Statistical Issues in Developing Sample Designs for Surveys Requiring Rapid Response

Allison M. Deal and William D. Kalsbeek
University of North Carolina at Chapel Hill Survey Research Unit

Abstract

Special design issues arise when adapting ongoing data gathering systems for quick turn-around assessments, for example during public health emergencies. One of these issues is determining adequate sample size for subgroup estimates. For example, if a comparison between overlapping population subgroups is desired (e.g. when comparing a state coverage rate to one of its region's), special consideration must be given to the overlap when determining sample size requirements. The need to estimate influenza vaccination coverage, in the context of adapting CDC's Behavioral Risk Factor Surveillance System (BRFSS), is used to illustrate these issues and their resolution.

Key Words: Rapid Response, Overlapping Samples, BRFSS, Sample Size

1. Introduction

The relative statistical and practical feasibility of using the Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) to deliver rapid turn-around survey findings is investigated, in addition to considering modifications to its design and data production process. This is done by focusing on the production of rapid response influenza vaccination coverage estimates in North Carolina, but a broader plane of application, such as any public health emergency, underlies this effort.

1.1 Rapid Response Surveillance

In public health emergencies, such as natural disasters, terrorist attacks, and shortages of vaccines, surveillance information is needed quickly and standard survey procedures cannot always be used because of their slow turn-around time. There is little opportunity to assemble substantial financial and organizational resources, and there is limited time for deliberate planning and careful execution. These time limits have implications on the development, evaluation, and implementation of data gathering methods, and they leave little time to generate findings and employ methods to compensate for known study limitations (Brookmeyer and Stroup, 2004).

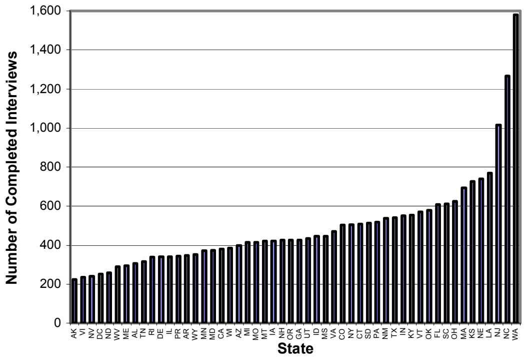
When emergencies like these arise, resources need to be pooled quickly in order to accomplish the same goals in a much shorter time frame. If some of these steps are done and/or planned for ahead of time, surveys following emergencies can be conducted faster and more efficiently, especially if pre-existing survey systems are used. In the case of a natural disaster, public health officials are immediately interested in its impact on the members of the population in the affected region. When terrorists attacked the World Trade Center in September 2001, health officials wanted to know how it affected the people of New York City (CDC, 2002). In the case of the shortage of vaccine supply during the 2004-2005 influenza season, the CDC needed to know quickly and continuously thereafter how well the vaccines were being distributed, so they could assess the situation and adjust their strategies accordingly (CDC, 2005).

1.2 BRFSS Background

The BRFSS is an ongoing monthly telephone survey conducted by individual state health departments with technical and methodological assistance provided by CDC. States conduct monthly telephone surveys using a standardized questionnaire to learn about the risk behaviors and health practices among adults. The states send their data to CDC, where the monthly data are combined, weighted, returned with standard tabulations, and published at the end of the year by each state. There are three parts to the BRFSS; the core sections, optional modules, and state-added questions, but not all states include state-added questions. Information from the core sections is collected by every state every month with every state asking the exact same questions in the exact same format. The optional modules and state-added questions are included at the state's discretion, where the optional modules must be asked exactly as CDC has written them, but the state-added questions adhere to no such guidelines (CDC, 2005).

As seen in Figure 1, all of the states complete a different number of interviews; in the 2004 calendar year, Washington completed 18,957 (~1,580 monthly), North Carolina completed 15,205 (~1,267 monthly), Alaska completed 2,688 (~224 monthly), and the Virgin Islands completed 2,812 (~234 monthly). The average number of completed interviews was 5,919 (~500 monthly) (CDC 2004). Washington and North Carolina are exceptions; almost all of the other states are in the range of about 300 to 800 monthly completed interviews.

