summary	2
I. Overview of the WTC Health Registry Eligible Population and Sample Frame	3
II. Outcome Rates Definitions.....	7
A. Contact Rate.....	7
B. Cooperation Rate	11
C. Response Rate	12
D. Eligibility Rate	13
E. Coverage Rate	13
F. Enrollment Rate	16
III. Final Outcome Rates.....	17
IV. Discussion	19
V. References.....	25
Appendix A	26
Appendix B	27
Appendix C.....	28

World Trade Center (WTC) Health Registry

Explanation and Calculation of Outcome Rates

Executive Summary

The purposes of this report are to document and explain the calculation of the final outcome rates for the WTC Health Registry. The outcome rates indicate coverage of the study population on several different dimensions. Where appropriate, standard formulas endorsed by the American Association for Public Opinion Research (AAPOR) are employed. The measures detailed in this report include the contact, cooperation, response, eligibility, coverage, and enrollment rates. The definition, formula, and interpretation, and final values for each rate by sample group and type is presented. A summary of the final outcome rates and important points from the report follows:

- The enrollment rate for the WTC Health Registry was 17.38%. That is, out of the 526,269 total person-type combinations³ in the true eligible population, the Registry obtained completed interview data for 91,463, or 17.38 %.
- The overall coverage rate was 32.93%. This means that the sample frame covered 32.93% of the true eligible population.
- The overall eligibility rate was 76.94%, meaning that 76.94% of preregistrants for whom eligibility was definitively determined were eligible.
- The overall contact rate as shown in Table 2 was 73.76%. This means that 73.76% of the estimated eligibles in the sample frame were contacted.
- The overall cooperation rate was 80.52%. This means that of the estimated eligible who were contacted, an interview was completed with 80.52%.

³ The total overall true eligible population is estimated at 410,761 persons. The total of 526,269 represents person-types for enrollment calculation. See Section E of this report for a comparison of these figures.

- The overall response rate was 59.40%. This means that interviews were completed with 59.40% of the estimated eligible sample. This surpassed the prespecified goal of 49%, which was based on contact and cooperation rate targets of 70%.
- The process of computing these rates made it evident that careful consideration of sample-specific attributes is very important. In a study where one person can represent up to four completed “interviews,” (i.e. a person could be included based on more than one eligibility requirement), care should be taken to assure that the correct base, or denominator, is used for measuring enrollment, coverage, or response.
- Although not investigated in this report, related research suggests those who were registered generally represent the true eligible population, as was intended.

This report received ATSDR peer review in accordance with the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA) section 104(i)(3), as amended by the Superfund Amendments and Reauthorization Act of 1986. All data are based on the November 2005 WTCHR database.

I. Overview of the WTC Health Registry Eligible Population and Sample Frame

The World Trade Center Health Registry is a database for tracking persons who were exposed to the WTC disaster on September 11, 2001. 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.

Eligibility for the WTC Health Registry was defined according to specific criteria. Preregistrants (entries in the sample database) were assigned to one of four sample types, based on the list source or information captured through the self-identification process:

- **Workers and volunteers** involved in rescue, recovery, or clean-up at the WTC site anytime from September 11, 2001, through June 30, 2002 and/or in debris handling at the Staten Island Landfill anytime from September 12, 2001, through June 30, 2002;
- **Residents** as of September 11, 2001 at addresses located south of Canal Street;
- **Students and school staff** enrolled/employed in schools or daycares south of Canal Street on September 11, 2001;
- **Building occupants, people in transit, and pedestrians**, employees, visitors, and passers-by present south of Chambers Street on September 11, 2001.

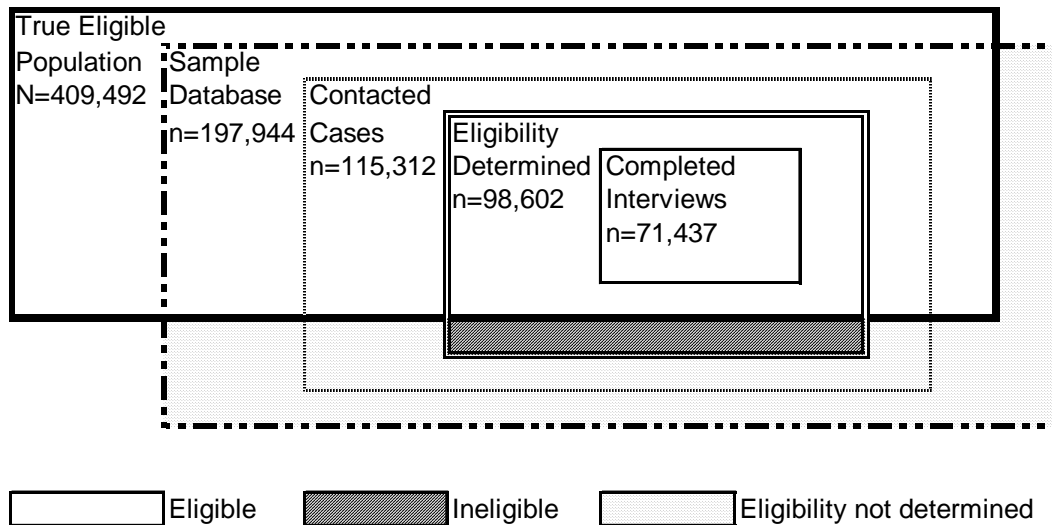
For two sample types, preregistrants were further stratified into higher and lower priority groups (Group 1 and Group 2, respectively). For occupants, Group 1 included employees of businesses in 38 damaged or destroyed buildings in and around the WTC site present on the morning of September 11, 2001. Group 2 included occupants, employees, visitors, and passers-by who were present south of Chambers Street on September 11, 2001, other than employees in the 38 damaged or destroyed buildings in and around the WTC site. For residents, Group 1 included residents as of September 11, 2001 at addresses located south of Chambers Street; Group 2 included residents located on or north of Chambers but south of Canal Street. A third resident group, Group 0, was defined upon recognition of the inclusion of respondents living on or North of Canal in ZIP codes overlapping the Canal Street boundary.

During sample frame construction, coverage of the true eligible population was maximized through list building activities and self-identification (Dolan et al, 2005). Interviews were completed with sample members in order to evaluate the extent of initial acute respiratory, psychological, and other injury across different populations affected.

Figure 1 shows the relationship between the true eligible population, sample database, contacted cases, and completed interviews. This diagram is included to aid the understanding of

the outcomes rate components, calculations, and interpretations. The diagram shows that the sample database of preregistrants covers some, but not all of the true eligible population of workers and volunteers, residents, students and school staff, and occupants described above. Because eligibility was not determined until the beginning of the interview, some cases in the sample database were found to be outside the true eligible population. For some cases, eligibility was not determined.

