MANAGING 78 SIMULTANEOUS RDD SAMPLES

Paul Buckley[1], J. Michael Dennis[1], Candice Saulsberry[1], Victor G. Coronado[2], Trena Ezzati-Rice[2], Edmond Maes[2], Ann-Sofi Rodén[1], and Robert A. Wright[2]
(1) Abt Associates Inc., (2) Centers for Disease Control
J. Michael Dennis, Abt Associates Inc., 640 N. LaSalle, Suite 400, Chicago, IL 60610

Key Words: RDD, Sample Management, National Immunization Survey

Introduction For the National Immunization Survey (NIS), Abt Associates Inc. has developed an array of systems for managing 78 random digit-dial (RDD) samples simultaneously. The NIS sample management system assures the even distribution of data collection for each of the 78 RDD samples and the completion of the appropriate number of cases per quarter. The system is based on the daily application of rigorous statistical techniques using historical and same-quarter information about each IAP. In addition, the system has automated reporting, allowing NIS managers to identify any problems at an early stage.

This paper describes some of the key aspects of the NIS sample management system and how it is used to inform decision-making regarding sample preparation, case release, and interviewer staffing.

Sponsored by the National Immunization Program and the National Center for Health Statistics (both of the Centers for Disease Control and Prevention), the NIS provides rolling four-quarter estimates of vaccination coverage for each of the 78 Immunization Action Plan (IAP) areas established by the National Immunization Program. The 78 IAP areas are geographical areas defined by the 50 states and 28 selected metropolitan areas (Ezzati-Rice et al., 1995). The sample for each of the 78 IAP areas requires separate monitoring since each IAP displays unique characteristics in terms of household eligibility and response rates.

The NIS is both a list-assisted random-digit dial (RDD) telephone survey of households having age-eligible children as well as a provider record check study that validates self-report data on children for whom consent has been acquired during the RDD interview. Households having children between the ages of 19 and 35 months are eligible. Data collection has been continuous since 1994.

The sample management system used on the NIS reflects the requirement to monitor independently each of the 78 IAP samples. In addition, the system reflects the large scale of the RDD study and the fixed quarter-based schedule for providing vaccination coverage estimates. The scale of the study in large measure is the result of the relatively low eligibility rate (less than 5% of screened households have age-eligible children). To complete approximately 35,000 interviews a year, the NIS places about 8 million calls to 1.8 million phone lines, screening approximately 800,000 households. These efforts require the preparation of more than 500,000 sample telephone lines per quarter. To conduct the work, a very large interviewing staff and numerous telephone supervisors are dedicated to RDD telephone production. Managing case release and the allocation of human resources to produce interview data on a quarterly basis is coordinated through the NIS sample management system.

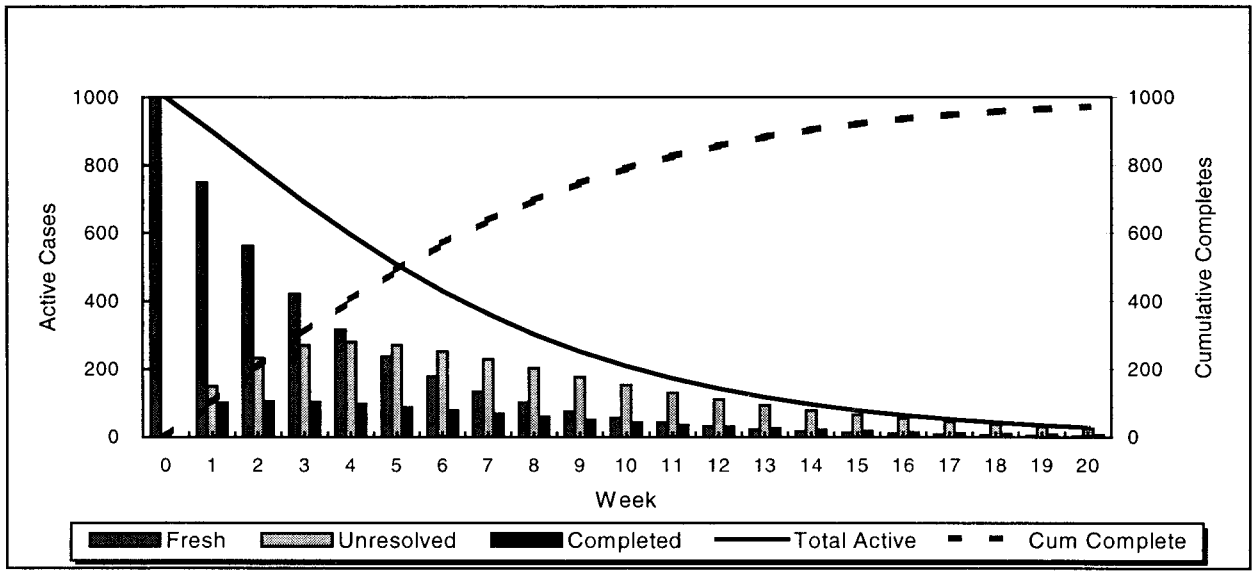
General RDD Sample Management Considerations

