Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 29–48, DOI: 10.2478/jos-2013-0003

The General Survey System Initiative at RTI International: An Integrated System for the Collection and Management of Survey Data

Lisa Thalji[1], *Craig A. Hill*[2], *Susan Mitchell*[2], *R. Suresh*[2], *Howard Speizer*[2], *and Daniel Pratt*[2]

There is a risk of introducing survey error at every stage of any study involving a survey: design, data collection, processing, and analysis. Effectively managing the survey sample from the time of sample selection through the survey lifecycle is essential to producing high-quality data on schedule and within budget. Managing the survey lifecycle using software systems that are not fully integrated can result in error and cost inefficiencies. The development of an integrated data collection and management system that supports monitoring of survey error has the potential to reduce errors and improve operational efficiencies. This system, referred to as Nirvana, uses a standardized database, protocol, and terminology. It integrates case status and history information across modes of data collection and tracing as well as sample and contact information. Nirvana also centralizes questionnaire development, quality monitoring, and support for real-time survey management decisions.

Key words: Total survey error; data collection systems; Adaptive Total Design; paradata; survey costs.

1. Introduction

Organizations that collect, process, and analyze survey data must design and implement systems and protocols that guard against the intrusion of survey error to produce data and statistics that are of sufficient quality. However, quality is a complex construct. By some definitions, quality data must be accurate and meet the user's needs or fitness for use (Juran and Gryna 1980). Biemer and Lyberg (2003) notes that when there are a variety of uses for survey data, fitness for use becomes multifaceted. He stipulates that survey data must be accurate, credible, comparable, useable, relevant, accessible, timely, complete, and coherent (Biemer 2010). In other work, this Total Survey Quality framework has been described as having three dimensions: accuracy, timeliness, and accessibility (Biemer and Lyberg 2003). In this article, we describe RTI International's (RTI) initiative to build an integrated survey management system designed to improve the accuracy, timeliness, and accessibility of the data we produce.

RTI is a nonprofit research organization that primarily conducts research for the U.S. federal government. As a research contractor, RTI responds to requests for proposals from

[1] RTI International 3040 East Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709-2194, U.S.A. Email: thalji@rti.org
[2] RTI International, 3040 East Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709-2194, U.S.A.

U.S. statistical agencies as well as other nonstatistical agencies. Research proposals are requested in response to policy or program questions (e.g., understanding incidence and prevalence rates for drug use and abuse) and the scope of research services requested varies widely. Some agencies request study designs and plans for data collection, data processing, and statistical analyses; others require only data collection and data processing proposals. RTI conducts studies using all types of survey modes, subject populations, and establishments. In all proposals, RTI describes how it will collect the relevant data and, most importantly, how it plans to ensure the quality of those data given schedule and budget constraints.

There is potential for introducing survey error at each stage of a survey – from sampling frame creation to data dissemination (United Nations 1999). Sampling error arises from observing a sample rather than the whole population, and further errors can be introduced through deficiencies in the sample frame or sample design. Nonsampling error can be introduced by questionnaire authors, interviewers, and/or respondents (measurement error); errors result from bias in the collected data (nonresponse error); and errors can occur in the coding, editing, and management of collected survey data (processing error). The accumulation of error across all survey stages is referred to as *total survey error* (Biemer and Lyberg 2003). Any error at any stage diminishes the accuracy and precision of survey estimates.

One common approach to minimizing survey error is to use, for each stage of a survey, a software-based solution designed to double-check, prevent, or alert survey managers to the potential for error. In the data collection process, for example, survey software can automatically disallow entry of out-of-range values, check for (and disallow) entries that are inconsistent with prior answers, and can keep interviewers from asking questions that are inappropriate for certain subsets of respondents (such as asking male respondents if they are pregnant). Another common approach is to use software to collect and analyze paradata (i.e., data about the data) during the data collection process that allows survey managers to adjust resources and approaches in near-real time to gain improvements in survey quality metrics, such as response rates, while balancing time and cost constraints. Paradata can range from simple summary measures such as response rates to something as specific as the elapsed time between call attempts.

At RTI, we call the process of using the paradata to make management choices that improve the quality of the survey data Adaptive Total Design (ATD). Evolving from the concept of responsive design (Groves and Heeringa 2006), ATD is a framework for monitoring key statistics that help survey managers decide how best, in real time, to modify protocols to minimize total survey error and maximize cost efficiency (Hunter et al. 2012). RTI's ATD system requires access to all of the data collected during survey administration and complex models for measuring the impact of these protocol changes on survey estimates.

Before this initiative, RTI built project-specific and independent software, systems, and protocols designed to prevent and minimize error. This process resulted in separate software-based systems for managing sampling, data collection, coding, editing, and data analysis processes. The information contained in these systems was not integrated (i.e., the systems did not talk to one another) within a given survey or across survey modes.

This lack of integration created the potential for undetected measurement and data processing error.

In addition, as is the case at many large survey research organizations, many of the survey management systems were developed for particular projects, which meant that RTI often had multiple software systems to support the data collection stages of the survey process. Since these systems were built independently of one another, standardization was inherently lacking, which also could create the potential for specification error. Finally, because many of these systems were built some time ago, it was proving increasingly difficult to generate the kind of near-real time paradata needed to fully use our ATD system.

In the data collection and data processing stage especially, we saw a need to create an integrated and standard system that ties together all stages of data collection and data processing by employing a common database structure, protocols and terminology across studies. Consequently, we developed an RTI-funded initiative, the General Survey System Initiative (GSSI). The GSSI's main goals are to improve the quality of the data we produce through means not achieved heretofore with our custom approach: (1) integration, (2) standardization, and (3) accessibility to necessary paradata.

The GSSI is governed by a group of senior managers in RTI's survey and programming groups. The senior managers are responsible for carrying out various aspects of the initiative, including technical design, budget and schedule.

2. Process for a New Approach

Our focus on customization had introduced several operational challenges. Although study staff were able to develop workarounds, concern grew that the lack of standardization and integration and the misallocation of resources could be causing unintended and undetected survey error. For example, if the tracing system and computer-assisted telephone interviewing (CATI) system are not integrated, a separate process is required to share information about the sample member between the two systems. The time delay to associate the sample member's information from two different systems impacts the survey manager's abilities to take immediate action and array resources appropriately. It also may result in some duplicative systems and/or manual operations in an attempt to improve the timeliness of information and decision making, thereby affecting overall operational efficiencies and costs. These types of challenges and inefficiencies made it increasingly difficult to maximize data quality within reasonable costs (i.e., within the budget provided by the government) as our survey efforts grew more and more complex.