Figure 1. Average Number of Monthly Completed Interviews



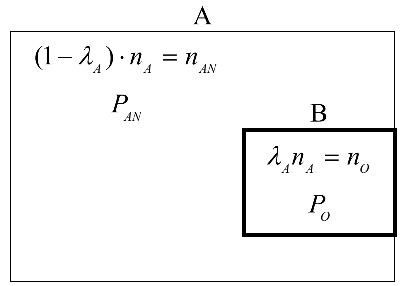
2. Statistical Issues

2.1 Overlapping Samples

In order to be sure that vaccinations are reaching all regions of a state, for example, officials would want to compare regional vaccination coverage estimates to the state coverage estimates as a whole. They want to know where in the state program implementation is ahead of or behind schedule; this would allow the appropriate redirection of resources. Comparing state to regional or city to county vaccination coverage rates involves overlapping, and thus non-independent, comparison groups, so the main statistical issue this paper addresses is measuring the precision of estimates when the samples used for comparison overlap. The developed theory is used to examine power and sample size relationships given the current monthly cycle, keeping in mind that for rapid response situations states would need to produce estimates faster, and their impact on the production of rapid response findings is subsequently discussed (see Section 3). These issues include producing biweekly estimates and modifications to questionnaire design.

2.2 Complete Overlap

If an estimate of a population difference between two overlapping samples, A and B (for example a state and one of its regions) is of interest, where the total sample sizes are n_A and n_B, respectively, let λ_A denote the proportion of people from A that are in the overlap. The proportion of people with the attribute of interest, for example having been vaccinated, will be referred to specifically as the "attribute proportion", i.e. as *P*; this is different from the proportion of the sample in the overlap, which is denoted by λ . The two overlapping samples can be viewed as having two independent components: those in A and *not* the overlap, denoted as "AN" and those *in* the overlap, denoted as "O". For example, P_{AN} is the attribute proportion for those in A but not the overlap, and n_{AN} is the sample size of that component. Note that $n_A = n_O + n_{AN}$ and $n_O = \lambda_A n_A$. The findings in this section, and further derivations, are equivalent to that in Deal (2007).



Letting P_O and n_O represent the attribute proportion being estimated and size of the overlap, the attribute proportion of people in A can be written as $P_A = \lambda_A P_O + (1-\lambda_A) P_{AN}$. If $\hat{P}_A$ and $\hat{P}_B$ are estimates for A and B, respectively, the attribute proportion for those not in the overlap of A and B can be estimated by $\hat{P}_{AN} = \frac{\hat{P}_A - \lambda_A \hat{P}_O}{(1-\lambda_A)}$ and the estimated difference shown to be $\hat{\delta} = \hat{P}_A - \hat{P}_B = (1-\lambda_A)\left(\hat{P}_{AN} - \hat{P}_O\right)$. The variance equation under the alternative hypothesis simplifies to

$$\hat{Var}_A(\hat{\delta}) = \frac{(1-\lambda_A)P_{AN}(1-P_{AN})}{n_A} + \frac{(1-\lambda_A)^2 P_O(1-P_O)}{\lambda_A n_A},$$

since the $(1-\lambda_A)$ term drops out. When calculating power and sample sizes relationships, using standard equations, the approximation by Johnson and Kotz,

$$1-\beta \doteq 1-\frac{1}{2}\left[1+\left\{1-\exp\left(\frac{-2Z_{1-\beta}^2}{\pi}\right)\right\}^{\frac{1}{2}}\right]$$

can be used.

2.2.1 Relative Differences

The following section shows power and sample size relationships for comparing vaccination coverage rates. As the influenza season progresses from October to January, the coverage rates will increase, since they are cumulative, and acceptable regional differences will change for each month. When both the state and regional rates are small, i.e. in the early months, a difference as small as a few percentage points could indicate that the region is lagging far behind in coverage. In the later months, a difference of a few percentage points could indicate less of a difference. For example, if in October, the state vaccination rate is 9.7% and the region is only at 6% coverage, the absolute difference is 3.7%. This is a very large difference from the state level, but only since the state level is fairly small. The relative difference in this case is 38.1%. If in January, the state rate is at 23.4% and there is an absolute difference of 3.7% in the region rate, the relative difference is only 15.8%.

As the state levels continue to rise, the same absolute difference from the regional level becomes less pronounced. Because of this, relative differences in coverage rates will be used since acceptable regional differences will change for each month. The relative difference can be calculated by dividing the difference between the state and regional coverage rate for the month by the state coverage rate for the month. When calculating the power and sample size relationships, the same absolute value of $\hat{\delta}$ was not used for all months. The same relative difference, i.e. the $\hat{\delta}$ that made the relationship between the state and regional coverage rates the same, was used. Graphs for relative differences of 20%, 35%, and 50% are shown to see the decrease in sample sizes as the difference in estimates increases.