Managing an RDD survey involves recognition that the work in the telephone center goes through three distinct phases. In the initial phase of data collection, there is a high proportion of fresh cases (i.e., cases that have not been dialed). Relatively few cases require interviewer specialists such as refusal conversion experts. In contrast, towards the end of data collection, there is high proportion of unresolved cases—partially completed interviews, refusals and hidden refusals, appointments, answering machines, and non-contacts. There are no fresh cases at this point. Therefore, much of the interviewing work at this point requires interviewer specialists such as refusal conversion experts. The duration of the initial and ending phases is variable, but can be controlled to a degree by altering the number and type of interviewers relative to the number of active cases.

The middle phase of data collection involves the main sequence for RDD work. At any point in time, there is a mix of fresh cases, those that have been fielded but not yet contacted (i.e., "non-contacts"), and some unresolved cases that have been contacted.

A simple sample management plan would specify the release of all fresh cases during the initial field period. As a result, the mix of fresh, non-contact, and other unresolved cases would change continuously, with the number of fresh cases constantly declining and the number of unresolved cases first increasing and then declining. Even with just three categories of cases

Figure 1. Case Mix Obtained by Releasing All Cases at the Start

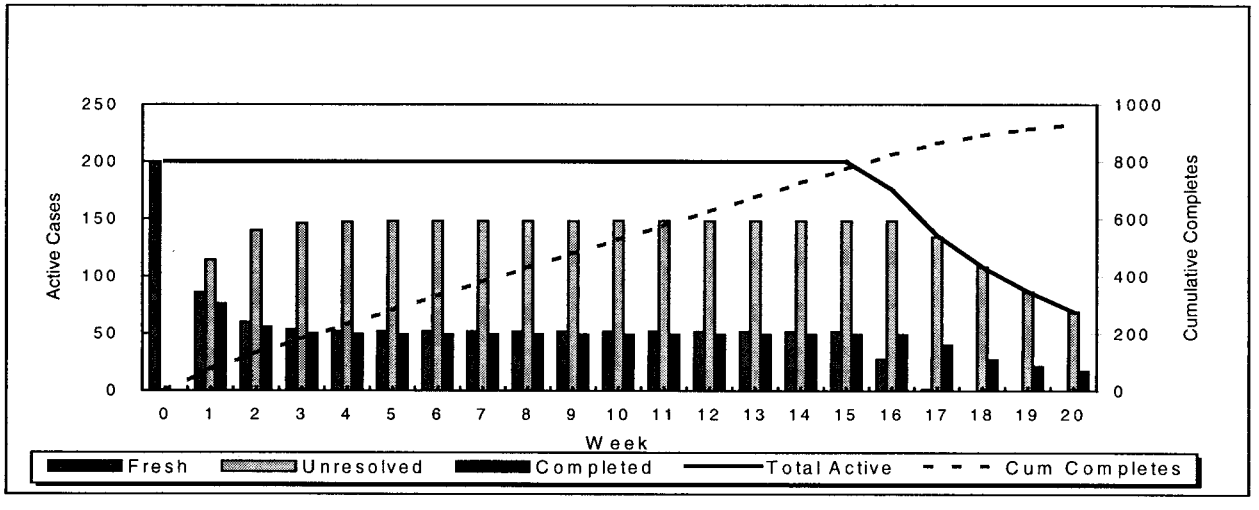


(fresh, resolved, and completed), Figure 1 illustrates the continuous changes in the sample queues that this approach generates. As the proportion of each type of case changes at its own unique rate throughout the field period, the number of interviewers, language specialists, and refusal converters changes as well.

As shown in Figure 2, the preferable alternative to this simple management plan is to maintain the total number of cases and the proportions of each type of case constant over time. This can be accomplished by releasing fresh cases into the telephone center on a continuous basis with the number of new cases determined by the number of cases that are completed or otherwise finalized. This sample management plan leads to an initial start-up phase with all fresh cases, a longer middle period characterized by constant proportions of fresh, unresolved, and completed cases, and a close-down period dedicated to finalizing unresolved cases (fresh, resolved, and completed).