To understand the challenges of working with RTI's systems, the GSSI steering committee decided to gather information by surveying RTI project leaders, survey managers, and programmers. A web survey was developed to (1) inventory all of the software and related tools used to support surveys; (2) evaluate the efficiency of deploying these tools; and (3) evaluate the overall usability of software and related tools. These areas typically contribute 20–30% of the total survey costs. Over the span of several months, the data were mined for common themes and challenges. The committee methodically reviewed and discussed these data to ensure a shared understanding of each problem before proposing solutions. What emerged through these discussions was a clearer understanding that the root cause of our problems could be attributed to the lack of standardization, integration of survey systems, and access to survey data and paradata (Couper 1998).

Another consideration that factored into our discussions was the need to better position ourselves to address the changes and challenges in the field of survey research. Response rates have been declining (see, e.g., Curtin et al. 2005; de Heer 1999; Steeh et al. 2001; Tortora 2004) across all modes of data collection: in-person, telephone, and mail surveys. To counter this trend, the conduct of multimode surveys has increased as a means of obtaining higher response rates than those using one mode alone. Often, telephone surveys are combined with an in-person follow-up. In other instances, in-person surveys are combined with shorter telephone surveys and/or mail surveys to bolster the response rates. To continue to provide timely quality data in a fast-changing survey environment, we needed robust systems that could operate under various conditions and provide the type of paradata required to efficiently manage survey operations.

Table 1 summarizes our root cause analysis and serves as the basis for developing a new approach to address these objectives. These objectives are guided by our total quality management framework and desire to improve the accuracy of the data we collect by reducing and controlling nonresponse error, measurement error, and data processing error.

The next step in our process was to convene a multidisciplinary task force to explore the feasibility of building or buying a single, unified data collection system to manage surveys across data collection modes and activities. The task force developed a list of general and function-specific requirements to guide the evaluation. General requirements included the following:

- Integrate case management systems across all modes of data collection and tracing.
- Integrate software systems for questionnaire development.
- Integrate data quality metrics.
- Store survey and paradata for all data collection modes and tracing in a single database.
- Standardize event and status codes to conform to survey industry practices for reporting and tracking survey progress.
- Allow for easy configuration to support a wide range of studies.
- Provide for easy-to-use tools to access paradata for survey management.

The task force evaluated several commercial off-the-shelf (COTS) data collection software systems. Where possible, questionnaires were programmed and case management systems were tested in the COTS software. The questionnaire programming systems met our requirements, but the case management systems did not. Given the level of complexity and details required to effectively manage the types of large-scale surveys RTI conducts for the U.S. federal government, a robust and refinable case management system was deemed the most important criterion. Thus, after careful consideration of these COTS data collection software systems, RTI decided to invest overhead funding to develop its own integrated data collection system to maximize design control over the integration of the case management systems. In making this decision, we knew that a commitment to such an endeavor was a long-term investment that would involve staff participation across disciplines for a two- to three-year period, followed by ongoing enhancements.

Table 1. Objectives for new approach

Objectives	Challenges	Desired outcomes
Standardization	Custom approach causes inefficiencies Maintenance of multiple systems raises costs Diversity of systems requires additional training and knowledge transfer	Eliminate duplicative systems Improve operational efficiencies Decrease learning curve for users and programmers Reduce level-of-effort for system specification, setup, and support Provide more flexible staffing options Leverage reusable components of the system Increase consistency of implementation
Integration	Difficulties in identifying next actions for particular cases result in duplicative work Delays in sharing information across systems reduce efficiency and increase cost	Reduce elapsed case processing time Eliminate case processing steps Improve sample management information Eliminate redundant data Reduce reliance on programmer intervention to move information between systems
Access to data (survey and paradata)	Lack of real-time access to data increases difficulty and cost of survey management Changing priorities quickly during data collection is difficult without access to flexible, dynamic data reporting tools	Increase availability of data to survey managers Improve quality of decisions Improve timeliness of decisions Develop better and more consistent data access and reporting tools

Once this decision was made, the GSSI steering committee convened another task force whose mission was to develop the ideal structure, flow, and functionality of a unified data collection system that would meet our objectives of standardization, integration of survey systems, and access to survey data and paradata. Their initial focus was limited to the data collection process, thereby addressing the most immediate costs and quality concerns and challenges of operating in our current environment. On large-scale surveys, data collection costs can be as high as 50–60% of the total costs of the survey, and can contribute to errors in the accuracy of the data collected. The design for the unified data collection system resulted in the development of the Nirvana system.

3. Design and Architecture of Nirvana

The integrated data collection system – Nirvana – is designed around a central database and repository. Data are stored in the standardized database in a table structure and are accessible by all of the component systems that support case flow as the sample moves through different stages of activities. Each project using Nirvana has its own database that holds all data, including:

- sample data, which includes strata used for sample selection and other characteristics of the population of interest.
- subject contact data, which includes contact information as well as past activities specific to each sample member.
- metadata that specifies details about the questionnaire instrument, such as question text, response options, and other attributes of the questions, as well as the configuration settings for the stages in the data collection process, and is comparable to the structural metadata used in Statistical Data and Metadata Exchange (SDMX).
- paradata which includes a wide array of operational and observational data about data collection, such as cost and production data, record of calls data, information about the sample frame and sample, data quality measures, interviewer characteristics, and interviewer observations, and
- survey data regardless of data collection mode.

The Nirvana database has a core SQL schema that holds all of the sample and subject contact data. All other modules tap these centralized schema tables to provide the necessary data for contacting sample members by telephone, by mail, or in person. When new contact information becomes available, these same tables are updated, regardless of the source of the information; thus, all of the modules receive the updated information at the same time.

Another key feature of the Nirvana database is that it maintains a single database table that holds the paradata, such as the history of all of the efforts expended in contacting the sample member. Thus, whether an invitation is sent via e-mail, telephone, or in-person visit to the sample member's home, these activities are recorded in the same table, and that information is available to all users. This feature ensures that the interviewer, supervisor, and survey manager are fully aware of all of the efforts made to reach the sample member.

