Demographic Research a free, expedited, online journal
of peer-reviewed research and commentary
in the population sciences published by the
Max Planck Institute for Demographic Research
Doberaner Strasse 114 · D-18057 Rostock · GERMANY
www.demographic-research.org

DEMOGRAPHIC RESEARCH
VOLUME 2, ARTICLE 6
PUBLISHED 27 JUNE 2000
www.demographic-research.org/Volumes/Vol2/6/
DOI: 10.4054/DemRes.2000.2.6

The Household Registration System: Computer Software for the Rapid Dissemination of Demographic Surveillance Systems

James F. Phillips

Bruce B. MacLeod

Brian Pence

The Household Registration System: Computer Software for the Rapid Dissemination of Demographic Surveillance Systems

James F. Phillips [1]

Bruce B. MacLeod [2]

Brian Pence [3]

Abstract

Although longitudinal experimental community health research is crucial to testing hypotheses about the demographic impact of health technologies, longitudinal demographic research field stations are rare, owing to the complexity and high cost of developing requisite computer software systems. This paper describes the Household Registration System (HRS), a software package that has been used for the rapid development of eleven surveillance systems in sub-Saharan Africa and Asia. Features of the HRS automate software generation for a family of surveillance applications, obviating the need for new and complex computer software systems for each new longitudinal demographic study.

[1] James F. Phillips is Senior Associate, Policy Research Division, Population Council, One Dag Hammarskjold Plaza, New York, NY 10017, email: jphillips@popcouncil.org.

[2] Bruce B. MacLeod is Associate Professor, University of Southern Maine, 96 Falmouth Ave., Portland, ME 04103, email: macleod@usm.maine.edu.

[3] Brian Pence is Data Analyst, Policy Research Division, Population Council, One Dag Hammarskjold Plaza, New York, NY 10017, email: bpence@popcouncil.org.

1. Introduction

Population laboratories have been an invaluable resource for research on the determinants of demographic dynamics, the prevention of disease, the impact of health technologies, the efficacy of family planning services, and the consequences of demographic change. Since longitudinal data are complex to compile and manage, however, most population laboratory-based research has been produced by a few generously funded, rigorously managed, and well-equipped research stations located in developing countries. Of these, the Matlab Demographic Surveillance System in Bangladesh exemplifies characteristics of population laboratories that justify their cost: Individuals in a large population are registered in a computerized database with field surveillance designed to record all events and changes in household relationships over time. Procedures record person-days of risk accruing to individuals so that rates, intervals, and relationships are readily linked with covariates of interest, thereby providing the machinery for a wide range of practical health and demographic research applications [18,51].

The logical rigor of the Matlab data management system which makes this important longitudinal health research possible is also a factor constraining its replication. The posting of demographic information to computer files is associated with a myriad of checks on the logical consistency of new events with extant registered information. Over 4000 lines of computer code are required to check that computer registers correspond to logically possible relationships, event patterns, and dynamics. Generating a system of this size and complexity requires a team of skilled computer scientists, demographic research guidance, and field capacities to respond to queries and advise system managers of needed changes. The Matlab system was developed over a period of three decades of intensive field research, but few contemporary studies have this luxury of time and resources. Systems must be developed rapidly for focused investigations. This typically leads to compromising simplifications that inadvertently lead to system limitations, eventually producing intractable data management challenges [Note 1]. The long delays involved in developing surveillance systems and utilizing surveillance data have elevated research costs and diminished the usefulness of data for policy. As a consequence, the field station model has been criticized as too costly and complex for practical applications. The few sites where systems are functioning are overburdened with multiple research protocols.

This paper presents an overview of an initiative addressed to the problem of generating low-cost and rigorously designed demographic surveillance systems. The Household Registration System (HRS) is a software system currently in use in eleven research sites in Africa and Asia. The HRS exploits recent developments in object-oriented programming and automated program generation to simplify the process of developing data management systems for a diverse collection of longitudinal household studies. The computational foundation of the HRS is a relational database system that resolves many of the complex data management issues associated with monitoring births, deaths, marriages, and migrations in a fixed geographical area. Experienced or novice programmers can build on this foundation by changing or adding to a collection of visually presented objects that correspond to the various items on data entry screens.