Balance in the mix of cases also assures that each case is worked each day, preventing cases from becoming "stale." When too many unresolved cases are active, there is risk that some cases will not be dialed on a given day. As a result, irregular holes in the calling patterns may create "lumpiness" that could increase average latencies from case release to case resolution as well as increase the average number of call attempts per case.

Figure 2. Case Mix Obtained by Release of Fresh Cases Over Time



An additional consideration is that when there are too many active cases, the likelihood of a missed appointment increases because an interviewer may not be available at the appointed time—all interviewers may be working less critical cases. When an appointment is missed, the probability of resolving a case diminishes substantially. Respondents may use the missed appointment as a justification for refusing to participate.

Because the NIS requires four complete surveys per year, these predictable patterns displayed in Figure 2 are especially important. By maintaining essentially a flat level of production over most of the quarter, proper sample management allows for projections of the number of completed interviews required per day in order to finish each quarter on schedule. Moreover, such sample management allows the NIS to take advantage of overlaps between quarters—one quarter can be in the close-down phase while the next one is starting. By overlapping a known and controlled start-up phase for one quarter while the previous quarter is in its final close-down phase, the NIS maintains a nearly flat (and highly efficient) staffing plan throughout the year.

These considerations underlie the NIS sample management plan, enabling a rational allocation of interviewing labor and the completion of work across all IAP areas in the same time frame.

Sample Management Database The NIS sample management system is based on cumulative project data and daily updating of the sample management database in a given quarter. The sample management database maintains statistics on the status of data collection for each IAP area as well as historical information from all prior quarters used for estimating eligibility and response rates. The database contains, among other statistics, numbers of completed interviews, numbers of active cases, historical and current eligibility rates, and estimates of completed interviews to be garnered from the active cases that are released but not yet finalized in the system. The sample management database is linked to daily production tables so that decisions regarding case release are based on the most current information. Fresh sample is organized into discrete, replicate-IAP combinations, the release of which is recorded in a case-release matrix. In a given quarter, over 1,400 replicate-IAP combinations are tracked. Another source of information for sample management is the queue report. The report shows the number of cases in each CATI queue (e.g., fresh, non-contacts, refusals, appointments, etc.) at the start of each day or at any point during the day. This information shapes decisions on the quantity of fresh sample to release. After a case is dialed, the call attempt's event information is stored and processed in sample and production databases, which then update the sample management database and refresh the queue-building report.

The sample management database is used to address variability between and within IAP areas. As mentioned earlier, the NIS is 78 simultaneous RDD surveys, with independent samples drawn for each of the 78 Immunization Action Planning areas. Variability due to IAP-specific characteristics can complicate the goal of collecting data at the same rate in each IAP—from the first to the last completed case, the interview completion rate should be comparable across all IAP areas.

To address between-IAP variability, the sample management database is used to craft IAP-specific replicates. Over several quarters, for example, the database indicated that the eligibility rate in Utah is substantially higher than the rate in the District of Columbia. The sample management database also takes into account historical variations in response rates across IAP areas. Based on these known differences, variably sized replicates for each IAP are designed so that each will yield the same number of completed interviews. In principle, each IAP needs exactly the same number of these properly specified replicates to obtain the quarter's completion goal.

The sample management database also allows the NIS to address within-IAP variability. At issue is that the rate of completion can vary significantly across replicates drawn for a single IAP. For example, one replicate sampled in Hawaii contained a retirement community, producing a completion rate considerably lower than predicted. The sample management database contains monitoring tools that all NIS managers use to dampen the effects of such intra-IAP variability. The database provides information needed to determine on a daily basis the release of cases to the telephone center by replicate within each IAP and to view on a daily basis the effects of that release.

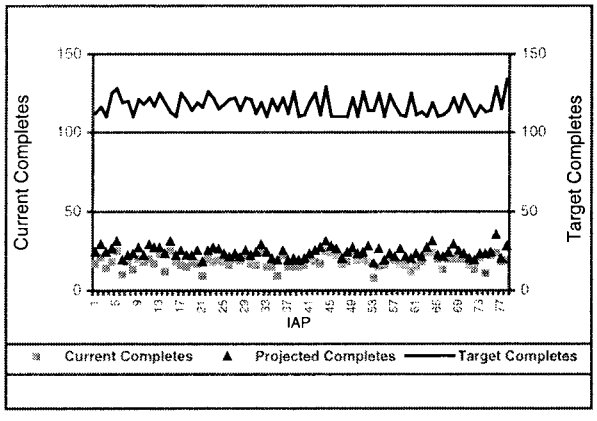
The sample management database is used to develop and refine a set of formulas for projecting the final number of completed interviews for each IAP on a daily basis. This program estimates the number of interviews that would be completed if no additional sample were to be released and from this the total number of fresh cases that are needed to meet the target number of completes for each IAP in the current quarter.