2.3 Sample Size and Power

The following calculations give power and sample size estimates that any state interested in monitoring influenza vaccination coverage during October through January could use. The state of North Carolina, which currently has one of the highest monthly sample sizes, and an "average" state are used for discussion. States could use these graphs to see how much they would need to increase their sample sizes to achieve acceptable power.

The following graphs show what sample sizes are needed to achieve certain levels of power. From Figure 2, it can be seen that 80% power for a relative difference of 20% cannot be achieved for any month by a sample of fewer than 3,000 respondents. To achieve 80% power in October, a state would need 8,500 respondents, which is unreasonable. This would require North Carolina to increase their sample size by six fold, and the smallest state to increase its sample size by over forty fold. The required sample sizes for this small relative difference are much larger than the states currently have, and possibly larger than they can feasibly achieve. From Figure 3, we can see that to achieve 80% power with a relative difference of 35%, a large sample would still be needed in October, about 2,700, but fewer than 1,500 would be needed for each of the remaining months. Finally, as seen in Figure 4, if all states could increase their sample size to that of North Carolina, they could all achieve 80% for $R_i = 50\%$.

Figure 2. Sample Size and Power Curves for Relative Difference of 20%

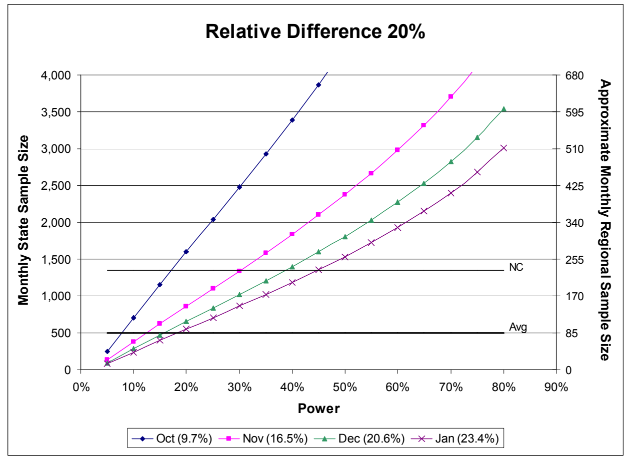


Figure 3. Sample Size and Power Curves for Relative Difference of 35%

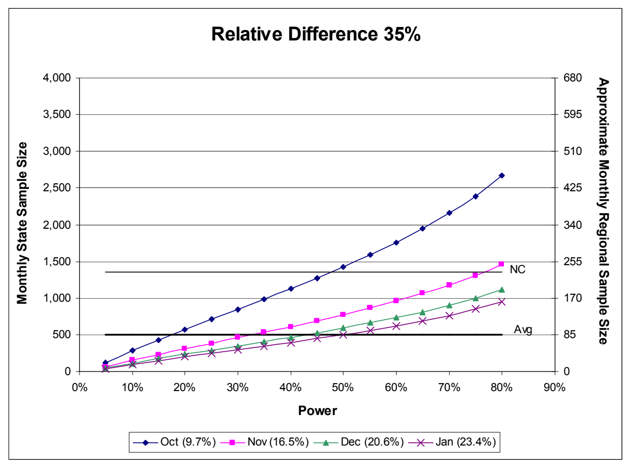
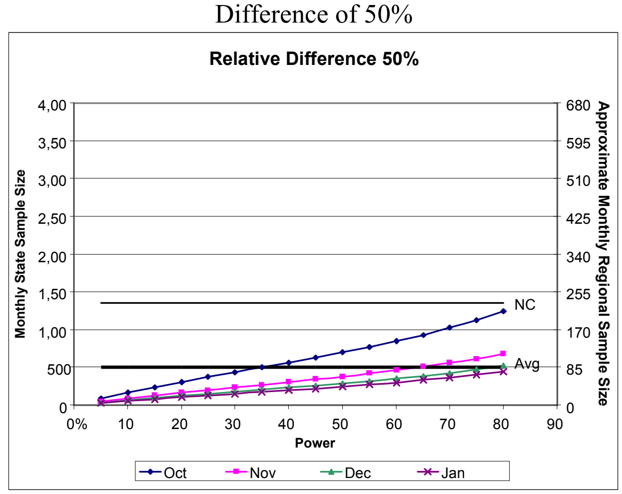