Nirvana uses an administrative database that maintains linkage information in an encrypted form for all of the project-specific Nirvana databases. The administrative database also maintains the role and access control for each user for each project. When a Nirvana application or website is invoked, it queries the administrative database to ensure that the current user has access to the project in question and automatically connects to the desired project. This technique makes it easy for survey managers to access their projects and provides an extra level of data security for the projects. Nirvana design took into consideration resource allocation, especially developer resources, with the hope of reducing costs. For example, even though the initial setup of the database might require an experienced SQL programmer, their involvement will be minimal after the initial database setup phase. The configuration of the project and other setup work can be done by a programmer with less experience and hence at lower cost.

The database is set up in such a way that each survey process is defined as a *stage*; for example, a lead letter mailing; a telephone, field, or web-based interview; or the tracing or locating of sample members are each defined as a stage. The current state of each stage is reflected by the current *status* code for that stage. For any given case, a specific status code initializes the stage. The user posts an *event* to indicate that a specific activity had occurred, and the system automatically derives the new status based on the posted event, the prior status, and other indicators. Note that it is possible for a status code in one stage to trigger the initialization status code in another stage, which is a simple but powerful technique that is employed to seamlessly link different modes of data collection. The Nirvana events, status, and stages essentially codify the steps and process in the workflow of a project and to some extent automatically document the same.

For example, for an in-person study that requires confirming telephone and address information in an automated fashion as its first step, the survey sample is loaded into Nirvana tracing stage number 400 with a status code of 1001 to indicate that the tracing stage has been initialized. When sample cases are sent to tracing (an activity), the status code within that stage changes to 1101 to indicate that sample cases have been sent to tracing. The results from tracing are reflected in the status code of that stage. When all pre-field tracing activities are completed, the sample case moves to the next stage in the survey process: the lead letter mailing stage. The completion status of pre-field activities triggers the initialization of the lead letter mailing stage. As letters are mailed on a flow basis, the status code changes to lead letter mailed. The lead letter mailed status automatically triggers the initialization of the next stage, field interview, and so on throughout the data collection process. An advantage of this scheme is that at any time the user can easily review the history of a sample case and know which activities have been conducted to date. In addition, the user knows the total number of sample cases ever initiated at each stage. The above examples illustrate the adaptability of the Nirvana system to changes in workflow processes required by any survey project. A data collection or monitoring process can be described as a combination of actions (Nirvana events) and resulting steps (Nirvana status); this implies that the process can be defined as a Nirvana stage. In addition, if a sequence of tasks can be automated, a Nirvana module can be developed to standardize the implementation of those tasks for that project and then the module becomes re-usable for future projects.

Figure 1 depicts the data flow among the major components of Nirvana.

Case Management Systems. As an overarching system, Nirvana incorporates the previously separate case management systems for CATI, computer-assisted personal

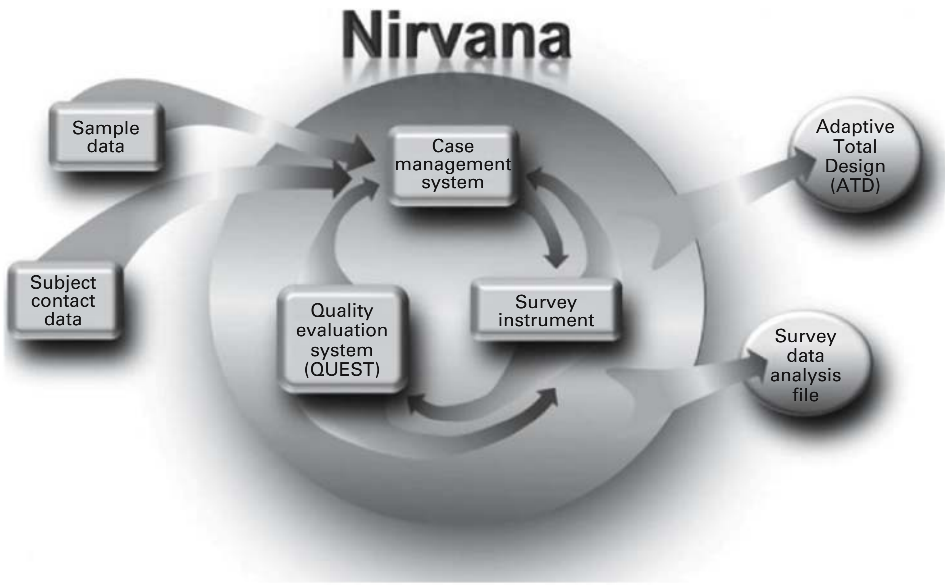

Fig. 1. Nirvana data flow

interviewing (CAPI), and tracing as well as modules that facilitate document mailing and receipt, e-mail invitations for web-based self-interviewing, biospecimen tracking, and so on. CATI and CAPI are for managing telephone and field data collection respectively; tracing is for identifying the best phone number or address for a sample member in a list survey or for panel maintenance. These case management systems and other modules use the same underlying tables to access the sample data and to record the paradata on the status of data collection. In some organizations, these systems use separate databases and processes that provide status information to a central database or reporting structure, but with Nirvana, the same database tracks the progress of data collection regardless of the mode, facilitating a seamless change in data collection mode. The integration of case management systems reduces programmer labor and minimizes the potential for processing errors. For example, if a sample member in a mixed-mode survey chooses to complete the case via the web rather than by CATI, the system will immediately detect that the survey was completed via the web so that no additional effort is expended in CATI for that sample member. Programmer involvement is not required to move completed surveys from one database to another, thereby decreasing the risk of processing errors.

Survey Instruments. The questionnaire development system is used to develop survey instruments for telephone, field, web, and hard copy paper data collection modes. This system allows for real time edit and consistency checks. Using the same system to author the instruments reduces training costs for the specification writers and programmers. This system provides special interfaces for translation specialists to develop instruments in other languages, including Spanish and Chinese. The system also facilitates recording of the questions for audio-computer-assisted self-interviewing instruments as well as mode-specific variations in question text and/or response options. The metadata for the survey instruments and survey data are stored in the Nirvana database. This facilitates the generation of codebooks to accompany analysis datasets as well as the generation of reports that can be stratified using key items in the questionnaire data.