Some of these objects have small amounts of computer code (ten to twenty lines) associated with them that control, for instance, the range of legal values, legal relationships with other variables, and the circumstances under which the field can be edited. Programmers who modify and extend the system need only to specify variable names, legal values for variables, and screen layouts for additional data. The programmer can also give rules governing how new data must relate to other variables or to previously compiled information, either within the same record or in a different record. A user can then regenerate the system with project-specific data and new code integrated with the HRS core to create a project-specific demographic surveillance system.

The code generated by the HRS maintains, retrieves, and reports on cross-sectional and longitudinal data associated with studies of households and their members. Investigators are freed from the need to rewrite and modify large amounts of computer code. The generated database programs incorporate rigorous design principles, enforce logical checks on much of the database information, and are capable of reporting key demographic rates. The amount of error-free program code generated for a particular application is substantial, fulfilling the requirements of demographers, epidemiologists, and social scientists for comprehensively cross-referenced information on households and members.

De novo development of data management software required for longitudinal demographic surveillance can require years of technical assistance from highly trained specialists; the end result is often a system that no one else can understand, modify, or manage. This “expert assistance” model extracts ownership from the research institution and reduces the policy relevance of research. The HRS, in contrast, embodies an empowering form of technology transfer to developing country institutions. While the system is appropriate for research in the developed world as well, the lack of permanent identification numbers and computerized records in most developing countries vastly complicates the data management of longitudinal research. The HRS is built around a core structure that is flexible enough to accommodate most longitudinal studies of populations. Innovations in object-oriented programming permit even beginning programmers to easily modify these core specifications and tailor the system to their needs. Finally, the structure of the HRS lends itself to a remote technical assistance model, whereby data managers can email difficulties to an HRS “expert” who can in return email short code segments to be incorporated into that site’s HRS. After reviewing the technical requirements of longitudinal population studies [section 2] and various alternatives to the HRS data model [section 3], this paper discusses the structure of the HRS and describes the process of tailoring and operationalizing the system to a specific study site [section 4]. Finally, in "Next Steps," we address plans to further enhance the system through modifications that are being incorporated into a new version, the HRS-3, which is currently under development [section 5].

2. Longitudinal Household Studies

2.1 The role of field stations in health research

The data management issues associated with monitoring births, deaths, marriages, and migrations in a fixed geographical region are complex and can complicate health research and intervention efforts. Despite the intrinsic complexity of longitudinal data systems, investment is justifiable on scientific grounds. Demographic events of interest, such as neonatal, post-neonatal, child, and maternal mortality are sufficiently rare as to require precise information on the events of interest and the population at risk. Cross-sectional survey techniques, so widely used in fertility research, are ill-suited to mortality studies because the determinants of mortality, morbidity, nutritional status, lactational behavior, etc. are either intervening events or longitudinal processes that can only be effectively studied in concomitant event history analyses. Recall biases associated with retrospective studies seriously compromise inference. Furthermore, health service interventions may vary over time, and their impact may be a function of their timing relative to seasonality of adversity and specific episodes of illness. When household relationships and characteristics are maintained longitudinally, time-referenced data can clarify the causes and consequences of adult, maternal, and child morbidity and mortality.

For many health studies, members of a household must be studied as a group. Strong observed correlations between social and economic status and health outcomes attest to the value of information on household member relationships, customs, and behaviors in research on the determinants of survival. It is crucial for research on health interventions to account for household-level characteristics, since health service effectiveness is determined as much by household social and behavioral factors as by the efficacy of medical technology.

Health research protocols require data on specific determinants of illness and mortality over time. Because mortality events are rare, research that focuses on mortality outcomes requires observation of large study populations. Morbidity events of interest are intermittent, requiring prospective studies rather than retrospective surveys. Field stations are established where individuals in populations can be observed in laboratory fashion, and interventions can be assigned to individuals in randomized trials or to groups in factorial experiments. Research on mortality requires a defined population at risk, and this requirement obliges investigators to monitor all components of population dynamics: births, deaths, in-migrations, and out-migrations [Note 2]. Monitoring population dynamics with demographic surveillance is thus the core scientific resource for field station-based epidemiological research.