Figure 1. Components of the True Eligible Population and Sample Database

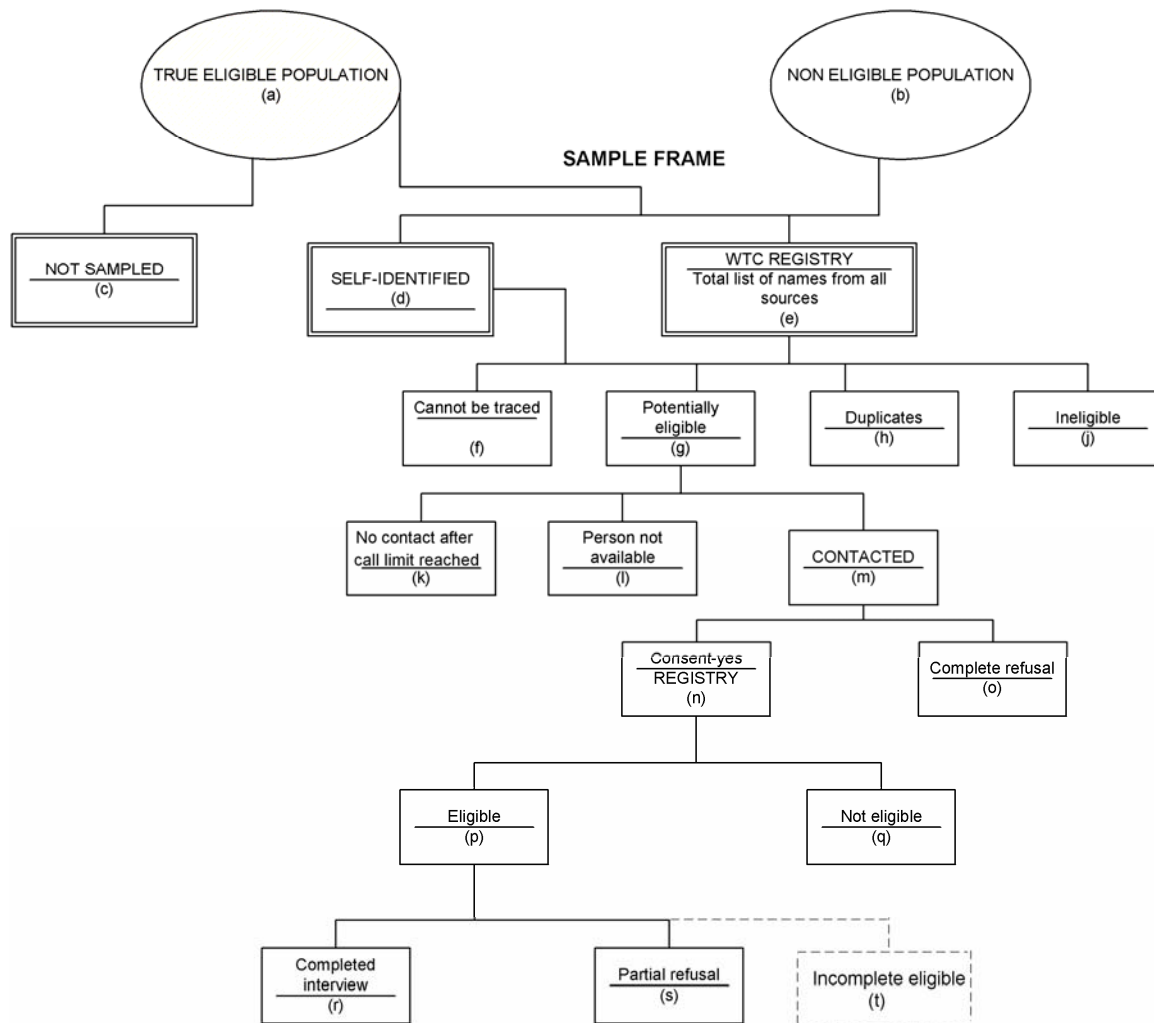


The degree to which eligible members of the sample database cover the true eligible population is called the *coverage rate*. Some, but not all, of the preregistrants were contacted by the study, meaning a human contact was made with the sample member. The degree to which eligible sample members were contacted by the study is called the *contact rate*. Of those contacted, some were determined to be eligible, some ineligible and others were indeterminate. The degree to which those for whom eligibility was determined were eligible is called the *eligibility rate*.

A completed interview was obtained from some, but not all, preregistrants. The degree to which those contacted were interviewed is called the *completion rate*. The degree to which eligible members of the sample database were interviewed is called the *response rate*; this is mathematically equivalent to the product of the contact and cooperation rates. Finally, the degree to which the completed interviewers cover the true eligible population is called the *enrollment rate*.

Figure 2 from the World Trade Center Registry Protocol (Brackbill and Thomas, 2003) also summarizes the process by which elements of the sample frame related to the true eligible population. One addition was made to this figure to help with the conceptualization of all rates in this report. Cases of known eligibility for which an interview or partial interview was *not* completed are referred to as (t). Box (t) comprises a third subset of the eligible population (p) and is located next to box (s).

Figure 2. Summary of Enrollment and Data Collection Tracking from the World Trade Center Registry Protocol



In order to help understand the potential for non-response and coverage error present in the Registry, contact, cooperation, response, eligibility, coverage, and enrollment rates were calculated and reported for each sample type and group and overall. This report defines these

outcome rates and the methods and rationale for calculating them, and provides a brief interpretation of the final outcome rates.

II. Outcome Rates Definitions

For the calculation of outcome rates for surveys, RTI's standard is the American Association for Public Opinion Research's (AAPOR) *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys* (AAPOR, 2000). This document provides comprehensive methods for calculating outcome rates for surveys conducted by random-digit dialing (RDD) telephone, for personal interviews in a sample of households, and for mail surveys of specifically named persons. While the WTC Health Registry does not neatly fit into one of these three categories, it can be described primarily as a telephone survey of specifically named persons (a combination of the first and third types listed above). As is such, the AAPOR standards serve as the correct guidelines for the calculation of contact, cooperation, and response rates for the Registry and as the base information source for calculation of the eligibility, coverage, and enrollment rates.

A. Contact Rate

AAPOR definition. The *contact rate* measures the proportion of all cases in which an eligible member of the sample was reached by the survey. The components of the contact rate are:

I = Complete interview

P = Partial interview

R = Refusal and break-off

O = Other

NC = Non-contact

UH = Unknown if household/occupied household

UO = Unknown, other

E = estimated proportion of cases of unknown eligibility that are eligible

AAPOR defines these components for telephone surveys in the following way: Interviews are divided into complete (I) and partial (P) based on pre-determined completion

criteria. Refusals and break-offs (R) are eligible cases for which no interview is obtained. Other (O) cases represent instances in which there is an eligible respondent who did not refuse the interview, but no interview is obtainable. This includes situations involving death of an eligible respondent, physical and/or mental inability, and language problems. Non-contacts (NC) include cases determined to be eligible for participation, but the respondent is never available. Cases where it is unknown if a household was reached or if the household is occupied (UH) are of unknown eligibility. Unknown, other (UO) cases include situations where it is unknown if an eligible respondent resides in a household.

For the Registry, eligibility was not determined until the interview was conducted. Therefore, non-contacts (NC) for which eligibility was not determined were better defined as being of unknown eligibility (UO). Cases where the household status (whether it is indeed a household and not a business, etc.) was not determined were a subset of non-contacts and could also be classified as unknown, other (UO). The estimated proportion of cases of unknown eligibility that are eligible (E) is often computed by dividing the number of known eligibles (in this case, I and P) by the known eligibles plus known ineligibles. For the Registry, this proportion was based on cases for which eligibility was definitively determined. Estimating eligibility for non-contacts based on the rate for those of known eligibility can bias the contact rate if the two groups are systematically different in terms of eligibility. For example, if non-contacts were in fact eligible at a lower rate, using the contact rate for those where eligibility was determined would inflate overall eligibility estimates. Regardless, the most reasonable assumption for the Registry is that the eligibility rate for UO cases was similar to that of cases where eligibility had been determined.

Rate calculation. Table 1 presents the correspondence of the AAPOR components to those detailed in Figure 2 and the WTC Health Registry final disposition codes: Some components of the different nomenclatures appear similar but actually stand for different things. For example, the AAPOR component P represents partial interviews, while p in the WTC Health Registry Enrollment Tracking diagram stands for eligible cases. Readers should use caution when reviewing the components and refer to Table 1 and Figure 2 for appropriate definitions.

Table 1. AAPOR Outcome Rate Components by Registry Dispositions

AAPOR Component	Enrollment Tracking Equivalent	WTC Health Registry Final Disposition*
I (completed interview)	r – completed interview	657 (Interview Completed)
P (partial interview)	s – partial interview	656 (Partial Interview Completed)
R (refusals and break-offs)	t - eligibility determined, but interview not completed	839 (Final Eligible Breakoffs)
NC (eligible non-contacts)	<i>not applicable – eligibility is determined after contact is made</i>	<i>not applicable – eligibility is determined after contact is made</i>
O (eligible other non-interviews)	<i>not applicable – eligibility is determined after interview begins</i>	<i>not applicable – eligibility is determined after interview begins</i>
UH (unknown if household is occupied)	<i>not applicable – sample is composed of individuals, not households</i>	<i>not applicable – sample is composed of individuals, not households</i>
UO (eligibility unknown, other)	f (cannot be traced) + k (never contacted) + i (person not available) + o (complete refusal)	<i>f: can not be traced</i> 662 (Final Unlocatable) 740 (Insufficient info to trace or contact) + <i>k: never contacted</i> 822 – (Finalized Pending Group 2 Residents - No Human Contact) 824 – (Finalized Pending TIO Cases - No Human Contact) 828 – (Finalized Port Authority Phone and Address - Did not pursue, resource constraints) 829 – (Finalized Port Authority Address Only, Did not Pursue, resource constraints) 835 – (Finalized Pending Field Cases - No Human Contact) 836 (Finalized - Pending Field Cases, Out of NYC Metro, did not pursue, resource constraints) 663 (Final Noninterview – Other) 810 – (9/11 casualty, possible duplicate) + <i>l: person not available</i> 840 (Person not available) + <i>o: refusal</i> 658 (Final Refusal by R) 842 (Final Refusal by Other)
E (estimated proportion of unknown eligibility that are eligible)	p (eligible) / n (consent to participate)	<i>p: eligible [657 + 656 + 839] / n: consent to participate [657 + 656 + 839 + 659 (Final ineligible – screened) + 808 (Final ineligible – postcard) + 841 (Final unable to enroll in Registry)]</i>

* The full list of Registry final disposition codes is included in Appendix B.