Quality Evaluation System (QUEST). This component supports monitoring of interview quality for telephone or field data collection. QUEST addresses measurement error through its ability to listen to recorded excerpts of the questionnaire administration and provide immediate feedback to interviewers. With Nirvana, the QUEST tables are in the project database, and the results of the review can be easily combined with other information such as interviewer demographics, their caseload, and item nonresponse to provide a richer quality control report.

4. GSSI and ATD

Perhaps the most notable outgrowth of the GSSI is the enablement of ATD because this ties the initiative to the Total Quality Survey Framework that underpins RTI's approach to surveys. As described earlier, ATD is a survey management tool that monitors cost, productivity, and data quality on a real-time basis during data collection. It differs from electronic reports and quality checks because visual displays (i.e., charts, graphs, and dashboards) are used to exhibit quantitative information in multiple dimensions, including over time. ATD provides a snapshot of the progress and quality of a data collection that reveals deviations from expectations, prior observations, and protocols in a convenient and accessible way. It is designed to show and compare relationships and changes across data elements. While there are costs associated with its development and implementation, these costs can be offset by the reduced risk to data quality, and the ability to ward off problems that may be expensive to fix (for example, retrieving missing data or replacing interviews that were fabricated).

ATD is integral to the Total Quality Survey Framework because it exposes survey errors that are common across studies, as well as specific to a single study design. When survey errors are identified, survey managers intervene to reallocate resources or address problems before they become entrenched. For example, survey managers may shift interviewing resources from one sampling area to another to level response rates, retrain interviewers to remedy errors in interview administration, introduce changes to the data collection protocol (such as an alternate mode of data collection or an incentive) to increase response, or alter CATI programs to correct or prevent measurement or interviewer errors.

ATD development depended on an integrated system of data collection and management provided by the GSSI. First, paradata needed to be collected systematically and stored in a common format to enable adoption across a wide range of surveys; it also had to be linked to survey data. Data standardization and linkage were essential first steps. The ATD system was developed from the GSSI to be:

- the monitoring tool for total survey quality.
- the support for data-driven decision making.
- accessible by a wide range of project staff and clients, each with different informational needs.
- graphical in display but allowing ready access to underlying data and numeric reports.

- supportive of cross-survey methods research so that metadata could be created and analyzed to examine patterns across surveys.
- user-friendly and intuitive, giving end users the ability to create custom graphs and dashboards with minimal programming knowledge.
- accessible on a variety of platforms, including laptops, smartphones, and tablet computers, and
- affordable.

The ATD system arrays dashboards and charts under five topical headings: (1) production and cost data, (2) sample status, (3) record of calls data, (4) data quality, and (5) timing. Information about the interviewers – identification numbers, gender, age, years of experience, and education – is used to supplement and subset data in these categories, comparing, for example, the productivity of female interviewers with male interviewers. Survey managers have the ability to identify and investigate potential problems by examining related graphs and drilling down to the underlying data itself. Thus it is possible to view survey status and changes in both visual and numeric form. For example, Figure 2 shows the overall mean survey length by day and the interviewers who deviated by more than 1 standard deviation from the aggregate mean. The chart shows a fairly consistent length of about 60 minutes with convergence around the mean as the data collection period lengthens and interviewers gain experience. However, interviewer "Bill" exhibits a pattern of longer interviews, and interviewer "Cindy" exhibits a pattern of shorter interviews. These findings merit further investigation. In surveys with sensitive questions, for example, a longer administration time may mean that the interviewer is dwelling on questions that make the sample member uncomfortable, potentially leading to an interview break-off or item refusal. Survey managers have the ability to confirm findings by examining graphs showing the section timings of sensitive question groupings, as well as the underlying data showing the number of interviews Bill completed and his tenure as an interviewer. If the number of completed interviews is small (for example, one per shift) or Bill is inexperienced, the observed pattern may be anomalous and expected to change over time.

To illustrate other uses, charts can be displayed by sample geography so that resources can be redirected to areas where production is lagging. For example, Figure 3 displays

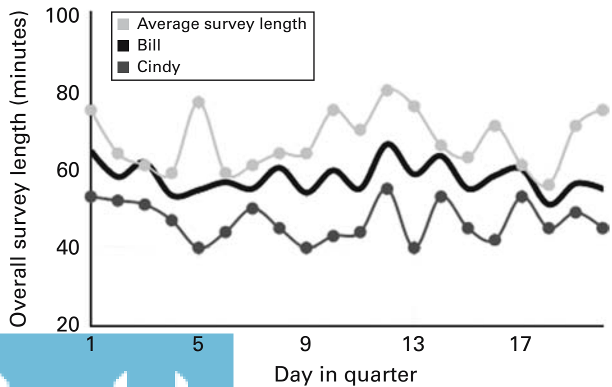

Fig. 2. Overall survey length and outlier interviewers

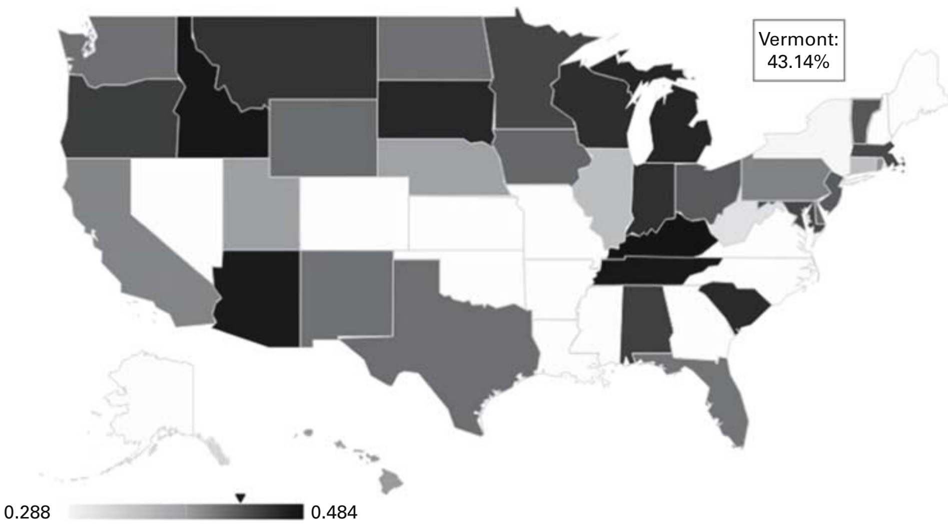

Fig. 3. Response rates by state

a map showing response rates by state for a national field study. Low-responding states can quickly be identified and investigated. Clusters of low-responding states can also be observed, indicating potential problems within regions that may be related to staffing or resources. By moving the cursor over the state, the actual response rate is displayed. As an enhancement, the percentage point improvement in response from the prior week or month can also be programmed to display.