Figure 4. Sample Size and Power Curves for Relative Difference of 50%



These graphs demonstrate that in order to achieve reasonable power, large sample sizes are needed, especially with small relative differences, which might be problematic for smaller states. Detecting small differences in the early months of the season may not be of as much interest since time is needed for the vaccine to start being distributed, and officials would only need to know if there is a huge discrepancy between state and regional rates. If sample sizes are not increased, the power associated with these hypothesis tests will be very low and the tests may be worthless. In order to achieve acceptable levels of power, sample sizes would need to be boosted, especially if officials want to monitor the rates biweekly. The sample sizes are currently not even adequate for monthly estimates, so asking the states to complete the same amount of interviews in half the time given their current resources, for biweekly estimates, is unreasonable

3. Operational Aspects

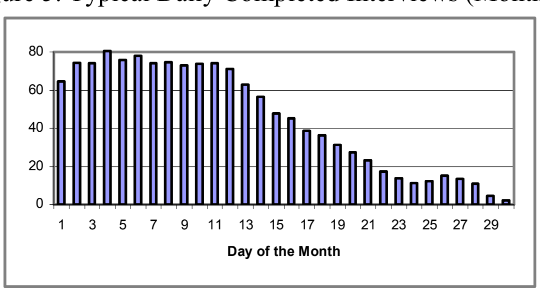
3.1 Biweekly Estimates

Since information is needed more quickly during public health emergencies, operational changes to the survey design would be required in order to produce estimates more frequently; monthly data may not be enough, but biweekly estimates might be appropriate. Investigating the operations of North Carolina's State Center for Health Statistics' (SCHS) BRFSS office allowed the creation of a model for the daily frequency of completed interviews in a typical monthly cycle of calling, and this information provided the basis for an alternative calling cycle. The authors would like to acknowledge with thanks the data supplied by the SCHS.

3.1.1 Current Calling Cycle

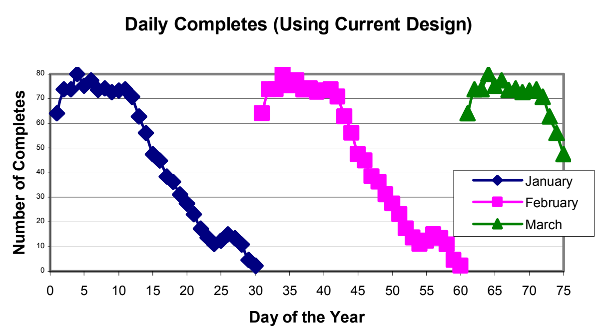
The current monthly cycle, as seen in Figure 5 seems to have a start-up phase, during the first two weeks, and then the daily number of completes drops drastically during the last two weeks, the wrap-up phase.

Figure 5. Typical Daily Completed Interviews (Monthly)



The goal of producing biweekly estimates cannot be done under the current design of a monthly calling cycle. One way it could be done is to split the monthly sample into two halves – those who completed their interview in the first half of the month (early respondents) and those who completed it in the second half (late respondents). This would lead to unequal sample sizes because so many people complete their interviews in the beginning of the month. Link and Beimer (2007) looked at the BRFSS data from 2004 and found that about 80% of adults aged 65 or older were early respondents compared to about 62% of adults aged 18-34 years. The early respondents were more likely to respond that they had received a vaccination than late respondents since there were so many elderly in the early respondent group. These findings suggest that simply dividing the sample into the first and second halves of the month is not ideal, since sample sizes for second half estimates would be substantively smaller than the first half estimates; moreover, since early and late responders are different, they should not be used to estimate the same value. If, for example, calling started on the first day of the year, Figure 6 shows the daily number of completes for the first two and a half months of calling. These values are based on the model found earlier.