When using the estimated proportion of cases of unknown eligibility that are eligible (E) component, the appropriate AAPOR contact rate is Contact Rate 2 (CON2), which is defined as follows:

$$\text{CON2} = \frac{(I + P) + R + O}{(I + P) + R + O + \text{NC} + E(\text{UH} + \text{UO})}$$

Using the nomenclature from Appendix A and accounting for the redundancy of NC and UH with UO, the AAPOR contact rate is defined as:

$$\text{Contact Rate} = \frac{(r + s) + t}{(r + s) + t + (p / n)(f + k + l + o)}$$

Because they were not eligible for inclusion in the Registry, duplicates (h), ineligible (j), and consenting ineligible (q) are not included in the contact rate denominator.

To account for the fact that a portion of the contacted sample members were of unknown eligibility, the estimated proportion of eligibility must be applied to the numerator of the contact rate calculation in addition to the denominator. Using the Registry terminology the contact rate is defined as:

$$\text{Contact Rate} = \frac{(\text{Eligibility Rate})\text{Total Contacted}}{\text{Known Eligibles} + (\text{Eligibility Rate})\text{Indeterminate Eligibility}}$$

The full definition by Registry disposition codes is too unwieldy to include in the body of this report but is included with all other component and rate definitions in Appendix B.

B. Cooperation Rate

The *cooperation rate* is the proportion of all cases interviewed of all eligible units ever contacted. For list-based or random digit dial studies as opposed those where respondents self-select into the sample, the cooperation rate is an important indicator of success in convincing respondents to participate. Because the Registry is only partially list-based, the cooperation rate can appear positively skewed because a significant portion of the sample volunteered, meaning they were very likely to cooperate. For the Registry, the response rate (described in Section II.C. of this report) is the best measure of overall success in completing interviews with sample members because it includes all sample members in the denominator, not just those who were contacted. The denominator for the cooperation rate is equal to the numerator of the contact rate – the number of eligible contacted cases. Using the AAPOR components defined in the previous section, the AAPOR Cooperation Rate 2 (COOP2) formula is as follows:

$$\text{COOP2} = \frac{(I + P)}{(I + P) + R + O}$$

Using the nomenclature from Figure 2, the WTC Registry cooperation rate is defined as:

$$\text{Cooperation Rate} = \frac{(r + s)}{(r + s) + t}$$

To account for the fact that a portion of the contacted sample members were of unknown eligibility, the estimated proportion of eligibility must be applied to the denominator of the cooperation rate. Using the Registry terminology the contact rate is defined as:

$$\text{Cooperation Rate} = \frac{\text{Completed Interviews}}{(\text{Eligibility Rate})\text{Total Contacted}}$$

The full definition by Registry disposition codes is included with all other component and rate definitions in Appendix B.

C. Response Rate

The *response rate* is the number of completed interviews divided by the number of interviews plus the number of non-interviews plus cases of unknown eligibility. Its value is equal to the product of the contact and cooperation rates, which for the Registry is equivalent to AAPOR Response Rate 4 (RR4):

$$RR4 = \frac{(I + P)}{(I + P) + R + O + NC + E(UH+UO)}$$

Using the nomenclature from Figure 2, the response rate is defined as:

$$\text{Response Rate} = \frac{(r + s)}{(r + s) + t + (p / n)(f + k + l + o)}$$

Using the Registry terminology the contact rate is defined as:

$$\text{Response Rate} = \frac{\text{Completed Interviews}}{\text{Known Eligibles} + (\text{Eligibility Rate})\text{Indeterminate Eligibility}}$$

The full definition by Registry disposition codes is included with all other component and rate definitions in Appendix B.

D. Eligibility Rate

The primary goal of the list building task for the Registry was to create a sample frame adequately covering the population at risk. It was also desired that a very high proportion of those identified for inclusion would actually be eligible. The *eligibility rate* serves as an indicator of list providers' and self-identifiers' accuracy in identifying eligible participants. It is equivalent to the number of eligibles (p) among those who were successfully contacted and agreed to participate (n) and is defined as follows:

$$\text{Eligibility Rate} = \frac{p}{n}$$

Using the Registry terminology the eligibility rate is defined as:

$$\text{Eligibility Rate} = \frac{\text{Known Eligibles}}{\text{Known Eligibles} + \text{Known Ineligibles}}$$

The full definition by Registry disposition codes is included with all other component and rate definitions in Appendix B.

E. Coverage Rate

The extent to which the sample frame covers the true eligible population is referred to as the *coverage rate*. For the Registry, this is equivalent to the total number of eligible registrants in the sample frame compared to the true eligible population. Using the nomenclature from Figure 2, the coverage rate can be calculated by dividing the number of known eligible sample members (p) plus a proportion of sample members of unknown eligibility (f, k, l, and o) divided by the true eligible population (a). As with the contact rate, the eligibility rate is applied to the numerator:

$$\text{Coverage rate} = \frac{p + (p/n)(f + k + l + o)}{a}$$

Using the Registry terminology, the coverage rate is defined as:

$$\text{Coverage Rate} = \frac{\text{Completed Interviews} + (\text{Eligibility Rate})\text{Indeterminate Eligibility}}{\text{True Eligible Population w/ Hierarchy}}$$

The denominator in this equation accounts for sample type and group overlap according to the Registry hierarchy. Basically, the hierarchy defines a case according to the highest priority sample group and type definition that fits the case. The highest priority group and type combination under the hierarchy is Group 1 Workers. Cases that are exclusively Group 0 Residents are excluded because they are not considered part of the original true eligible population. The full hierarchy, from top to bottom, is as follows:

1. Group 1 Workers
2. Group 1 Residents
3. Group 1 Students/School Staff
4. Group 1 Occupants
5. Group 2 Residents
6. Group 2 Occupants

To account for the fact that the coverage rate numerator defines cases according to the hierarchy, the denominator is adjusted at the sample type and group level to fit the hierarchy. Sample group and type completes with the hierarchy are divided by completes without the hierarchy and that number is then multiplied by the sample group and type denominators. The sum of the group and type specific denominators with the hierarchy equals the overall true eligible population with hierarchy. The formula for this calculation is:

$$\text{True Eligible Population w/ Hierarchy} = \sum(r'_x/r_x * a_x)$$

Where:

r'_x = completed interviews of type x with the WTCHR hierarchy

r_x = completed interviews of type x without the WTCHR hierarchy

a_x = true eligible population estimate for type x without the WTCHR hierarchy

The formula can be used to compute coverage rate denominators for specific groups and types as well. As an example, the solution for Group 1 Occupants is as follows:

$$\begin{aligned} & \text{True Eligible Population for Group 1 Occupants w/ Hierarchy} \\ & = 7,905 \text{ Completes w/ Hierarchy} / 10,393 \text{ Completes w/o Hierarchy} \\ & \quad * 62,092 \text{ True Eligibles w/o Hierarchy} \\ & = 47,228 \end{aligned}$$

Summing over all sample groups and types, an overall coverage rate denominator of 365,269 is obtained. This number is used as the overall denominator for calculating the coverage rate because the numerator also follows the WTCHR hierarchy. It could also be used as an overall estimate of the true eligible population accounting for overlap of sample types, but different figures could be obtained sorting the hierarchy in a different order.