Field stations have been used for epidemiological research ever since the 1920s [Note 3]. Early epidemiological research demonstrated the influence of sociological and demographic factors on the occurrence of disease and illness, establishing the importance of social research in understanding morbidity risks [Note 4]. In the 1950s, the use of field stations was expanded from investigating disease to conducting the controlled trial of health interventions [Note 5]. Field stations were developed in Guatemala, India, and Pakistan for international health research programs [3,27,28,31,35,52,54,55,63]. Field stations were also employed for population research in the 1950s. The Khana study, launched in India in 1953, was a test of the impact of family planning service delivery on fertility in eleven rural Punjab villages. A similar study was launched in Singur, West Bengal in 1956 [Note 6].

2.2 The Matlab station in Bangladesh

Of the field stations that have been established in developing countries for health research, the most productive has been the Matlab station of the International Centre for Diarrhoeal Disease Research, Bangladesh. Originally established in 1960 as a project for testing vaccines against cholera, the Matlab station eventually became a population laboratory for a wide range of research on the epidemiology of enteric disease, the determinants of health behavior and survival, and the impact of community health and family planning services [Note 7].

The Demographic Surveillance System (DSS) is the core scientific resource of the Matlab field station. The DSS enables researchers to conduct research on a defined population over time. The DSS identifies the population in study areas at any point in time and monitors all components of demographic dynamics over time: Births, deaths, migration into study areas and migration out of study areas. In the Matlab DSS, risk is measured at the individual level and calculation of person-days of observation permits individual-level longitudinal studies of greatest scientific interest: Individual-level randomized trials, causal analysis of determinants of demographic outcomes, the analysis of concomitant events, and other statistical studies of covariates of rates. By following individuals over time, the Matlab DSS thus provides the basis for a wide range of research protocols and activities in health and population [16-18,40].

Although Matlab has been a productive source of research on health and population issues, Matlab data management technology has not been utilized by other field stations. For the DSS operation, several thousand lines of database computer code are managed on a mainframe system in Dhaka, geographically remote from Matlab field operations. Field work and computing are managed separately, an operation that requires a large clerical staff for managing data correction operations.

Managing this complex computer and field system requires a team of highly trained experts. Changing the simplest parameter of system code involves time-consuming and expensive expert assistance, and replicating Matlab computer operations demands a substantial investment in system development. New systems, in turn, have little to gain from using Matlab software designs that require expensive mainframe hardware.

The HRS is designed to resolve limitations of the Matlab system while replicating features of Matlab technology that make it a productive scientific center. The HRS preserves Matlab principles of structure, linkage, and checking, while introducing modern principles of database management, low-cost microcomputer hardware, and flexible and extensible software. We turn next to a discussion of alternatives to the Matlab/HRS data model before describing the HRS software.

3. Data Management for Longitudinal Systems

The Matlab data management system, while flawed as a model for replicable surveillance, represents an appropriate model for longitudinal community health research [Note 8]. At least four alternatives to the Matlab population-based surveillance approach have been proposed. Each has been associated with technical pitfalls.

3.1 Cohort studies

The most common longitudinal studies are cohort studies in which individuals at risk are followed over time and events are monitored. Individuals are lost to observation through out-migration and death, but life table methods are used to adjust for possible biases that may arise from sample loss. The main limitation of cohort studies is attrition bias. This is particularly problematic in places like West Africa where migration is extensive. In the HRS design, monitoring social units provides a continuing basis for renewing the population at risk. In an HRS cohort, data on individuals leaving observation through out-migration are recorded together with information on individuals migrating into the study area. By recording both in-migration and out-migration, the HRS preserves an element of representativeness that is not feasible in cohort studies. Moreover, designs which use households rather than individuals as the primary level of organization provide means of generating data on migration that can be used to match records of in-migrants with out-migrant records, thereby diminishing censoring of individual event histories. Even with such aides, matching is a complex and time-consuming process. However, as discussed in section 4.2, features of the HRS in conjunction with appropriate field procedures can prove effective in resolving internal and return migration.