These predicted values are computed using the percent of released telephone numbers in each IAP that yield completed interviews, the percent of released cases that will never be resolved, the current number of completed interviews, telephone numbers released, and currently active cases in each of several categories. These predictions are reviewed both in the aggregate and on the IAP-level. Using the aggregate values, the system estimates how many additional fresh cases should be released to the telephone center each day. By examining the individual IAP areas, the system identifies the IAP areas for which fresh sample should be released.

Examples of Sample Management Tools Some of the sample management tools supported by the database are described below.

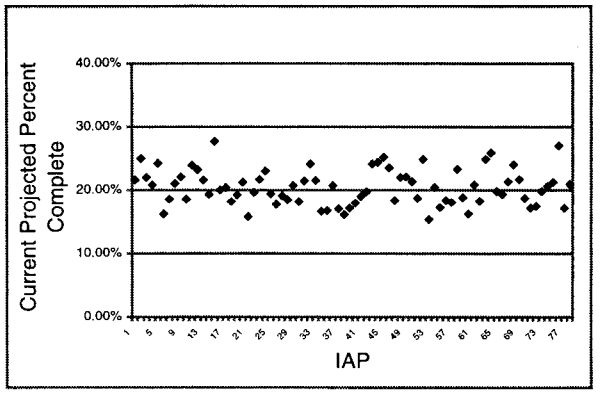
Figure 3 shows our overall completion targets for each IAP, the current number of completed interviews for each IAP (boxes in graph), and the projected number of completed interviews (triangles) to be obtained from the currently active telephone numbers. Since the data are displayed from early in a quarter, the gap between what is targeted and what is completed is large. The aim is to manage sample release such that the number of current completes is proportional to the target number and the gap between current and projected completes is relatively uniform across IAP areas.

Figure 3: Current, Projected, and Targeted Number of Completed Interviews



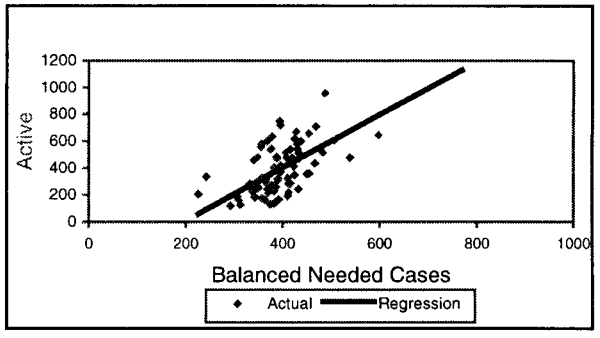
From the values underlying Figure 3, the projected percent of completed interviews is calculated (i.e., the projected number of completed interviews divided by the target number of completed interviews). The results are displayed in Figure 4. These data are used to decide where to release fresh cases—those IAP areas with the lowest projected percent complete on a given day are the ones most in need of additional fresh cases.

Figure 4: Projected Percent of Target Obtained, by IAP



A second monitoring tool highlights IAP areas by differential need for case release. As mentioned above, the production characteristics of each IAP and even IAP-replicate intersections are unique. Their varying characteristics influence the number of completed interviews per replicate. Deciding how many cases are needed for case release, therefore, requires that we compute how many cases will be needed to finish all the work in each IAP and then compare this figure to how many are currently active. An example of this exercise is displayed in Figure 5. IAP areas above the diagonal need more cases than are currently active, while those below do not. The further to the right of the diagonal an IAP is, the more total cases an IAP requires to complete the quarter.