Figure 4 is a chart showing the number of interviews completed by day compared with the 7-day moving average. The graph shows how production declines over the course of data collection, and how daily results differ from the moving average, indicating a more or

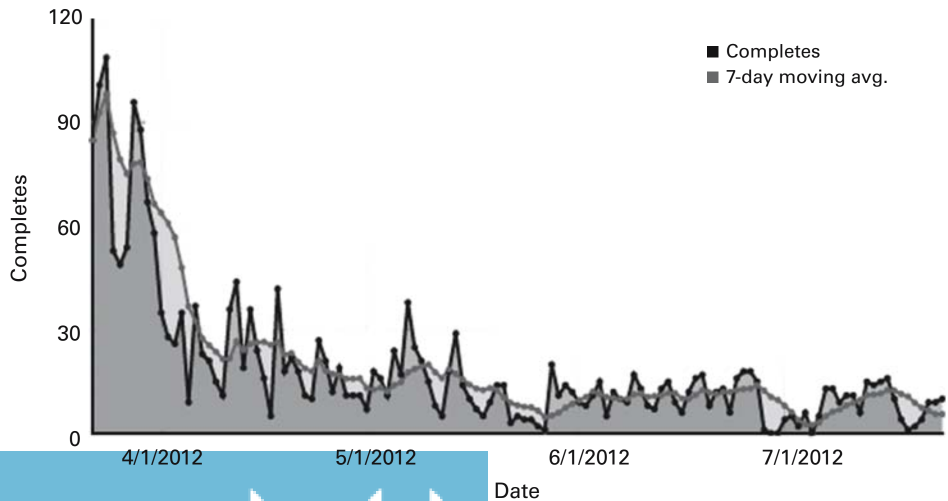

Fig. 4. Completed interviews by day compared with 7-day moving average

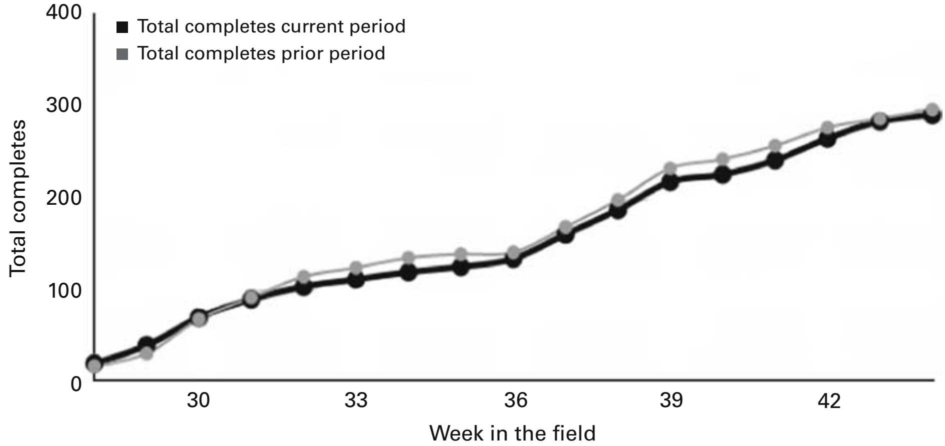

Fig. 5. Completed interviews by week compared with prior period

less productive day. Information in this chart can be used to model and predict final production, and to efficiently plan and staff future rounds of data collection.

In another view, the chart in Figure 5 shows how interview production in a current survey round compares to interview production in a previous round. In general, the chart shows that production increased at the same pace and rate in each round, which is evidence that field conditions remained stable and there were no external "shocks" to the environment that might impact data comparability. Expected production by week can be overlaid on the chart to compare planned and actual progress.

Because the charts incorporate the element of time, survey managers can immediately gauge the effect of changes to the data collection protocol or unanticipated field conditions. The chart in Figure 6 shows how the number of CATI interviews completed by

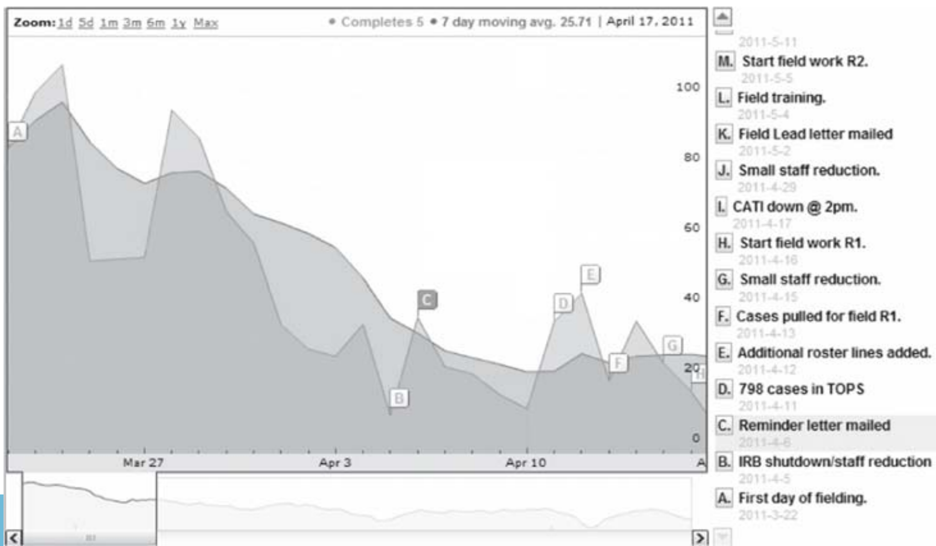

Fig. 6. Annotated timeline of production by day

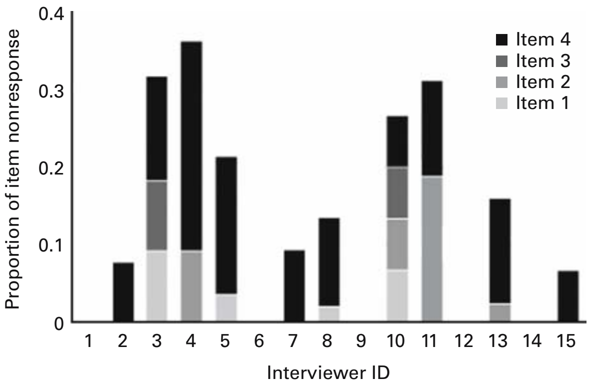

Fig. 7. Item nonresponse by interviewer

month tracked with interventions (reminder letters, incentives, sample releases) and external conditions (weather, early closings, system downtime). This chart provides a record of events that helps to explain the variation in production over time.