It is not ideal to base this important estimate on the hierarchy imposed by the study, so a different method is used to obtain the overall estimate of the true eligible population, accounting for overlap of sample types. Completed interviews of each group and type were classified by the total number of types under which they were eligible. This distribution was then applied to the marginal counts of true eligibles for each sample group and type to predict the number that would be classified under multiple sample types in the population. Because multiple types represent duplication in the count of unique persons, counts of cases identified under more than one sample type were divided by the number of types under which one would be classified, thereby converting person-type counts into simply person counts. The formula for this calculation is.

$$\text{Overall True Eligible Population} = \sum \sum [(r_{xy}/r_x * a_x)/y]$$

Where:

r_{xy} = completed interviews of type x eligible under y (1-4) number of types

r_x = the marginal number of interviews of type x

a_x = the marginal number of true eligibles of type x

As an example, the solution for Group 1 Occupant component of this formula follows:

$$\begin{aligned} & \text{True Eligible Population Component for Group 1 Occupants} \\ & = [(7,838 \text{ G1 Occupant completes of one sample type} / 10,393 \text{ G1 Occupant completes} \\ & \quad * 62,092 \text{ G1 Occupant True Eligibles}) / 1 \text{ sample type}] \\ & + [(2,432 \text{ G1 Occupant completes of two sample types} / 10,393 \text{ G1 Occupant completes} \\ & \quad * 62,092 \text{ G1 Occupant True Eligibles}) / 2 \text{ sample types}] \\ & + [(122 \text{ G1 Occupant completes of one sample type} / 10,393 \text{ G1 Occupant completes} \\ & \quad * 62,092 \text{ G1 Occupant True Eligibles}) / 3 \text{ sample types}] \\ & + [(1 \text{ G1 Occupant completes of one sample type} / 10,393 \text{ G1 Occupant completes} \\ & \quad * 62,092 \text{ G1 Occupant True Eligibles}) / 4 \text{ sample types}] \\ & = 54,337 \end{aligned}$$

Summing over all sample types and groups, an estimated overall denominator of 410,761 persons is obtained. While this number does not figure into any of the final WTC Health Registry outcome rates, it is important as a stand-alone figure as it represents the best estimate the project can provide of the true number of persons exposed to the events of September 11, 2001, as exposure is defined by the WTC Health Registry.

F. Enrollment Rate

To specifically evaluate the recruitment of eligible registrants from the true eligible population, an enrollment rate was calculated. The enrollment rate measures the proportion of the true eligible population for which an interview was completed. This rate accounts for sample type overlap in the completed interviews by using person-type combinations as the unit of analysis. This is appropriate since the denominator is also calculated in terms of person-type combinations. For instance, because of overlap, 50 completed interviews may actually represent 75 person-type combinations if half of the completed interviews were done with persons eligible under one sample type and half were done with persons eligible under two sample types. In terms of the nomenclature from Figure 2, the enrollment rate is defined as:

$$\text{Enrollment rate} = \frac{(r + s)}{a}$$

Using the Registry terminology the Enrollment rate is defined as:

$$\text{Enrollment rate} = \frac{\text{Completed Interviews without Hierarchy}^4}{\text{True Eligible Population without Hierarchy}}$$

The full definition by Registry disposition codes is included with all other component and rate definitions in Appendix B.

III. Final Outcome Rates

Table 2 presents the final WTC Health Registry outcome rates. Table 3 presents the final outcome rates excluding cases added to the sample frame from the Port Authority WTC tower occupants list. Additional tables of this type are also included. Table 4 presents the final outcome rates for list cases and Table 5 for self-identified cases. A copy of the SAS program used to produce the inputs for these tables is included in Appendix C. This program can easily be modified to include or exclude additional sample sources by inserting the appropriate “listsourcoid” filter.

Table 2 shows that enrollment for the WTC Health Registry was 17.38%. That is, out of the 526,269 total person-type combinations in the true eligible population, the Registry obtained completed interview data for 91,463, or 17.38 %. By sample type and group, enrollment was highest for Group 1 residents (37.26%) and lowest for Group 2 occupants (11.03%). At first glance, an enrollment rate of under 20% may appear to be too low from which to draw meaningful conclusions about the true eligible population. However, given the wide variation of

⁴ Each unit represents a person-type combination. So, for instance, one person could be included twice if he or she was determined eligible in the interview as a resident and student.

the types of people included in this population and the fact that many had moved from the area since September 11, 2001, completion of a 30-minute interview with almost 1 in 5 of those exposed can be seen as a major accomplishment. Whether these 91,463 person-type combinations adequately represent the true eligible population is a question beyond the scope of this report, but a preliminary analysis suggests that while unadjusted adverse health outcomes may be slightly inflated, the degree of bias is not statistically significant (Murphy, 2005).

The overall coverage rate was 32.93%. This means that the sample frame covered 32.93% of the true eligible population. Coverage was highest for Group 1 residents (66.16%) and lowest for Group 2 occupants (12.31%). Because Group 2 occupants were self-identifiers, it was expected that coverage may be relatively low for this group and type. Including nearly 1 in 3 eligible persons in the preregistrant database presents questions regarding the degree to which the sample represents the true eligible population, but again, given the challenge presented by the nature of this population, this level of coverage represents a successful effort including exposed individuals in the sample.

The overall eligibility rate was 76.94%, meaning that 76.94% of preregistrants for whom eligibility was definitively determined were eligible. This was highest for Group 2 occupants (98.51%) and lowest for Group 1 occupants (40.81%). It was expected that eligibility would be highest for groups and types consisting of self-identifiers like the Group 2 occupants. The Group 1 occupants include many cases originating from the Port Authority list. The eligibility rate for this list was exceptionally low and this contributed to the low eligibility rate for Group 1 occupants. Table 3 shows that when Port Authority cases are excluded, the eligibility rate for Group 1 occupants rises to 85.00% and the overall eligibility rate rises to 85.26%. When examining only list cases as shown in Table 4, the overall eligibility rate is only 57.34%. Table 5 shows that the overall eligibility rate for self-identified cases is much higher at 91.59%. The fact that eligibility among self-identifiers would be much higher than among those included on lists is not surprising. In order to self-identify, a preregistrant would likely have the information necessary to determine whether he/she would be eligible for the study.

The overall contact rate as shown in Table 2 was 73.76%. This means that 73.76% of the estimated eligibles in the sample frame were contacted. This rate was highest for Group 2 occupants (98.51%) and lowest for Group 2 residents (47.38%). Contact rates for Group 2 were expected to be very high, since the majority of Group 2 preregistrants were self-identifiers (they

initiated the contact, thereby becoming part of the sample). In fact, the eligibility rate for all self-identifiers presented in Table 4 was 93.41%. On the other hand, the contact rate for list cases was 60.36%. This means that more than 3 in 5 estimated eligible cases obtained from lists resulted in a successful contact. This rate surpassed the prepecified goal of 70% set by the study, which was based on an estimation of a reasonable target to provide adequate precision in the survey estimates.

The overall cooperation rate was 80.52%. This means that of the estimated eligible who were contacted, an interview was completed with 80.52%. This rate surpassed the prepecified goal of 70% set by the study. This rate was highest for Group 2 occupants (91.34%) and lowest for Group 2 residents (61.17%). As expected, the cooperation rate was higher among self-identifiers (92.53%) than among list cases (66.05%).

The overall response rate was 59.40%. This means that interviews were completed with 59.40% of the estimated eligible sample. This surpassed the prespecified goal of 49%, which was based on contact and cooperation rate targets of 70%. Response rates by Group also surpassed 49% (57.73% for Group 1 and 64.87% for Group 2). Response rate was highest for Group 2 occupants (95.69%) and lowest for Group 2 residents (29.10%). As expected, the response rate for self-identifiers (86.44%) was much higher than that for listed cases (39.86%).

IV. Discussion

This report presented explanations of and calculations for the outcome rates used in the World Trade Center Health Registry. The process of computing these rates made it evident that careful consideration of sample-specific attributes is very important. In a study where one person can represent up to four completed “interviews,” care should be taken to assure that the correct base, or denominator, is used for measuring enrollment, coverage, or response. The calculation of the overall true eligible population showed that different techniques can provide different results and unless care is taken to use the correct denominator for each rate, results can be misleading.

In the end, the Registry obtained completed interviews with nearly 1 in 5 eligible sample members. The sample database covered nearly 1 in 3. Of those for whom eligibility was determined, more than 3 in 4 were eligible. Of those believed to be eligible, almost 3 in 4 were

contacted and more than 4 in 5 of the estimated eligible contacted were interviewed. These results provide some context for the interpretation of rates and estimates produced by the study, but they do not provide an indication of how well those sampled, contacted, or interviewed represent the true eligible population. Direct measures of this are very costly to obtain and are beyond the scope of the project. Indirect measures are discussed elsewhere (Murphy, 2005) and it is believed that those who were registered generally represent the true eligible population, as was intended.