Although cohort studies are a seemingly efficient means of conducting longitudinal research, they are often so focused that a given study contributes little to other longitudinal studies that may be planned in the same population [Note 9]. In general the combined cost of a series of independent cohort studies is much higher than an HRS core which monitors demographic dynamics in a defined study area, provides a platform for the rapid implementation of multiple cohort studies, and allows new initiatives to build on a wealth of pre-existing data for the study population.

3.2 Longitudinal events with cross-sectional censuses

Experimental studies requiring rates on aggregate populations have been conducted with surveillance systems that monitor births and deaths only, avoiding the difficult data collection and data management problems associated with migration, under the assumption that periodic censuses provide accurate estimates of populations at risk [4,58]. This is a reasonable design if all that is needed are accurate rates for aggregate areal data. However, its usefulness is limited by the fact that individual level analysis is not possible, since risk accumulating to individuals cannot be monitored unless all components of demographic dynamics are observed. Although new “verbal autopsy” techniques, in which structured reporting by non-medical staff is used as a cheap way of identifying cause of death in broad categories, can simplify event surveillance and produce fairly reliable data, the logistical demands of conducting repeat censuses and event surveillance remain complex. Longitudinal studies require reinterviewing, and the process of recruiting, training, and deploying workers in annual rounds can actually be more costly than maintaining a team of interviewers who interview respondents in shorter rounds. The reason is that periodic restaffing of surveys in developing countries can occupy the time of expensive senior scientific staff and incur high overhead costs associated with recruitment and repeated training. These designs lead to rather superficial descriptive analyses of experimental demographic endpoints, without ancillary research on determinants or covariates of project results. Since risk at the individual or family level is unknown, most statistical methods, including regression analyses, are not applicable. Research is thus confined to the narrow goals and purposes of experiments, failing to capitalize on the full potential of field station-based studies for research on demographic determinants or interrelationships.

3.3 Relational files linked from batch files

Longitudinal studies are often designed as systems in which separate batches of data are managed for each type of demographic event, with periodic episodes of merging, matching, and linking of batches to construct relational files of longitudinal event histories for analysis. The batch-and-episode approach invites designs in which computer operations are not coordinated with field operations, allowing flawed data to accumulate before errors are corrected. The record of such studies suggests that economies achieved by simplifying data entry have been offset by the long-term costs and complexity of managing, linking, and cleaning data. If logical problems are not detected until the longitudinal files are constructed, even minor logical lapses can lead to major delays and difficulties, in some cases preventing use of data for any purpose other than rudimentary descriptions of study populations [45,62].

3.4 Panel surveys

Throughout Africa, field experiments in family planning have been implemented to test the demographic impact of service delivery systems. In the most common type of field experiment family planning is monitored as the outcome variable, although in a few studies fertility histories are compiled in the baseline and end-of-project surveys. This design permits use of statistical controls for baseline reproductive motives in the analysis of program effects. While panel designs provide important insights into the impact of family planning experiments, their demographic impact cannot be precisely gauged unless longitudinal surveillance approaches are used.

First, retrospective recall of events is always flawed, particularly in traditional nonnumerate societal settings. Panel survey rounds typically take place a year or more apart. Retrospective recall of births and deaths may be reliable over such durations, but complex procedures for recording proximate fertility determinants require information that is not reliably recalled. Family planning use, for instance, overlaps with post-partum amenorrhea or substitutes for traditional fertility regulation practices. Prospective observation is required in order to accurately measure and elucidate these dynamics.

Second, panel designs administered in annual survey rounds have proved to be no less complex to conduct than prospective surveillance studies administered in quarterly prospective cycles. In rural traditional settings where field stations are based, the task of hiring and training teams for annual panels is difficult and costly; a professional team working in a continuous interviewing cycle can be less expensive to manage and develop than recurrently convened interviewing teams because interviewer salary costs are low relative to the costs of senior project leaders and supervisors. A continuous interviewing design provides data that are relatively free of recall bias and free of the administrative complexities of regenerating interviewing teams for successive panels.

Finally, the complexity of panel designs has proved to be difficult to manage in the absence of surveillance data. As the gap between interviews increases, censoring increases owing to migration and data linkage problems. Fieldwork associated with resolving censoring problems in panel designs is more costly than corresponding field costs of surveillance systems with interviewing rounds of 90 days or less. Panel designs are selected to save costs; however, it is likely that repeat survey costs will be lower than surveillance costs only if no attempt is made to link responses over time.