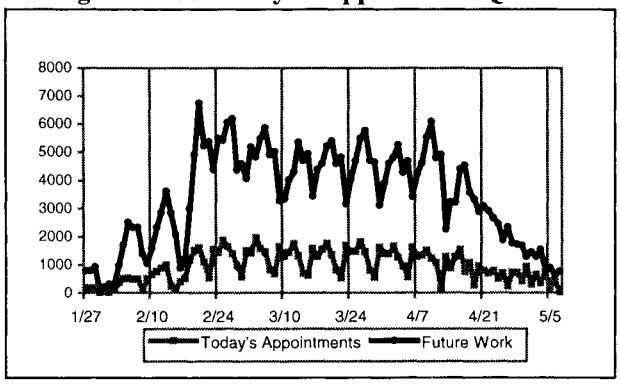
Figure 5: Number of Active and Needed Cases for Each IAP



Other monitoring tools examine specific queues of CATI cases. Figure 6 shows the day-by-day values for a recent quarter of two critical queues—the current appointments and the amount of future work (which includes both appointments after the current day and initial refusal cases that are on hold). Sudden changes in values indicate that the system could be out of balance. Intra-week fluctuations are expected as a result of specifications in the CATI scheduler,

respondent availability, and predictable intra-week variations in the number of interviewers.

Figure 6: Summary of Appointment Queues



Finally, two other tools demonstrate whether the system is in control in terms of maintaining a constant rate of production and balance across all IAP areas. Figure 7 displays the cumulative number of completed interviews for a recent quarter, showing a fairly constant rate of production after the initial start-up period and before the closing weeks of the quarter.

Figure 7: Cumulative Completed Interviews

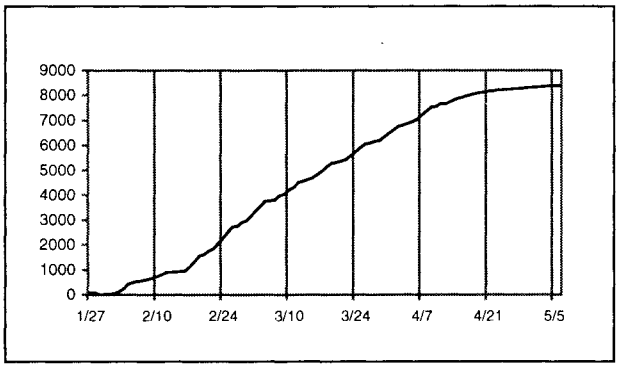


Figure 8 demonstrates an approximate balance in the number of completed interviews at the end of a recent quarter. Some of the inter-IAP variation reflects differences across IAP areas in the completion targets (which are set to yield a rolling four-quarter average of 110 completed interviews for each IAP). As can be seen in the two exhibits, the NIS sample management system produced consistent results.

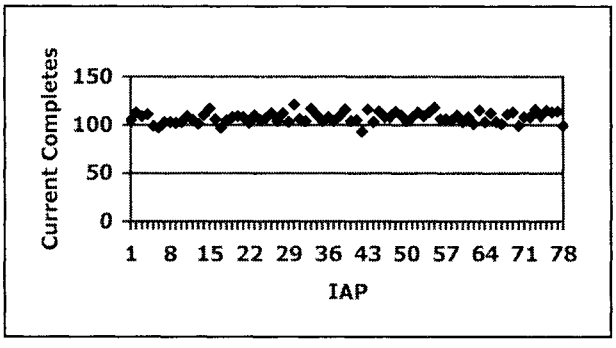
Conclusion The NIS sample management system was designed to manage 78 independent RDD samples simultaneously. The aims of the system include the completion of interviews across all IAP areas at an even rate (controlling for inter- and intra-IAP variation), adherence to a fixed quarterly schedule of data delivery, and the supply of a constant mix of cases in the telephone center. Achieving the latter aim allows the NIS to distribute the work evenly on a year-round basis to interviewing generalists and specialists and is crucial to maintaining a high level of production efficiency.

The sample management system contains a number of graphical tools for early detection of any problems. These tools are also used for making decisions on a daily basis regarding the quantity of cases to release to CATI and the specific IAP-replicate intersections to be used for daily case release. Mid-quarter corrections are sometimes made in the case release ratios on the basis of these tools in order to preserve an appropriate mix of cases in the CATI system.

While the NIS sample management system is used to manage 78 identical surveys, its functionality has general applicability for managing different RDD surveys at the same time. The system—and the rigorous statistical methods on which it is based—could be used for managing all sample in a telephone center and has applicability for the management of field surveys as well.

With each succeeding the quarter, the intelligence of the sample management system increases because the underlying database is updated with IAP-specific information that enhances the forecasting power of the predictive formulas embedded in the system. This provides the basis for more precise specifications for sample generation as well as management of case release to CATI.

Figure 8: Number of Current Completes by IAP at End of Quarter



References

Ezzati-Rice, T.M., Zell, E.R., Battaglia, M.P., Ching, P.L.Y.H., and Wright, R.A. 1995. "The Design of the National Immunization Survey." *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 668-672.