Figure 6. Daily Completed Interviews (Current)

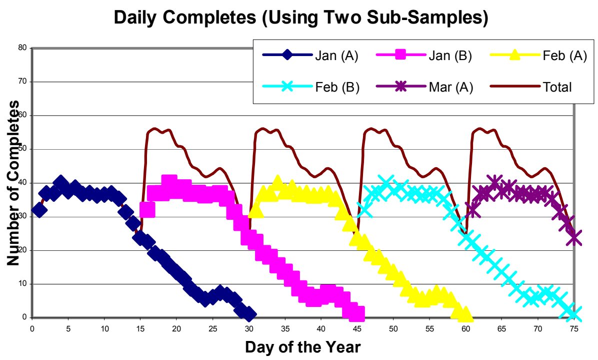


3.1.2 Proposed Calling Cycle

Another possible change in the calling cycle to allow the production of biweekly estimates would be to randomly divide the entire monthly sample into two equal sized sub-samples, say (A) and (B), before calling ever begins. For this example, also assume that the process would start at the beginning of the year. As seen in Figure 7, calling would start on the January (A) sample on Day 1, and would continue for 30 days. On the 16th day of the year, calling would start on the January (B) sample. During Days 16 through 30, calling would be going on simultaneously for both parts of January's sample; the January (A) sample would be in its wrap-up phase, and the January (B) sample would be in its start-up phase. Bi-weekly samples under this plan would have both first and second half sample members, from (A) and (B) portions;

there are just half as many completes in each sub-sample. This graph also shows the expected total daily number of completes based on this model.

Figure 7. Daily Completed Interviews (Proposed)



This process would take some extra time when it first starts, because only the January (A) sample would be finished on Day 30, not the entire month's sample. Also, extra time might be needed because the calling room staff will be managing two samples at any given time. The first set of biweekly estimates could be produced after 30 days, and data for subsequent estimates would be available every two weeks after that. This process would allow biweekly and monthly estimates to be produced; monthly estimates could be found by combining sub-samples. Estimates for January would be based on only January's (A) sample, but February's estimate would be based on the second half of January's (B) sample, February's entire (A) sample, and the first half of February's (B) sample. Since CDC weights the data using an automated procedure, calculating weights more often, for the biweekly samples, should not take too much extra time.

This sub-sample idea would result in a more constant number of completed interviews being made daily with only a few days in the middle of each month with fewer than 30 completes. There would be no time period of the month when hardly any completed interviews are being made. This sub-sample method would also allow good biweekly estimates to be obtained since the samples are randomly selected, unlike simply splitting the month into early and late respondents. The previous results about power and sample size requirements would apply similarly to biweekly estimates. The sample sizes required for desired power would need to be met for both samples, not just the total monthly sample. Although this process provides good statistical quality, managing two samples simultaneously presents some challenges operationally. Also, due to the change in timing of data collection, there may be an impact on the time references of certain questions.

3.2 Questionnaire Modifications

The final issue to be discussed is the design of the questionnaire and any possible modifications to the instrument. There are essentially two main parts of the BRFSS survey – 1) the core and 2) the optional modules and state-added questions. The core consists of 20 sections with questions related to general health which the states are required to run by CDC. The modules and state-added questions, however, are optional, and the states can include as many or as few as they choose. The data from these questions are important to the health programs that sponsor them because they use the data to help track the effectiveness of their programs and secure funding for them. The module and state-added questions are often a supplementary source of revenue for the state BRFSS programs because local agencies often pay to have their questions included.

If the instrument has to be changed in order to accommodate new questions for a public health emergency, either the core or the modules would need to be changed. Questions could be added to or substituted in either one, but some problems may emerge. If CDC decided to change the core by adding questions and increasing the length of the survey, the states may not be supportive of the longer interview length. Respondents might also get discouraged and not be as cooperative in completing the entire survey, so response rates could also drop. On the other hand, the respondents may be agreeable if they think valuable information is being collected. Another possibility would be to permanently reduce the number of questions on the core, so there is always room to add a few questions in the case of an emergency.

If CDC chose to substitute questions within the core it would have to be careful about introducing any contextual effects (Sudman and Bradburn 1982). These are the effects of the question placement within the questionnaire, which can influence the outcome of the survey. Sometimes asking a certain question can influence and change how the respondent answers subsequent questions. If CDC removed some questions and added the new questions at the end of the core, this would avoid impacting any of the previous questions. CDC would need to develop a plan for instrument modification before any emergencies occur, so the new questions could be added without delay. They could rotate the questions they drop from year to year, so no set of questions is skipped two years in a row. Keeping the number of questions the same would not place any extra pressure on the states.

Changes to the modules could also be made. If CDC adds a module for the specific event, and requires all the states to administer it, they would have to provide supplemental funding. The states may be hesitant to drop current modules and added questions from their survey because the health programs rely on their data. Some states may welcome the extra business, but others may feel they cannot fit in any more modules. For some states, CDC adding a module could be seen as an addition that would not make their survey time too long, but in other states they may need to substitute the module for one they are currently running. In this case, the states that add the module would be generating more revenue than usual, and the states that substitute might be generating less if CDC is paying them less than a supporting agency. If CDC reimbursed the states for their expenses and any loses they may sustain, the states may be cooperative.