4. The HRS Software System [Note 10]

4.1 A model for household registration and surveillance

The HRS is a software system currently operating in eleven research centers in Asia and Africa [see Table 1] that maintains a consistent record of significant demographic events that occur to a population in a fixed geographic region, generates up-to-date registration books that are used by the field workers, and computes basic demographic information (age-specific birth, death, and migration rates; age and sex distributions of the population; and life table functions). Investigators add to and sometimes modify this core set of data and software to tailor the HRS to their particular projects. The investigator can insert data fields, define logical checks, and interactively specify the screen layout. Core data and code specification and the ability to modify and add to the core specification all combine to give the user the flexibility to develop a site-specific household registration program [37-38].

4.2 Integrated data collection and data management

The HRS data management system operates in conjunction with a field system for collecting data on household members. The HRS requires an integrated field operation and data management system, as shown in Figure 1. No particular requirement for the duration of the work interval is specified in the HRS. However, 90 days has been selected for most HRS applications because this interval is short enough to ensure that all pregnancies can be seen by interviewers within a round, but long enough to ensure that all data can be entered, checked, processed, and reported in the work cycle. The central management task in HRS operations involves ensuring that field operations are linked to computer operations so that errors and problems are noted, fed back to interviewers, and corrected within the routine work cycle. Transactions with paper "Household Registration Book" (HRB) registers are designed to match computer database transactions. In a typical HRS application, detailed instructions are issued to ensure standardized interviewing and recording, with registers designed to maintain four rounds of interviews to facilitate probing and data checking at the time of visitation (Item a, Figure 1). Supervisors check a random sub-sample of the study population. This re-interview is used to detect problems to be discussed in monthly staff meetings. The HRB is arranged by household in the order that households are contacted.

Figure 1 about here

Upon completion of a round of interviewing, registers are passed to data entry staff for updating the computer database (b). Upon entry, a series of logical checks are imposed to assess the consistency of the new information with data already present in the database. Clean data are archived (c), and an error report is generated (d) and reported to the field (e). Supervisors examine error reports and review actions to be taken with relevant staff. The HRB is manually corrected (f), and relevant corrections on event forms are communicated to the data management center for processing (g). Data passing all tests are archived, and information from households failing logical tests is printed with relevant error messages for field diagnosis and correction. Thus each cycle generates fully edited and cleaned data before updated HRBs are printed (h) or a new cycle of data collection begins (a).

The critical event in the Figure 1 cycle is the shaded box between (b) and (c), which imposes uniform logical rules on data as they are entered and provides reports to field supervisors and data managers at the time of data entry [29]. This step ensures both that the clerk is entering legal values for the variables and that any information recorded by the system is logically consistent with demographic information compiled in the past. For instance, the visit date must lie within the three-month period of the round in question, and the date of a recorded event cannot fall after the visit date. A woman who was observed as pregnant during the previous round must either be observed as still pregnant or have a pregnancy outcome (birth or miscarriage) recorded. A member who has migrated out of the study area cannot have any events recorded unless s/he has migrated back in.

Accurately monitoring migration represents one of the most daunting challenges to longitudinal research, especially in developing country settings that lack computerized records and permanent individual identification numbers. Monitoring the population in a fixed geographic area can reduce censoring by tracking migration and linking individual histories from in- and out-migration records. In practice, of course, the field process of resolving internal and return migration is difficult, time-consuming, and error-prone. The most significant problems occur when departure is not detected in a timely fashion, i.e. when an individual who moves within the study area is observed to have arrived in a new location before the corresponding departure from the previous residence is registered. Rapid resolution of this double entry is necessary in order to minimize the artificial inflation of the population at risk.

Features of the HRS help streamline the reconciliation of migration movements. The system generates reports of all in-migrants in a round who are not recognized as previous members of the study population and all out-migrants who have not been recorded as an in-migrant elsewhere in the same round. These reports can be generated for specific subsets of the population such as men, children, or residents of particular geographic areas, in order to facilitate the matching process. Plans for subsequent versions of the HRS include features to further simplify this process. Even with these features and personal identification information (such as names and dates of birth), resolving migration remains a difficult problem. It is essential that the software operate in conjunction with well-designed and effective field procedures to follow up and resolve migration inconsistencies.