As an indicator of data quality, Figure 7 shows item nonresponse rates by interviewer for key items. It is possible to quickly discern which items are higher in nonresponse (e.g., Item 4) and which interviewers are experiencing item nonresponse at higher levels and across multiple items (e.g., Interviewers 3 and 10). This information can be used to coach interviewers to improve their performance. It can also be used to alert survey designers and analysts to items that may need imputation or re-evaluation for future use.

While cost and productivity measures are relatively limited and straightforward to chart and monitor, data quality measures can span the range of survey errors, including measurement errors, nonresponse errors, sampling errors, and processing errors. ATD has the potential to influence the occurrence and magnitude of all these error types, but is perhaps most practical for measurement error because it can be tied to paradata emanating from survey systems for reinterviewing, reviewing recorded interviews, and matching with administrative records – approaches commonly used for estimating measurement error. ATD is also well suited to identifying potential nonresponse error because it is capable of monitoring not only item nonresponse rates, but also survey nonresponse (weighted and unweighted) both overall and by key subgroups. ATD is best suited to applications where the underlying paradata exists in a standardized way in other survey systems, and when complex, study-specific programming (for example, the calculation of variances estimates) is not required.

To date, RTI has integrated ATD into a number of large-scale in-person and telephone surveys. In implementing ATD, the largest cost investment was in standardizing the data elements, and combining survey data and paradata in the same database, as described in Section 3. There were also costs associated with educating end users about ATD – the selection of relevant charts, a protocol for monitoring outcomes, and appropriate interventions to take when problems are discovered. For this reason, we found it cost-efficient to develop charts and graphs that could be used repeatedly over time or across multiple surveys. The creation of the graphs was straightforward, using commercially available charting software that did not require specialized training.

5. Implementation and Evaluation of Nirvana

Nirvana has been implemented with a phased approach. A "beta" or pilot version of Nirvana was implemented in the spring/summer of 2011 in the field test phase of a repeating cross-sectional education study: the field test of the National Postsecondary Student Aid Study of 2012 (NPSAS:12) conducted for the National Center for Education Statistics (NCES). The system was refined during this operational field test and rolled out on a second project in the summer/fall of 2011: the field test of the Baccalaureate and Beyond: 2008/12 Longitudinal Study, also conducted for the NCES. The ATD component was not available for this early evaluation period. Nirvana has been implemented in three full-scale studies in the first few months of 2012 with two more planned for later in 2012.

NPSAS:12, as a representative example of the projects handled by Nirvana, has multiple modes of data collection (web, telephone, in-person, paper) with many hierarchically linked survey participants (postsecondary institutions, students, and parents). The sample members are selected from school enrollment data and added to the survey on a flow basis. Survey data are collected over a nine-month period from over 100,000 students and their families and more than 1,500 postsecondary institutions. Data collection is monitored across all modes and by multiple subgroups. NPSAS:12 is used as the base year for a longitudinal study that follows first-time beginning students over time as they continue their education and work experiences.

The Nirvana system architecture, development environment, and software tools were finalized in the early stage of the NPSAS:12 field test implementation period. The critical component of this architecture was the database schema, including the metadata for all of the database variables. After finalizing the design and the database structure, the team produced a design document and scheduled a series of discussions with the system development team to form a common understanding of the technical design. The project manager established system development and change management protocols and assigned system modules for development to team members. The GSSI steering committee approved an implementation schedule with concrete resource requirements.

Although the system design decisions were made early in the implementation process, the data flow specifications were more difficult to finalize. As described earlier, the Nirvana system uses a three-tier case status hierarchy (i.e., stage, event, and status) to define case flow through all survey steps. Developing, standardizing, and mapping the three-tiered code structure to the American Association for Public Opinion Research codes was time intensive, but will yield benefits as all projects will report on the same standardized status codes.

The pilot projects generated a set of user-driven recommendations for enhancements to Nirvana and its associated data collection processes. The primary system implementation challenge, as demonstrated in the pilot studies and in later projects, has been to effectively automate the case flow protocol. Although Nirvana establishes a common case flow protocol, survey specifications for a particular project invariably require augmentations to the standard that need to be programmed and tested. As the system is enhanced to support multiple protocols, less augmentation and adaption is required, and this challenge is mitigated.

Conducting an evaluation of Nirvana is an integral part of our continued development of the system. An evaluation of the Nirvana system will assess the extent to which the objectives and desired outcomes have been reached with its implementation. The first 18 months of Nirvana roll-out entail a gradual ramp-up to ensure that various operational components are well-tested before a more widespread implementation. Our evaluation plan focuses on efforts during and after this roll-out period and consists of evaluating elements of each of the key objectives for Nirvana: standardization, integration of survey systems, and access to survey and paradata. The primary approach for the evaluation will be qualitative, though we will conduct descriptive analysis using some quantitative data available. For each of the key objectives, Table 2 specifies certain desired outcomes and the associated metrics included in the evaluation plan to assess the level to which those outcomes have been achieved.

To evaluate the standardization objective, RTI will track the number and types of projects and number of personnel that adopt and implement Nirvana. Although RTI senior management has the expectation that each new survey project will adopt and implement Nirvana, the extent to which this happens – and reasons for doing so or not doing so – will be monitored. For each implementation, we will track the set-up costs, including the time and effort to develop system specifications and to program and test all system modules. These costs will be compared against similar costs for prior projects that did not use Nirvana. Set-up costs will also be compared over time among projects implementing Nirvana. Also measured will be the extent to which one-time customization of processes is required for projects using Nirvana, with the goal that one-time customization is less necessary over time. Such customization will be measured within projects as well as across projects over time. Note that certain customizations will not be one-time-only efforts; they will comprise desired expansions of modules that will be made available for future projects – that is, incorporated into Nirvana's expanded capabilities. The evaluation necessarily will include qualitative as well as quantitative elements, given that no two projects are the same and there will not be directly comparable projects. In as much as quantitative data are available for comparisons and analysis, qualitative data obtained through focus groups will provide valuable context on several of the proposed outcomes. The perceptions of the end users regarding several of the proposed outcomes contribute to the determination of whether goals were achieved in implementing the Nirvana system. Focus groups will be used to engage in discussion and elicit feedback from various end users with varying roles across the projects.

