and Marine Biology & Fisheries, Miller School of Medicine and Rosenstiel School of Marine and Atmospheric Sciences, Clinical Research Building, 10th Floor (R669), 1120 NW 14th St, Miami, FL 33136 (e-mail: lfleming@med.miami.edu).

## Contributors

## Acknowledgments

## Human Participant Protection

## References

1. American Cancer Society. *Cancer Facts and Figures 2007*. Atlanta, GA: American Cancer Society; 2007.

2. Smith RA, Mettlin CJ, Davis KJ, Eyre H. American Cancer Society guidelines for the early detection of cancer. *CA Cancer J Clin.* 2000;50:34–49.

3. Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2005. *CA Cancer J Clin.* 2005;55:31–44.

4. US Preventive Services Task Force. *The Guide to Clinical Preventive Services*. Washington, DC: Agency for Healthcare Research and Quality; 2006. Available at: http://www.ahrq.gov/clinic/pocketgd/pocketgd.htm. Accessed May 19, 2008.

5. Healthy People 2010. Available at: http://www.healthypeople.gov. Accessed August 8, 2008.

6. Barbeau EM, Krieger N, Soobader MJ. Working class matters: socioeconomic disadvantage, race/ethnicity, gender, and smoking in NHIS 2000. *Am J Public Health.* 2004;94:269–278.

7. Caban AJ, Lee DJ, Fleming LE, Gomez-Marin O, LeBlanc W, Pitman T. Obesity in US workers: the National Health Interview Survey, 1986 to 2002. *Am J Public Health.* 2005;95:1614–1622.

8. Sorerholm SC. National Occupational Research Agenda (NORA). Cross-sector research in the second decade. Presented at the 2006 NORA Symposium; Washington, DC; April 2006. Available at: http://www.cdc.gov/niosh/nora/symp06/pdfs/cross06present.pdf. Accessed May 19, 2008.

9. National Center for Health Statistics. *Data File Documentation, National Health Interview Survey, 2000* (machine-readable data file and documentation). Hyattsville, MD: National Center for Health Statistics; 2002.

10. SAS Institute. *Survey Data Analysis*. Available at: http://support.sas.com/rnd/app/da/new/dastat91.html. Accessed May 19, 2008.

11. Jabine JB. *Reporting Chronic Conditions With NHIS*. Hyattsville, MD: National Center for Health Statistics, 1987. Series 2, no. 105. DHHS publication PHS 87−1379.

12. Behrens V, Seligman P, Cameron L, Mathias CG, Fine L. The prevalence of back pain, hand discomfort, and dermatitis in the US working population. *Am J Public Health.* 1994;84:1780–1785.

13. Health and Human Services, Centers for Disease Control and Prevention. *Health Risks in the United States Behavioral Risk Factor Surveillance System at a Glance, 2007.* Available at: http://www.cdc.gov/nccdphp/publications/aag/pdf/brfss.pdf. Accessed May 19, 2008.

14. Carrière P, Baade P, Newman B, Aitken J, Janda M. Cancer screening in Queensland men. *Med J Aust.* 2007;186:404–407.

15. Levy BT, Nordin T, Sinift S, Rosenbaum M, James PA. Why hasn't this patient been screened for colon cancer? *J Am Board Fam Med.* 2007;20:458–468.

16. Miller DP Jr, Brownlee CD, McCoy TP, Pignone MP. The effect of health literacy on knowledge and receipt of colorectal cancer screening: a survey study. *BMC Fam Pract.* 2007;8:16.

17. Bagai A, Parsons K, Malone B, Fantino J, Paszat L, Rabeneck L. Workplace colorectal cancer screening awareness programs: an adjunct to primary care practice? *J Community Health.* 2007;32:157–167.

18. Sorensen G, Barbeau EM. Integrating occupational health, safety and worksite health promotion: opportunities for research and practice. *Med Lav.* 2006;97:240–257.