Table 2. WTC Health Registry Final Outcome Rates

Component/Rate	Group 1:	Group 1:	Group 1:	Group 1:	Group 1:	Group 2:	Group 2:	Group 2:	Group 0:	Unknown	Total	Total	Total
	Workers	Residents	Schools	Occupants	Total	Residents	Occupants	Total	Residents	Group / Type	Residents	Occupants	Overall
Total Preregistrants	51,899	16,354	2,608	84,435	155,297	16,835	20,933	37,767	898	3,982	34,087	105,368	197,944
Total Eligibles	32,705	7,832	1,833	7,940	50,310	4,680	19,974	24,654	898	0	13,410	27,914	75,862
Completes	30,665	7,511	1,736	7,905	47,817	4,327	18,477	22,804	816	0	12,654	26,382	71,437
Incomplete Eligibles	2,040	321	97	35	2,493	353	1,497	1,850	82	0	756	1,532	4,425
Completes w/o Hierarchy	30,665	8,170	2,646	10,393	51,874	5,319	33,094	38,413	1,176	0	14,665	43,487	91,463
Known Ineligibles	4,841	1,772	154	11,516	18,283	597	303	900	0	3,557	2,369	11,819	22,740
Indeterminate Eligibility	14,351	6,750	621	23,363	45,085	11,558	655	12,213	0	425	18,308	24,018	57,723
Total Contacted	42,027	12,196	2,229	25,675	82,127	7,976	20,535	28,511	898	3,776	21,070	46,210	115,312
Contacted - Indeterminate Eligibility	4,481	2,592	242	6,219	13,534	2,699	258	2,957	0	219	5,291	6,477	16,710
True Eligible Population**	91,469	21,926	15,197	62,092	190,684	35,585	300,000	335,585	0	0	57,511	362,092	526,269
w/ Hierarchy Applied	91,469	20,157	9,971	47,228	168,825	28,948	167,496	196,444	0	0	49,106	214,723	365,269
Enrollment Rate	33.53%	37.26%	17.41%	16.74%	27.20%	14.95%	11.03%	11.45%	.	.	25.50%	12.01%	17.38%
Coverage Rate	49.42%	66.16%	24.13%	37.00%	49.39%	51.58%	12.31%	18.55%	.	.	58.99%	20.86%	32.93%
Eligibility Rate	87.11%	81.55%	92.25%	40.81%	73.35%	88.69%	98.51%	96.48%	.	.	84.99%	70.25%	76.94%
Contact Rate	80.98%	74.58%	85.47%	59.96%	72.25%	47.38%	98.10%	75.49%	.	.	61.81%	72.49%	73.76%
Cooperation Rate	83.77%	75.52%	84.43%	75.44%	79.38%	61.17%	91.34%	82.90%	.	.	70.67%	81.26%	80.52%
Response Rate	67.83%	56.32%	72.16%	45.24%	57.35%	28.98%	89.61%	62.58%	.	.	43.68%	58.90%	59.40%

* Final resident completes in ZIP Codes 10002 and 10013 whose residence on September 11, 2001 was on or North of Canal

** The total overall true eligible population is estimated at 410,761 persons. The total of 526,269 in this table represents person-types for enrollment calculation.

Data are based on the November 2005 WTCHR database.

Table 3. WTC Health Registry Final Outcome Rates, Excluding Port Authority Cases

Component/Rate	Group 1:	Group 1:	Group 1:	Group 1:	Group 1:	Group 2:	Group 2:	Group 2:	Group 0:	Unknown	Total	Total	Total
	Workers	Residents	Schools	Occupants	Total	Residents	Occupants	Total	Residents	Group / Type	Residents	Occupants	Overall
Total Preregistrants	50,877	16,320	2,599	9,394	79,190	16,812	19,168	35,980	870	3,982	34,002	28,562	120,022
Total Eligibles	31,687	7,799	1,825	6,192	47,503	4,657	18,213	22,870	870	0	13,326	24,405	71,243
Completes	29,827	7,482	1,730	6,169	45,208	4,308	17,243	21,551	796	0	12,586	23,412	67,555
Incomplete Eligibles	1,860	317	95	23	2,295	349	970	1,319	74	0	740	993	3,688
Completes w/o Hierarchy	29,827	8,140	2,640	8,501	49,108	5,297	31,649	36,946	1,148	0	14,585	40,150	87,202
Known Ineligibles	4,840	1,771	153	1,093	7,857	597	303	900	0	3,557	2,368	1,396	12,314
Indeterminate Eligibility	14,350	6,750	621	2,109	23,830	11,558	652	12,210	0	425	18,308	2,761	36,465
Total Contacted	41,007	12,162	2,220	8,203	63,592	7,953	18,773	26,726	870	3,776	20,985	26,976	94,964
Contacted - Indeterminate Eligibility	4,480	2,592	242	918	8,232	2,699	257	2,956	0	219	5,291	1,175	11,407
True Eligible Population**	91,469	21,926	15,197	62,092	190,684	35,585	300,000	335,585	0	0	57,511	362,092	526,269
w/ Hierarchy Applied	91,469	20,157	9,971	47,228	168,825	28,948	167,496	196,444	0	0	49,106	214,723	365,269
Enrollment Rate	32.61%	37.12%	17.37%	13.69%	25.75%	14.89%	10.55%	11.01%	.	.	25.36%	11.09%	16.57%
Coverage Rate	48.25%	65.98%	24.05%	16.91%	40.25%	51.48%	11.26%	17.62%	.	.	58.79%	12.58%	28.02%
Eligibility Rate	86.75%	81.49%	92.26%	85.00%	85.81%	88.64%	98.36%	96.21%	.	.	84.91%	94.59%	85.26%
Contact Rate	80.60%	74.52%	85.42%	87.32%	80.30%	47.31%	97.94%	74.28%	.	.	61.72%	94.45%	79.12%
Cooperation Rate	83.85%	75.49%	84.46%	88.48%	82.85%	61.11%	93.38%	83.81%	.	.	70.63%	91.75%	83.43%
Response Rate	67.58%	56.26%	72.14%	77.26%	66.53%	28.91%	91.45%	62.25%	.	.	43.59%	86.66%	66.01%

* Final resident completes in ZIP Codes 10002 and 10013 whose residence on September 11, 2001 was on or North of Canal

** The total overall true eligible population is estimated at 410,761 persons. The total of 526,269 in this table represents person-types for enrollment calculation.

Data are based on the November 2005 WTCHR database.

Table 4. WTC Health Registry Final Outcome Rates, List Cases Only

Component/Rate	Group 1:	Group 1:	Group 1:	Group 1:	Group 1:	Group 2:	Group 2:	Group 2:	Group 0:	Unknown	Total	Total	Total
	Workers	Residents	Schools	Occupants	Total	Residents	Occupants	Total	Residents	Group / Type	Residents	Occupants	Overall
Total Preregistrants	31,127	9,245	753	78,665	119,790	10,250	4,353	14,603	144	0	19,639	83,018	134,537
Total Eligibles	12,892	2,224	346	2,886	18,348	1,353	4,344	5,697	144	0	3,721	7,230	24,189
Completes	11,387	2,045	298	2,861	16,591	1,167	3,371	4,538	110	0	3,322	6,232	21,239
Incomplete Eligibles	1,505	179	48	25	1,757	186	973	1,159	34	0	399	998	2,950
Completes w/o Hierarchy	11,387	2,241	397	3,367	17,392	1,398	6,492	7,890	211	0	3,850	9,859	25,493
Known Ineligibles	4,643	1,617	40	11,274	17,574	420	2	422	0	0	2,037	11,276	17,996
Indeterminate Eligibility	13,590	5,404	367	22,889	42,250	8,477	6	8,483	0	0	13,881	22,895	50,733
Total Contacted	21,739	5,776	513	20,139	48,167	3,422	4,348	7,770	144	0	9,342	24,487	56,081
Contacted - Indeterminate Eligibility	4,204	1,935	127	5,979	12,245	1,649	2	1,651	0	0	3,584	5,981	13,896
True Eligible Population**	91,469	21,926	15,197	62,092	190,684	35,585	300,000	335,585	0	0	57,511	362,092	526,269
w/ Hierarchy Applied	91,469	20,157	9,971	47,228	168,825	28,948	167,496	196,444	0	0	49,106	214,723	365,269
Enrollment Rate	12.45%	10.22%	2.61%	5.42%	9.12%	3.93%	2.16%	2.35%	.	.	6.69%	2.72%	4.84%
Coverage Rate	25.02%	26.56%	6.77%	15.99%	23.65%	27.02%	2.60%	6.92%	.	.	25.84%	7.53%	14.59%
Eligibility Rate	73.52%	57.90%	89.64%	20.38%	51.08%	76.31%	99.95%	93.10%	.	.	64.62%	39.07%	57.34%
Contact Rate	69.84%	62.48%	68.13%	54.36%	61.62%	33.39%	99.91%	53.21%	.	.	47.57%	59.15%	60.36%
Cooperation Rate	71.25%	61.15%	64.81%	69.70%	67.44%	44.69%	77.57%	62.73%	.	.	55.03%	65.14%	66.05%
Response Rate	49.76%	38.20%	44.15%	37.89%	41.55%	14.92%	77.49%	33.38%	.	.	26.18%	38.53%	39.86%

* Final resident completes in ZIP Codes 10002 and 10013 whose residence on September 11, 2001 was on or North of Canal

** The total overall true eligible population is estimated at 410,761 persons. The total of 526,269 in this table represents person-types for enrollment calculation.