Improving quality is a multidimensional goal that is part of each of the three objectives for building Nirvana. As such, measuring improvement in quality is an important component of our evaluation plan. The standardization and system integration objectives are designed to provide a common, easier-to-use framework for managing and tracking survey activities. Providing richer information to survey managers theoretically allows them to avoid unforeseen problems and proactively develop effective survey strategies. By achieving these objectives, we hope to significantly reduce the number of opportunities for mistakes (i.e., nonadherence to particular specified steps in the process) and consequently the number of errors committed by the survey personnel. Therefore, our plan is to measure quality improvement by tracking mistakes and, in particular, the amount of rework required by these mistakes. We plan to count mistakes that require repeating processes or correcting steps that have already been implemented and track the amount of labor that is invested in this rework. The amount of labor required to correct mistakes is a proxy

Table 2. Metrics for evaluating Nirvana

Objectives	Desired outcomes	Metrics
Standardization	Eliminate duplicative systems	Inventory systems deployed across projects
	Improve operational efficiencies	Count data repositories, files, and programs Count process steps
	Decrease learning curve for users and programmers	Assess extent of customisation required Evaluate usability
	Reduce level-of-effort for system setup, support, and maintenance	Assess cost and effort to specify, set up, and support the system
	Provide more flexible staffing options	Count the number of staff trained to set up Nirvana Count the number of staff that have used Nirvana to manage a survey
	Leverage reusable components of the system	Compare modules deployed across projects
	Increase consistency of implementation	Count the number of projects deploying Nirvana
Integration	Reduce elapsed case processing time	Assess case processing times for critical functions
	Eliminate case processing steps	Count processing steps
	Improve sample management information	Ask survey managers
	Eliminate redundant data	Count repositories, files, redundancies
	Reduce reliance on programmer intervention to move information between systems	Measure programmer time for system support
Access to data	Increase availability of data to survey managers	Ask survey managers
	Improve quality of decisions	Ask survey managers Ask clients
	Improve timeliness of decisions	Ask survey managers Ask clients
	Develop better and more consistent data access and reporting tools	Inventory reporting tools

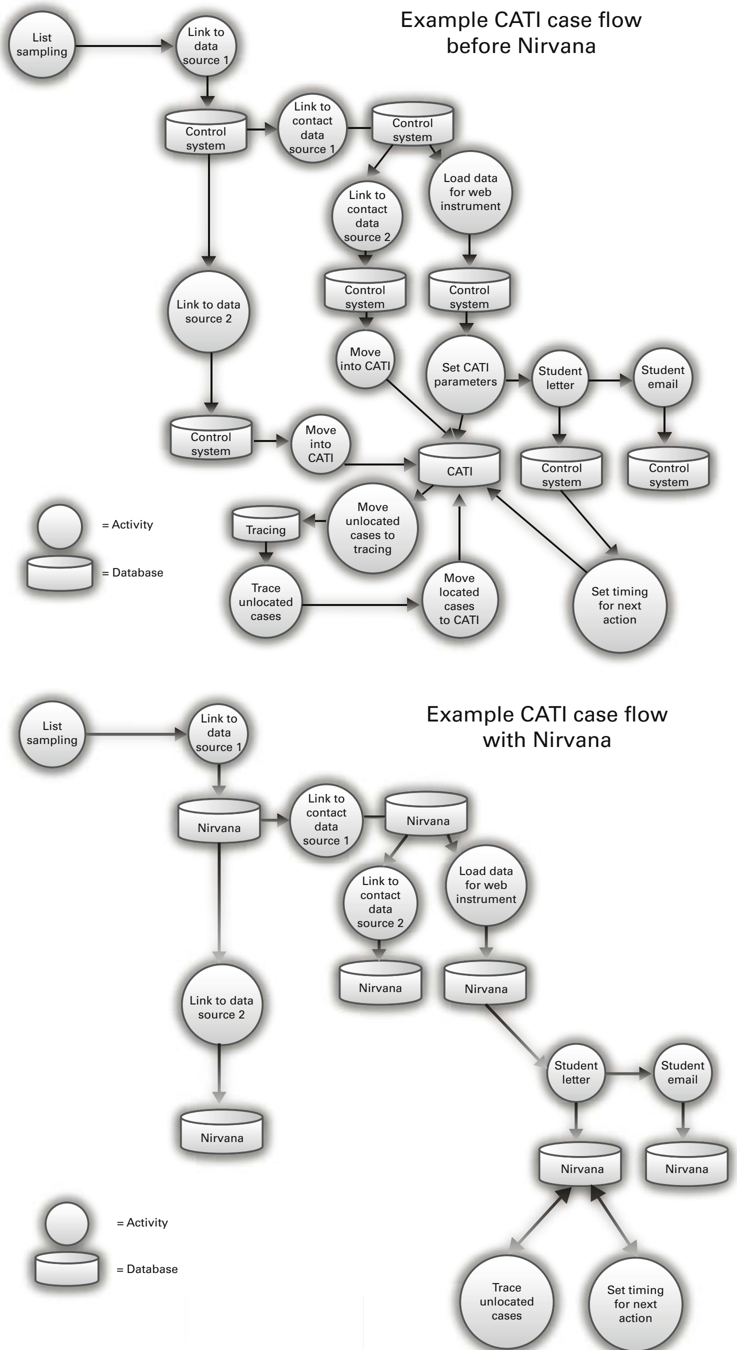

Fig. 8. Example process-flow diagrams

measure for the size of the mistake; therefore, with further use of the system we expect to see a lower mean mistake count and less effort invested to correct mistakes over time.

The system integration achieved through Nirvana addresses a number of important issues: reduction of steps in the process, reduction of the number of opportunities to make errors, increased efficiency, elimination of redundant data and workaround steps associated with multiple systems, and reduction of programmer intervention. Figure 8 depicts a pair of process-flow diagrams associated with the initiation of CATI data collection to illustrate this integration.