Currently, for influenza vaccination coverage estimates, the survey question asks if the respondent has been vaccinated in the last 12 months; this may include people vaccinated during the previous season in addition to the current season. In this case, the existing questions could be replaced and asked in a way that allowed cumulative rates to be found.

4. Conclusion

The BRFSS design is capable of making adaptations to meet rapid response needs, and based on research from MMWR, discussions with public health state officials and BRFSS personnel, and research done at the Survey Research Unit, it appears that the BRFSS should be used for these situations because of its design scope and context. However, if with-in state comparisons are needed, the current state sample sizes will not provide adequate power to test for differences in monthly vaccination coverage rates between states and their regions. In order to test these differences with adequate power, the BRFSS sample sizes would need to be increased. Even once they have been increased a reasonable amount, it would still be hard to detect relative differences of 20% and sometimes 35%, especially in the early months.

When biweekly estimates are of interest, a change in the monthly calling cycle is needed. One idea is to take the monthly sample sent from CDC and randomly split it into two sub-samples. If the first monthly half-sample is put into calling on the first day of the month, the second added on the 16th day of the month, and each is allowed to run for its typical month of calling, reasonable biweekly estimates could be produced from each monthly half-sample. By combining two biweekly samples into one large sample, monthly estimates can still be calculated. Since these biweekly estimates would be needed in response to a current, and possibly new, public health emergency, an efficient way to fold in the rapid response questions must be devised. New questions would need to be added or substituted into either the core or module sections of the BRFSS.

The possible changes suggested in this paper would preserve the statistical quality of the BRFSS, and their effects on practicality have been taken into consideration. The BRFSS can be made more useful in the ways suggested, but there is a need for additional, and perhaps substantial, resources. This need must be addressed in addition to the increased complexities of integrating a rapid response component into the BRFSS design, but it certainly seems achievable.

The theory developed in this paper examines the differences in populations where there is complete overlap, and a generalization to this can be seen in Deal (2007). For the general case of partial overlap, further research using this theory and applying it to a specific scenario, similar to the vaccine coverage shortage situation, can be done. Also, investigating how the proportion of overlap affects the power and sample size relationships would be an interesting topic.

Acknowledgements

The authors would like to acknowledge with thanks the support of CDC and RTI International (under subcontract #08654-016, Task E from RTI International) and the North Carolina State Center for Health Statistics.

References

Biemer, P.P., & Link, M.W. (2007). Evaluating and modeling early cooperator effects in RDD surveys. In Lepkowski, J.M., N.C. Tucker, J.M. Brick, E.D. DeLeeuw, L. Japec, et al. (Eds.), *Advances in telephone survey methodology*. Hoboken, NJ: John Wiley & Sons, Inc..

Brookmeyer, R. and D. Stroup (2004). "Monitoring the Health of Populations, Statistical Principles and Methods for Public Health Surveillance." New York, Oxford University Press.

CDC (2002). "Rapid Assessment of Injuries Among Survivors of the Terrorist Attack on the World Trade Center - New York City, September 2001." Morbidity and Mortality Weekly Report **51**: 1-5.

CDC (2005). Behavioural Risk Factor Surveillance System: Operational and User's Guide, Version 3.0. **2005**.

CDC (2005). "Estimated Influenza Vaccination Coverage Among Adults and Children - United States, September 2004 - January 31, 2005." Morbidity and Mortality Weekly Report **54**: 304-307.

Deal, Allison (2007). "Some Statistical Issues in Adapting Sample Designs for Rapid Response Surveillance when Comparing Overlapping Samples." Unpublished master's paper, Biostatistics Department, University of North Carolina at Chapel Hill.

Johnson, N. L. and S. Kotz (1970). "Continuous Univariate Distributions -1." Boston, Houghton Mifflin Company.

Kalsbeek, W. D. (2005). Complete as Part of Task E (Design of Protocol for Rapid Response Surveillance); RTI International Task Order Contract 08654-016 (Behavioral Risk Factor Surveillance System) with the Centers for Disease Control and Prevention.

Sudman, S. and N. Bradburn (1982). "Asking Questions: A Practical Guide to Questionnaire Design." San Francisco, Jossey-Bass.