# Estimating the World Trade Center Tower Population on September 11, 2001: A Capture–Recapture Approach

Joe Murphy, MA

I applied the capture–recapture method to estimate the World Trade Center tower population at the time of the September 11, 2001, terrorist attacks. Available lists helped identify 8965 survivors and 2152 confirmed casualties. The capture–recapture model suggested that an additional 4435 survivors were present, putting the total count of all present at 15 552 (95% confidence interval = 15 216, 15 897). An accurate estimate represents the potential number at risk for trauma as a result of direct exposure to the events of the day. (*Am J Public Health*. 2009;99:65–67. doi: 10.2105/AJPH.2007.124768)

On September 11, 2001, 2 airplanes crashed into the World Trade Center towers in New York City as part of a terrorist attack. In addition to the loss of life, survivors were exposed to hazardous dust and debris from the collapse of the buildings, and many people endured psychological trauma.[1–4] No one has definitively determined the number of persons present in the 2 towers at 8:46 AM on that day when the first airplane crashed into the North Tower, and no one may ever be able to do so. Several estimates have been made previously,[5–7] but none made use of capture–recapture methods, which are useful in estimating population size when enumeration by more-direct methods is not feasible.[8–10]

My objective was to apply the capture–recapture method to 3 list sources of individuals determined to be present inside the World Trade Center towers on September 11 to estimate the number present when the first airplane struck. This number represents a population at risk for long-term health effects as a result of direct exposure to the events of the day.

## METHODS

The World Trade Center Health Registry (WTCHR) is a database used for following up with individuals exposed to the disaster of September 11 to evaluate the short- and long-term physical and mental health effects. Exposed groups, including rescue and recovery workers, residents, students and school staff, building occupants, and passersby, were located and interviewed. Potentially eligible individuals were identified through a multitude of list sources and could also register through the project Web site or toll-free telephone number. In all, 71 437 interviews were completed, including 8965 with individuals present in the World Trade Center towers at 8:46 AM on September 11.

Those present in the towers—a group with high potential for physical and psychological trauma as a result of exposure to the events of the day—were identified from 3 different list sources for the WTCHR. The first list included 3622 individuals who volunteered by Web site or telephone to complete the WTCHR survey (the self-identified list), and were confirmed to be present in the buildings on September 11. The second list included data

**TABLE 1—World Trade Center Health Registry Tower Occupants on September 11, 2001, by List Source**

| List | No. (%) |
|---|---|
| 1 source | 6295 (70.2) |
| Self-identified individuals list (SI) only | 1702 (19.0) |
| Businesses list (BL) only | 270 (3.0) |
| Port Authority list (PA) only | 4323 (48.2) |
| 2 sources | 2496 (27.8) |
| SI and BL | 88 (1.0) |
| SI and PA | 1658 (18.5) |
| BL and PA | 750 (8.4) |
| 3 sources (SI, BL, and PA) | 174 (1.9) |
| Total | 8965 (100.0) |

**TABLE 2—Capture–Recapture Model Estimates of the World Trade Center Tower Population on September 11, 2001**

| Model | Akaike's Information Criterion, No. | Estimate of Survivors, No. (95% CI) | Estimate of Survivors and Casualties, No. |
|---|---|---|---|
| Independent (no interactions) | 324.5 | 12 646 (12 417, 12 914) | 14 798 |
| SI×BL interaction | 273.5 | 12 412 (12 167, 12 655) | 14 564 |
| SI×PA interaction | 220.0 | 11 010 (10 743, 11 315) | 13 162 |
| BL×PA interaction | 103.1 | 13 699 (13 350, 14 049) | 15 851 |
| SI×BL, SI×PA interactions | 98.8 | 10 517 (10 301, 10 763) | 12 669 |
| SI×BL, BL×PA interactions | 75.1 | 13 400 (13 064, 13 745) | 15 552 |
| SI×PA, BL×PA interactions | 104.6 | 14 161 (13 066, 15 642) | 16 313 |

*Note.* CI = confidence interval; SI = self-identified list source; BL = business list source; PA = Port Authority list source.