Data are based on the November 2005 WTCHR database.

Table 5. WTC Health Registry Final Outcome Rates, Self-Identified Cases Only

Component/Rate	Group 1:	Group 1:	Group 1:	Group 1:	Group 1:	Group 2:	Group 2:	Group 2:	Group 0:	Unknown	Total	Total	Total
	Workers	Residents	Schools	Occupants	Total	Residents	Occupants	Total	Residents	Group / Type	Residents	Occupants	Overall
Total Preregistrants	20,772	7,109	1,855	5,770	35,506	6,585	16,580	23,165	754	3,982	14,448	22,350	63,407
Total Eligibles	19,813	5,608	1,487	5,054	31,962	3,327	15,630	18,957	754	0	9,689	20,684	51,673
Completes	19,278	5,466	1,438	5,044	31,226	3,160	15,106	18,266	706	0	9,332	20,150	50,198
Incomplete Eligibles	535	142	49	10	736	167	524	691	48	0	357	534	1,475
Completes w/o Hierarchy	19,278	5,929	2,249	7,026	34,482	3,921	26,602	30,523	965	0	10,815	33,628	65,970
Known Ineligibles	198	155	114	242	709	177	301	478	0	3,557	332	543	4,744
Indeterminate Eligibility	761	1,346	254	474	2,835	3,081	649	3,730	0	425	4,427	1,123	6,990
Total Contacted	20,288	6,420	1,716	5,536	33,960	4,554	16,187	20,741	754	3,776	11,728	21,723	59,231
Contacted - Indeterminate Eligibility	277	657	115	240	1,289	1,050	256	1,306	0	219	1,707	496	2,814
True Eligible Population**	91,469	21,926	15,197	62,092	190,684	35,585	300,000	335,585	0	0	57,511	362,092	526,269
w/ Hierarchy Applied	91,469	20,157	9,971	47,228	168,825	28,948	167,496	196,444	0	0	49,106	214,723	365,269
Enrollment Rate	21.08%	27.04%	14.80%	11.32%	18.08%	11.02%	8.87%	9.10%	.	.	18.81%	9.29%	12.54%
Coverage Rate	22.48%	34.32%	17.28%	11.66%	20.57%	21.60%	9.71%	11.50%	.	.	28.45%	10.14%	15.90%
Eligibility Rate	99.01%	97.31%	92.88%	95.43%	97.83%	94.95%	98.11%	97.54%	.	.	96.69%	97.44%	91.59%
Contact Rate	97.67%	90.31%	92.51%	95.94%	95.65%	69.16%	97.63%	89.54%	.	.	81.17%	97.19%	93.41%
Cooperation Rate	95.97%	87.49%	90.22%	95.48%	93.99%	73.08%	95.12%	90.29%	.	.	82.30%	95.19%	92.53%
Response Rate	93.74%	79.01%	83.46%	91.60%	89.90%	50.54%	92.86%	80.84%	.	.	66.80%	92.52%	86.44%

* Final resident completes in ZIP Codes 10002 and 10013 whose residence on September 11, 2001 was on or North of Canal

** The total overall true eligible population is estimated at 410,761 persons. The total of 526,269 in this table represents person-types for enrollment calculation.

Data are based on the November 2005 WTCHR database.

V. References

- The American Association for Public Opinion Research (2000). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, Michigan: AAPOR.
http://www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/standard_definitions
- Brackbill, R.M., and Thomas, P. (2003). *Protocol for the World Trade Center Registry*. New York City Department of Health and Mental Hygiene in collaboration with Agency for Toxic Substances and Disease Registry. New York, NY.
- Dolan, M., Murphy, J., Thalji, L., and Pulliam, P. (2005). *World Trade Center Health Registry: Sample Building and Denominator Estimation*. RTI International. Research Triangle Park, NC.
- Murphy, J., Brackbill, R.M., Sapp, J.H., Thalji, L., and Pulliam, P. (2005). *An Analysis of Nonresponse Bias in the World Trade Center Health Registry*. Presented at the 2005 meeting of the American Statistical Association. Minneapolis, MN.

Appendix A

WTC Health Registry Final Disposition Codes

Disposition	Description
227	9/11 Casualty
622	Deduplicated - Duplicate
656	Partial Interview Completed
657	Interview Completed
658	Final Refusal by R
659	Final Ineligible - Screened
662	Final Unlocatable
663	Final Noninterview - Other
696	Duplicate Interview
705	Duplicate Deactivated
740	Not sufficient information for tracing or contacting
790	Deduplicated - Indeterminate - Duplicate
806	9/11 Casualty Duplicate
808	Final Ineligible - Postcard
810	9/11 Casualty Duplicate - Possible
821	Finalized - Pending Group 2 Residents - Human Contact
822	Finalized - Pending Group 2 Residents - No Human Contact
823	Finalized - Pending TIO Cases - Did Pursue, Human Contact
824	Finalized - Pending TIO Cases - Did Pursue, No Human Contact
828	Finalized - Port Authority - Phone and Address, Did not Pursue, Resource Constraints
829	Finalized - Port Authority - Address only, Did not Pursue, Resource Constraints
834	Finalized - Pending Field Cases - Did pursue, Human Contact
835	Finalized - Pending Field Cases - Did pursue, No Human Contact
836	Finalized - Pending Field Cases - Out of NYC metro area, Did not pursue, resource constraints
839	Final - Eligible Breakoffs
840	Final - Person not available
841	Final - Unable to enroll in the Registry
842	Refusal by other

Appendix B

Cell Definitions and Example Interpretations for Outcome Rates and Rate Components (Table 2)

Rate / Component	Definition	Example Interpretation
Total Preregistrants	All cases in the Preregistrant database (after deduplication).	There are 197,944 records in the preregistrant database after deduplication.
Total Eligibles	Cases with a finaldisposition of 656, 657, or 839.	There are 75,862 cases that are known to be eligible in the preregistrant database.
Completes	Cases with a finaldisposition of 656, 657.	A total of 71,437 people completed an interview.
Incomplete Eligibles	Cases with a finaldisposition of 839.	A total of 4,425 known eligible cases did not complete an interview.
Completes W/o Hierarchy	Cases with a sampletypedetail and group matching the column definition, and finaldisposition of 656, 657, or 839.	There are a total of 91,463 person-type combinations represented by the 71,437 completed interviews.
Known Ineligibles	Cases with a finaldisposition of 659, 808, or 841.	A total of 22,763 cases were determined to be ineligible.
Indeterminate Eligibility	Cases with a finaldisposition of 662, 740, 822, 824, 835, 836, 663, 810, 840, 658, 821, 823, 834, or 842.	A total of 57,723 cases in the preregistrant database were of indeterminate eligibility.
Total Contacted	Cases with a finaldisposition of 656, 657, 658, 839, 659, 808, 841, 821, 823 or 824.	A total of 115,312 cases were contacted by the Registry.
Contacted - Indeterminate Eligibility	Cases with a finaldisposition of 658, 821, 823, or 834.	Of those contacted, 16,710 were of indeterminate eligibility.
True Eligible Population	From denominator estimates.	A total of 526,269 person-type combinations were estimated to be eligible in the population.
True Eligible Population w/ Hierarchy	See Section II.E. of this report.	After adjusting for the sample type hierarchy, the sum of estimated true eligible individuals for use in the coverage rate is 365,269.
Enrollment Rate	Completes w/o Hierarchy / True Eligible Population	A total of 17.38% of the person-type combinations in the true eligible population were accounted for in Registry interviews.
Coverage Rate	$(\text{Total Eligibles} + (\text{Eligibility Rate} * \text{Indeterminate Eligibility})) / \text{True Eligible Population w/ Hierarchy}$	The preregistrant database included an estimated 32.93% of true eligibles.
Eligibility Rate	$\text{Total Eligibles} / (\text{Known Eligibles} + \text{Known Ineligibles})$	Of those for whom eligibility was determined, 76.94% were eligible.
Contact Rate	$(\text{Eligibility Rate} * \text{Total Contacted}) / (\text{Total Eligibles} + (\text{Eligibility Rate} * \text{Indeterminate Eligibility}))$	Of the estimated eligible in the preregistrant database, 73.76% were contacted by the Registry.
Cooperation Rate	$\text{Completes w/ Hierarchy} / (\text{Eligibility Rate} * \text{Total Contacted})$	Of the estimated eligible in the preregistrant database who were contacted, 80.52% cooperated to be interviewed.
Response Rate	$\text{Completes w/ Hierarchy} / (\text{Total Eligibles} + (\text{Eligibility Rate} * \text{Indeterminate Eligibility}))$	Of the estimated eligible in the preregistrant database, 59.40% cooperated to be interviewed.