The first diagram shows the process flow for a portion of the survey process with pre-Nirvana systems, and the second diagram illustrates the flow for the same activities with Nirvana. A comparison of the diagrams shows a reduction in the number of activities with Nirvana, from 15 to ten. The number of databases has been reduced from three different databases (control system, CATI, and tracing) to a single database (Nirvana). Duplicative information has been eliminated, and information is shared across processes rather than requiring data transfers. The fewer steps and reduction to a single database in turn improve efficiency and reduce the risk of error associated with data inconsistency and redundancy.

A fundamental objective of Nirvana has been providing access to survey data and paradata that is easy, direct, and comprehensive for all who need it (e.g., project leaders, survey managers, data collectors, and programmers). A crucial step in meeting this objective is the establishment of a shared database across all system components and processes. Previously, there were multiple databases that had existed with RTI's data collection. Having all information across processes (e.g., sampling, mailing, tracing, self-administered data collection, CATI, and CAPI) in one location enables survey managers to make decisions in an informed and timely manner due to the availability of trustworthy, real-time status details. Making such data easy for survey managers to use is a work in progress with Nirvana. RTI is in the process of adding query and reporting tools to its ATD framework for Nirvana. To evaluate the success of these tools and access to data for informed decision making, RTI will include discussion of these topics in the focus groups of survey managers, programmers, and project leaders for each of the projects implementing Nirvana during the 18-month initial phase of implementation. Additionally, a user survey similar to the system inventory survey will be administered at the end of the same 18-month phase. An important aspect of this survey will be the extent to which survey managers had easy, timely access to key information to inform the appropriate next steps and to ensure the success of their studies.

6. Conclusions

The overarching goal of the GSSI is to improve survey quality by reducing errors and assisting in the production of relevant, timely, accessible, and cost-effective data. A root cause analysis was undertaken to better understand the challenges associated with a custom systems design and development approach to collect and produce accurate study estimates. This analysis uncovered that the lack of standardization and integration of data collection systems could result in unintended and undetected survey error. The analysis further revealed that the tools necessary to access survey and paradata easily and in real time to manage survey operations were not in place.

The three objectives – standardization, integration of survey systems, and access to survey data and paradata – were among the key design requirements for developing Nirvana, a unified data collection and management system. The standardization and system integration objectives are designed to provide a common easier-to-use framework for managing survey activities. The access to survey data and paradata objective is designed to provide direct and uncomplicated access to this information. Together, these objectives will provide better, timelier, and easier access to data that will enable survey managers to make more informed decisions regarding survey strategies and avoid more unforeseen problems. By achieving these objectives, our goal is to significantly reduce the number of opportunities for error and consequently the number of errors committed.

The Nirvana evaluation plan is designed to measure the extent to which each objective has been met and survey quality has been improved. Our plan details the outcomes expected and the process by which we will assess and collect further information. The plan includes the use of focus groups, surveys, quality improvement metrics, and cost estimates where feasible. Further enhancements will be made based on the results of our evaluation plan.

ATD is a noteworthy improvement to the survey process that was made possible by the integrated system provided by the GSSI. ATD maximizes survey quality by allocating budget and personnel resources to survey activities depending on priorities and the potential to reduce total survey error. It has been used to identify interviewers who deviate from an expected norm, triggering the need for further investigation or retraining. It has been used to monitor the status of the sample to determine the timing and size of sample releases. It has been used to monitor item nonresponse rates and response distributions to key survey variables. ATD has also been used to compare data collection costs with budgeted costs to identify discrepancies and forecast final costs. For example, in a recent longitudinal, in-person survey on child health and well-being, ATD charts showed that labor hours per completed interview were less than budgeted and expected to increase only slightly through the remaining data collection period. The anticipated cost savings were used to plan and fund a study on nonresponse bias in the survey estimates. The potential uses of ATD are plentiful: comparing weighted and unweighted response rates by sample geography over time, modeling final response rates, tracking indicators of bias, comparing characteristics of respondents with the sampling frame, identifying interviewer falsification, and observing the real-time effects of protocol changes such as contacts by different modes or the offer of incentives.

Together, GSSI and ATD have improved our ability to take advantage of previously untapped data to improve the quality of survey estimates. While time and resource intensive to establish, our investment is expected to pay off in terms of reduced costs and errors over multiple projects, greater operational efficiencies, and enriched information for decision making.

7. References

Biemer, P. (2010). Total Survey Error: Design, Implementation, and Evaluation. Public Opinion Quarterly, 74, 1–35.

Biemer, P.P. and Lyberg, L. (2003). Introduction to Survey Quality. New York: John Wiley & Sons.

Couper, M.P. (1998), Measuring Survey Quality in a CASIC Environment. Invited paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, TX. Available at: http://www.amstat.org/sections/srms/Proceedings/ (accessed June 20, 2012)

Curtin, R., Presser, S., and Singer, E. (2005). Changes in Telephone Survey Nonresponse Over the Past Quarter Century. Public Opinion Quarterly, 69, 87–98.

de Heer, W. (1999). International Response Trends: Results of an International Survey. Journal of Official Statistics, 15, 129–142.

Groves, R. and Heeringa, S. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society Series A: Statistics in Society, 169, 439–457.

Hunter, D.C, Mitchell, S., Carley-Baxter, L., and Keating, M. (2012). Using Paradata to Monitor Survey Production, Cost, and Quality within an Adaptive Total Design Framework. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC. Available at: http://www.fcsm.gov/12papers/Hunter_2012FCSM_V-C.pdf (accessed June 20, 2012)

Juran, J.M. and Gryna, F.M. (1980). Quality Planning and Analysis: From Product Development Through Use. New York: McGraw-Hill.

Steeh, C., Kirgis, N., Cannon, B., and DeWitt, J. (2001). Are They Really as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century. Journal of Official Statistics, 17, 227–247.

Tortora, R.D. (2004). Response Trends in a National Random Digit Dial Survey. Metodolski zvezki, 1, 21–32.

United Nations (1999). Information Systems Architecture for National and International Statistical Offices – Guidelines and Recommendations. Available at: http://www.unece.org/stats/documents/information_systems_architecture/1.e.pdf (accessed June 20, 2012)

Received November 2011
Revised June 2012
Accepted October 2012