supplied by businesses (the business list), with office space in the World Trade Center towers identifying employees who were present on the morning of September 11. Interviews conducted with 1282 individuals from this list confirmed that they were present in the buildings on September 11. The third list, obtained from the Port Authority of New York and New Jersey (the Port Authority list), included individuals with security access to the towers as of September 11; from this list, 5247 were confirmed to have been present. I identified individuals included on multiple lists by comparing names, addresses, telephone numbers, dates of birth, and social security numbers with an algorithm developed by Choicemaker Technologies (New York, NY).[11]

Capture–recapture methods that used log-linear modeling estimated the number of individuals not on any list. The log-linear approach models the logarithm of the expected value of each observed list source. Seven samples were observed: $z_{100}$, $z_{010}$, $z_{001}$, $z_{110}$, $z_{101}$, $z_{011}$, and $z_{111}$, where, for instance, $z_{011}$ is the number of individuals in samples 2 and 3 but not in sample 1. The missing sample, $z_{000}$, refers to those not on any list. The most general log-linear model for the 3-sample case is as follows:

$$(1) \quad \log E(z_{ijk}) = u + u_1 I(i = 1)$$
$$+ u_2 I(j = 1) + u_3 I(k = 1)$$
$$+ u_{12} I(i = j = 1)$$
$$+ u_{13} I(i = k = 1)$$
$$+ u_{23} I(j = k = 1)$$
$$+ u_{123} I(i = j = k = 1),$$

where $\log E(z_{ijk})$ = logarithm of the expected value of each observable category, and $I(A)$ is an indicator function for event $A$.[8]

Models were fit with R version 2.6.0 (available at http://www.r-project.org.index.html). which uses the generalized linear model (glm) function to provide population estimates and fit statistics.[12] The models were run with all combinations of interactions between the list sources except for the 3-list interaction term, which was assumed to be 0 ($u_{123}=0$).[8] The model with the smallest Akaike's information criterion was selected as best.[13,14] The estimated number of people not on any list was added to the number on any list and the casualty count to obtain the overall population estimates for the World Trade Center towers at 8:46 AM on September 11.

## RESULTS

A total of 8965 persons identified for the WTCHR indicated being in one of the towers on the morning of September 11. Among this group, 3622 volunteered for the interview via Web site or telephone (the self-identified list), 5247 were included on the Port Authority list, and 1282 were included on the business list. Table 1 presents the number included in each combination of lists, and the results of the capture–recapture model are shown in Table 2. The model with SI×BL and BL×PA interactions had the lowest Akaike's information criterion and was selected as best. Adding the 2152 occupant casualties[15] to the 8965 included on a list source and the 4435 estimated individuals not included on any of the list sources resulted in an estimate of the World Trade Center tower population on the morning of September 11 of 15 552 (95% confidence interval [CI]=15 216, 15 897).

## DISCUSSION

Debate exists concerning the applicability of capture–recapture to epidemiology; indeed, capture–recapture estimates should be interpreted with great caution.[16] Although these methods are not directly comparable to those used in previous studies to measure the World Trade Center tower population on September 11, capture–recapture provided results similar to those from previous analyses. For instance, the Port Authority of New York and New Jersey (A. Reiss, Port Authority of New York and New Jersey, written communication, March 2005) counted 14 417 card accesses to the towers on

Tuesday, May 8, 2001, between 12:00 AM and 8:47 AM, the closest date to September 11 for which data were available. *USA Today* matched floor plans, architectural drawings, and photographs to the accounts of many survivors and estimated that between 12 000 and 16 000 people were in the towers at the time of the first crash.[5] The National Institute of Standards and Technology contacted a stratified random sample of individuals from a list of persons with building access as of September 11 and estimated that 17 400 (95% CI=16 260, 18 620) were present.[6]

The 3 necessary assumptions of the capture–recapture methodology are perfect matching, independence of lists, and homogeneity.[9] Perfect matching assumes that individuals on different lists can be matched without error. Despite considerable effort devoted to proper matching in this study, some true matches or nonmatches may have been missed. Independence of lists assumes that the probability of an individual appearing on 1 list does not depend on inclusion on another list. List dependence is controlled for when 3 or more lists are used and interactions between are included in the model, as was done in this study. Homogeneity assumes that the probability of inclusion on a list does not vary among individuals. Possible heterogeneity in the capture probabilities was explored with the Rcapture plot descriptive function,[12] which suggested that heterogeneity was not a problem for this analysis. However, individuals may have varied in their capture probabilities in an unmeasured way.