Appendix C

SAS Code for Outcome Rates Components

```
*****
*
* PROGRAM: OUTCOMES_RATES.SAS
* PURPOSE: TO CALCULATE OUTCOME RATES COMPONENTS
*           FOR THE WORLD TRADE CENTER HEALTH REGISTRY
*
*****

***STEP 1: DEFINE PATH AND FORMATS;

libname w '<INSERT CORRECT PATH HERE>';
options nofmterr;

proc format;
value typegroup
10=' 7-Group 0 Resident'
11=' 2-Group 1 Resident'
12=' 5-Group 2 Resident'
21=' 3-Group 1 Student'
31=' 4-Group 1 Occupant'
32=' 6-Group 2 Occupant'
41=' 1-Group 1 Worker'
55=' 8-Group/Type Unknown';

***STEP 2: READ IN THE DATA FROM CSV FILES;

*PREREGISTRANT DATA;
data w.v_preregistrantmaster; infile 'c:\w\y_preregistrantmaster.csv' dsd dlm=', ' missover pad firstobs=2;
input listsourceid $ caseid $ samplegroup $ samplotype $ samplegroupfinal $ samplotypefinal;
proc sort; by caseid; run;

*ADDITIONAL PREREGISTRANT DATA;
data w.tempnewpreregistrantcolumns; infile 'c:\w\tempnewpreregistrantcolumns.csv' dsd dlm=', ' missover pad
firstobs=2;
input preregistrantid $ caseid $ sampltypedetail $ prefinaldisposition $ finaldisposition $ contacted $;
proc sort; by caseid; run;

*QUESTIONNAIRE RESPONSES;
data w.transposedresponses; infile 'c:\w\transposedresponses.csv' dsd dlm=', ' missover pad firstobs=2;
input caseid $ resident occ_transit responder si_worker student staff build_code build_name $ build_address $;
proc sort; by caseid; run;

*ELIGIBILITY INDICATOR;
data w.questionresponses; infile 'c:\w\questionresponses.csv' dsd dlm=', ' missover pad firstobs=2;
input caseid $ eligible $;
proc sort; by caseid; run;

*QUESTIONNAIRE DATA FOR ELIGIBLE BREAKOFFS;
data w.eligiblebreakoffs; infile 'c:\w\eligiblebreakoffs.csv' dsd dlm=', ' missover pad firstobs=2;
input caseid $ resident occ_transit responder si_worker student staff;
proc sort; by caseid; run;

*DERIVED VARIABLE BUILD_CODE2 FROM DAVID WU;
data w.derived2; infile 'c:\w\derived2.csv' dsd dlm=', ' missover pad firstobs=2;
input caseid $ build_code2 census_tract census_block;
proc sort; by caseid; run;

*RESIDENT GROUP INDICATOR FOR 891 ELIGIBLE BREAKOFFS FROM DAVID WU;
data w.residents891; infile 'c:\w\residents891.csv' dsd dlm=', ' missover pad firstobs=2;
input caseid $ resident_group;
proc sort; by caseid; run;

*OCCUPANTS GROUP INDICATOR FOR 2408 ELIGIBLE BREAKOFFS FROM DAVID WU;
data w.occupants2408; infile 'c:\w\occupants2408.csv' dsd dlm=', ' missover pad firstobs=2;
input caseid $ build_code2;
proc sort; by caseid; run;

***STEP 3: COMBINE ALL DATA FROM ABOVE PLUS GEOCODE_RESIDENT_LOCATION FOR RESIDENT GROUP DATA FROM DAVID WU AND
CREATE NEW VARIABLES;
data all; merge w.derived2 w.eligiblebreakoffs w.residents891 w.occupants2408 w.questionresponses
w.tempnewpreregistrantcolumns
w.transposedresponses w.v_preregistrantmaster (in=inpr) w.geocode_resident_location; by caseid;

*KEEP ONLY THOSE CASES UNIQUE IN THE PREREGISTRANT DATABASE;
if inpr=1 and finaldisposition~='';
```

```

*DEFINE COMPLETES;
if final disposition in('656' '657') then comp=1;

*ASSIGN FORMAT FOR TYPE-GROUP VARIABLE;
format tg typegroup.;

*CREATE VARIABLES FOR THE FOUR SAMPLE TYPES W/O HIERARCHY BASED ON SAMPLETYPEDETAIL;
*OCCUPANTS/PASSERSBY;
if sampl etypedetail in('2','6','9','10','11','12','13','14','15','16','17','18') then occ=1; else occ=0;
*STUDENTS/STAFF;
if sampl etypedetail in('4','5','7','8','9','10','11','12','15','16','17','18','20','21','22','23') then
stu=1; else stu=0;
*RESIDENTS;
if sampl etypedetail in('1','2','3','4','5','6','7','8','9','10','11','12') then res=1; else res=0;
*WORKERS/VOLUNTEERS;
if sampl etypedetail in('3','6','7','8','9','10','14','15','16','19','20','21') then rrw=1; else rrw=0;

*RESIDENT GROUP INDICATOR USING ALL AVAILABLE INFORMATION - W/O HIERARCHY;
if res=1 then do;
  if resident_group=. then res_group=resident_group;
  if resident_group=. then do;
    if sampl etypefinal='1' and sampl egroupfinal='2' then res_group=2;
    if sampl etypefinal='1' and sampl egroupfinal='1' then res_group=1;
    if sampl etypefinal ~='1' then do;
      if sampl etype='1' and sampl egroup='2' then res_group=2;
      if sampl etype='1' and sampl egroup='1' then res_group=1;
      if sampl etype~='1' then do;
        if sampl egroupfinal='2' then res_group=2;
        if sampl egroupfinal='1' then res_group=1;
        if sampl egroupfinal=' ' then do;
          if sampl egroup='2' then res_group=2;
          if sampl egroup='1' then res_group=1;
        end;
      end;
    end;
  end;
end; end; end; end; end;