Although the capture–recapture method has limitations and may not provide a definitive estimate of the World Trade Center tower population on September 11, this study confirmed that it was a useful method in this context that provided estimates similar to those obtained in other studies. ∎

## About the Author
*The author is with RTI International, Chicago, IL.*
*Requests for reprints should be sent to Joe Murphy, MA, RTI International, 230 W Monroe St, Suite 2100, Chicago, IL 60606 (e-mail: jmurphy@rti.org).*
*This brief was accepted May 18, 2008.*

## References
1. Landrigan PJ, Lioy PJ, Thurston G, et al. Health and environmental consequences of the World Trade Center disaster. *Environ Health Perspect.* 2004;112:731–739.

2. Brackbill RM, Thorpe LE, DiGrande L, et al. Surveillance for World Trade Center disaster health effects among survivors of collapsed and damaged buildings. *MMWR Surveill Summ.* 2006;55(SS-02):1–18.

3. Galea S, Vlahov D, Resnick H, et al. Trends of probable post-traumatic stress disorder in New York City after the September 11 terrorist attacks. *Am J Epidemiol.* 2003;158:514–524.

4. Schlenger WE, Caddell JM, Ebert L. Psychological reactions to terrorist attacks: findings from the National Study of Americans' Reactions to September 11. *JAMA.* 2002;288:581–588.

5. Cauchon D. For many on September 11, survival was no accident. *USA Today.* December 20, 2001. Available at: http://www.usatoday.com/news/sept11/2001/12/19/usatcov-wtcsurvival.htm. Accessed August 16, 2007.

6. National Institute of Standards and Technology (NIST). Interim report on telephone interviews. Available at: http://wtc.nist.gov/progress_report_june04/appendixo.pdf. Accessed January 24, 2008.

7. Murphy J, Brackbill RM, Thalji L, Dolan M, Pulliam P, Walker D. Measuring and maximizing coverage in the World Trade Center Health Registry. *Stat Med.* 2007;26:1688–1701.

8. Chao A, Tsay P, Lin S, Shau W, Chao E. The applications of capture-recapture models to epidemiological data. *Stat Med.* 2001;20:3123–3157.

9. Darroch J, Fienberg S, Glonek G, Junker B. A three-sample multiple-capture approach to census population estimation with heterogeneous catchability. *J Am Stat Assoc.* 1993;88:1137–1148.

10. Feinberg S, Meyer M. Log-linear models and categorical data analysis with psychometric and econometric applications. *J Econom.* 1983;22:191–214.

11. Murphy J, Pulliam P, Lucas R. Sample frame deduplication in the World Trade Center Health Registry. In: *Proceedings of the American Statistical Association, Survey Research Methods Section.* Alexandria, VA: American Statistical Association; 2004:4072–4076.

12. Baillargeon S, Rivest L. Rcapture: loglinear models for capture-recapture in R. *J Stat Softw.* 2007;17:1–31.

13. Hook E, Regal R. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev.* 1995;17:243–264.

14. Aaron D, Chang Y, Markovic N, LaPorte R. Estimating the lesbian population: a capture-recapture approach. *J Epidemiol Community Health.* 2003;57:207–209.

15. National Commission on Terrorist Attacks. *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States.* New York, NY: WW Norton & Co; 2004.

16. Cormack RM, Chang YF, Smith GS. Estimating deaths from industrial injury by capture-recapture: a cautionary tale. *Int J Epidemiol.* 2000;29:1053–1059.