```

```

*OCCUPANT GROUP INDICATION USING ALL AVAILABLE INFORMATION - W/O HIERARCHY;
if occ=1 then do;
    if 1<=build_code2<=39 then occ_group=1;
    else occ_group=2;
end;

***STEP 4: CREATE VARIABLE TG FOR TYPE AND GROUP, USING THE BEST SOURCE OF INFORMATION LAST, SUCH THAT A VALUE
IS ASSIGNED AND THEN REASSIGNED IF BETTER INFORMATION IS AVAILABLE. START WITH SAMPLETYPE AND SAMPLEGROUP WHICH
WERE SET AS PRELOADS, THEN MOVE TO SAMPLETYPEFINAL AND SAMPLEGROUPFINAL WHICH WERE SET IN THE INTERVIEW, THEN
MOVE TO SAMPLETYPEDETAIL, RES_GROUP, AND OCC_GROUP WHICH WERE SET DURING POSTPROCESSING.;

*DEFINE BY SAMPLETYPE AND SAMPLEGROUP;
if sampletype='3' and samplegroup='2' then tg=32;
if sampletype='1' and samplegroup='2' then tg=12;
if sampletype='3' and samplegroup='1' then tg=31;
if sampletype='2' then tg=21;
if sampletype='1' and samplegroup='1' then tg=11;
if sampletype='4' then tg=41;

*SET NO GROUP/TYPE TO 55 WHERE APPLICABLE;
if tg=. then tg=55;

*DEFINE BY SAMPLETYPEFINAL AND SAMPLEGROUPFINAL;
if sampletypefinal='3' and samplegroupfinal='2' then tg=32;
if sampletypefinal='1' and samplegroupfinal='2' then tg=12;
if sampletypefinal='3' and samplegroupfinal='1' then tg=31;
if sampletypefinal='2' then tg=21;
if sampletypefinal='1' and samplegroupfinal='1' then tg=11;
if sampletypefinal='4' then tg=41;

*DEFINE BY SAMPLETYPEDETAIL, BUILD_CODE2, AND RESIDENT_GROUP. GO IN THE REVERSE ORDER OF THE HIERARCHY TO SET TG
EQUAL TO
THE LOWEST HIERARCHY VALUE THAT EVALUATES TRUE FIRST, THEN CHANGE IT TO THE NEXT HIGHEST HIERARCHY LEVEL THAT
EVALUATES
TRUE, AND SO ON.;
*GROUP 2 OCCUPANTS;
if occ=1 then do;
    if occ_group=2 then tg=32;
end;

*GROUP 0 AND 2 RESIDENTS;
if res=1 then do;
    if res_group=0 then tg=10;
    if res_group=2 then tg=12;
end;

*GROUP 1 OCCUPANTS;
if occ=1 then do;
    if occ_group=1 then tg=31;
end;

*GROUP 1 STUDENTS AND STAFF;
if stu=1 then tg=21;

*GROUP 1 RESIDENTS;
if res=1 then do;
    if res_group=1 then tg=11;
end;

*GROUP 1 WORKERS;
if rrw=1 then tg=41;

***STEP 5: CREATE PRIORITY_GROUP VARIABLE BASED ON TG (FINAL SAMPLE GROUP WITH HIERARCHY) AND CREATE OTHER
VARIABLES;
if comp=1 and tg=10 then priority_group=0;
if comp=1 and tg in(12,32) then priority_group=2;
if comp=1 and tg in(11,21,31,41) then priority_group=1;

*CREATE LIST VS. SELF-ID VARIABLE;
list=0;if listsourceid ~in(' -1' '527' '528' '529' '530' '531' '532' '533' '534' '535' '536' '537' '538' '539'
'540'
'541' '542' '543' '544' '545' '546' '547' '548') then list=1;

*NUMBER OF TYPES PER COMPLETE FOR OVERALL DENOMINATOR CALCULATION - SUBTRACT OUT GROUP 0 SINCE THEY ARE NOT
INCLUDED IN
THE OVERALL DENOMINATOR;
if comp=1 then do;
    types=occ+stu+res+rrw;
    if res_group=0 then types=types-1;
end;

*PRINT OUTCOMES RATES VALUES FOR ALL CASES;
proc freq order=formatted; tables tg; title 'Total Preregistrants';
proc freq order=formatted; tables tg; where finaldisposition in ('656', '657', '839'); title 'Eligibles';

```

```

proc freq order=formatted; tables tg; where finaldisposition in('656', '657'); title 'Completes';
proc freq order=formatted; tables tg; where finaldisposition in('839'); title 'Incomplete Eligibles';
proc freq order=formatted; tables rrw res*res_group stu occ*occ_group / list missing; where finaldisposition
in('656', '657'); title 'Completes w/o Hierarchy';
proc freq order=formatted; tables tg; where finaldisposition in('659', '808', '841'); title 'Known Ineligibles';
proc freq order=formatted; tables tg; where finaldisposition in('662', '740', '822', '824', '835', '836', '663',
'810', '840', '658', '821', '823', '834', '842'); title 'Indeterminate Eligibility';
proc freq order=formatted; tables tg; where finaldisposition in('656', '657', '658', '839', '659', '808', '841',
'821', '823', '834'); title 'Total Contacted';
proc freq order=formatted; tables tg; where finaldisposition in('658', '821', '823', '834'); title 'Contacted -
Indeterminate Eligibility';

*PRINT OUTCOMES RATES VALUES MINUS THE PA CASES;
proc freq order=formatted; tables tg; where listsourceid='490'; title 'Total Preregistrants - no PA';
proc freq order=formatted; tables tg; where listsourceid='490' and finaldisposition in('656', '657', '839'); title
'Eligibles - no PA';
proc freq order=formatted; tables tg; where listsourceid='490' and finaldisposition in('656', '657'); title
'Completes - no PA';
proc freq order=formatted; tables tg; where listsourceid='490' and finaldisposition in('839'); title 'Incomplete
Eligibles - no PA';
proc freq order=formatted; tables rrw res*res_group stu occ*occ_group / list missing; where listsourceid='490'
and finaldisposition in('656', '657'); title 'Completes w/o Hierarchy - no PA';
proc freq order=formatted; tables tg; where listsourceid='490' and finaldisposition in('659', '808', '841'); title
'Known Ineligibles - no PA';
proc freq order=formatted; tables tg; where listsourceid='490' and finaldisposition in('662', '740', '822',
'824', '835', '836', '663', '810', '840', '658', '821', '823', '834', '842'); title 'Indeterminate Eligibility -
no PA';
proc freq order=formatted; tables tg; where listsourceid='490' and finaldisposition in('656', '657', '658',
'839', '659', '808', '841', '821', '823', '834'); title 'Total Contacted - no PA';
proc freq order=formatted; tables tg; where listsourceid='490' and finaldisposition in('658', '821', '823',
'834'); title 'Contacted - Indeterminate Eligibility - no PA';

```



```

*PRINT OUTCOMES RATES VALUES FOR LIST CASES;
proc freq order=format; tables tg; where list=1; title 'Total Preregistrants - List';
proc freq order=format; tables tg; where list=1 and final disposition in ('656', '657', '839'); title 'Eligibles - List';
proc freq order=format; tables tg; where list=1 and final disposition in ('656', '657'); title 'Completes - List';
proc freq order=format; tables tg; where list=1 and final disposition in ('839'); title 'Incomplete Eligibles - List';
proc freq order=format; tables rrw res*res_group stu occ*occ_group / list missing; where list=1 and final disposition in ('656', '657'); title 'Completes w/o Hierarchy - List';
proc freq order=format; tables tg; where list=1 and final disposition in ('659', '808', '841'); title 'Known Ineligibles - List';
proc freq order=format; tables tg; where list=1 and final disposition in ('662', '740', '822', '824', '835', '836', '663', '810', '840', '658', '821', '823', '834', '842'); title 'Indeterminate Eligibility - List';
proc freq order=format; tables tg; where list=1 and final disposition in ('656', '657', '658', '839', '659', '808', '841', '821', '823', '834'); title 'Total Contacted - List';
proc freq order=format; tables tg; where list=1 and final disposition in ('658', '821', '823', '834'); title 'Contacted - Indeterminate Eligibility - List';

*PRINT OUTCOMES RATES VALUES FOR SELF ID CASES;
proc freq order=format; tables tg; where list=0; title 'Total Preregistrants - Self ID';
proc freq order=format; tables tg; where list=0 and final disposition in ('656', '657', '839'); title 'Eligibles - Self ID';
proc freq order=format; tables tg; where list=0 and final disposition in ('656', '657'); title 'Completes - Self ID';
proc freq order=format; tables tg; where list=0 and final disposition in ('839'); title 'Incomplete Eligibles - Self ID';
proc freq order=format; tables rrw res*res_group stu occ*occ_group / list missing; where list=0 and final disposition in ('656', '657'); title 'Completes w/o Hierarchy - Self ID';
proc freq order=format; tables tg; where list=0 and final disposition in ('659', '808', '841'); title 'Known Ineligibles - Self ID';
proc freq order=format; tables tg; where list=0 and final disposition in ('662', '740', '822', '824', '835', '836', '663', '810', '840', '658', '821', '823', '834', '842'); title 'Indeterminate Eligibility - Self ID';
proc freq order=format; tables tg; where list=0 and final disposition in ('656', '657', '658', '839', '659', '808', '841', '821', '823', '834'); title 'Total Contacted - Self ID';
proc freq order=format; tables tg; where list=0 and final disposition in ('658', '821', '823', '834'); title 'Contacted - Indeterminate Eligibility - Self ID';
run;

*PRINT CROSSTABS FOR OVERALL DENOMINATOR CALCULATION;
proc freq order=format; tables types; where comp=1 and rrw=1; title 'Number of types per complete - Workers w/o hierarchy';
proc freq order=format; tables res_group*types / list missing; where comp=1 and res_group in(1,2); title 'Number of types per complete - Residents w/o hierarchy';
proc freq order=format; tables types; where comp=1 and stu=1; title 'Number of types per complete - Students/Staff w/o hierarchy';
proc freq order=format; tables occ_group*types / list missing; where comp=1 and occ_group in(1,2); title 'Number of types per complete - Occupants w/o hierarchy';
run;

*CREATE DATA SET WITH PRIORITY GROUP VALUE;
data w.priority_group; set all (keep=caseid priority_group); if priority_group=.;
proc freq order=format; tables priority_group; title 'Priority Group';
run;

```