The design of data collection procedures has been informed by database concepts. Workers are equipped with a register, printed from the population database, that is designed to facilitate data management. Pages are arranged in the order that households are visited. Rows in the register correspond to individuals in the households, and page headings include the name of the household head and information about characteristics that household members share as a group, such as primary family religion and household wealth and size. Rows for individuals list names, ages, relationships, and other basic information. Each column in the worker register corresponds to a visit cycle, and space is provided in each column for workers to enter codes corresponding to vital events or household status changes observed during that round (births, deaths, marriages, migrations, and pregnancies). This procedure limits the flow of loose paper and enforces data linkage between observed and past events at the time of the visit. If the register lists a woman as being pregnant at the last visit, the interviewer will know to probe whether she is still pregnant or whether a birth or miscarriage occurred. Since event data are recorded together with data on the individuals at risk, the worker's register structures data collection in a manner that is conceptually similar to computer operations for linking and checking records in the database.

Data entry involves passing registers to the data entry clerks who key in the household numbers and event codes and perform other requisite transactions with the database [Note 11]. The entry of only the events that affect the structure of a household (births, deaths, marriages, migrations) represents one of the fundamental differences between the HRS and other batch-oriented systems. Rather than completely reinterview the sample population, create a new record for each household, and then, with the computer, link those records back to previous interviews (an error-prone process), individuals in the study area are linked to households and past-event histories with the paper registers carried by the interviewer (mirroring the linkages in the relational database). This substantially reduces the quantity of data entry and the subsequent costs of consistency checking. Most data inconsistencies are caught at data entry and reports of these inconsistencies are printed for supervisory action. These field operations are deemed sufficiently well developed at this point to be generic to any longitudinal study of demographic dynamics [Note 12].

4.3 The Core HRS Data

There are characteristics of households, members, relationships, and demographic events that are common to all longitudinal studies of human populations. The logic for these characteristics is embedded in the core system. The HRS structures data and maintains logical integrity on the following basic elements of a household unit:

- All households have defined members. Rules unambiguously exclude nonmembers.
- All households have a single head at a given point in time, and members relate to one another and to the head in definable ways.
- Members have names, dates of birth, and other characteristics that do not change.
- Events can occur to members, such as death, birth, in- and out-migration, and marital status change. Events change household membership or relationships according to fixed rules.
- Episodes (longitudinal events such as pregnancy, marriage, or residency) occur to individuals at risk (i.e. active members) and must follow simple logical relationships.

This core structure must be adapted to conform to the local population and area. A household member is usually defined as a person registered as resident in the household at the initial enumeration, born to a member of the household, or who migrated into the household. The definition of a household head is somewhat more complex. In most current HRS applications, for instance, the household head must be a member of the household. A study operating in a polygamous society might need to relax this requirement, since one man could be the head of several households of which he is not a resident.

Although the list above is seemingly trivial, everyday relationships tend to become complex and unwieldy when arrayed as a logical system of longitudinal population data. Portraying even simple relationships requires rigorous standards to avoid error. For example, to register a death in the population, a household member must be resident in the study area; a birth to a woman five months after she gave birth to another child is an inconsistent event. This logic may seem mundane, but lapses in the integrity of data management can generate deaths to individuals who are not logically members of the risk set, births to nonexistent mothers, or migrations among the deceased. The accumulation of minor logical lapses can render data useless for all but the most basic analyses. In addition, errors generated at one stage tend to cause additional errors in later stages; this compounding effect can quickly cause the database to become completely out of step with the study population. Longitudinal household research requires defining rigorous and unambiguous standards for data management. The logical integrity of the HRS core when paired with appropriate field procedures permits these standards to be met.

4.4 The functional components of HRS software

Viewing the HRS software from a user's perspective provides an overview of the HRS program structure. First, we present screens as they appear to a data entry clerk. Subsequently, we discuss data entry screens that a programmer would use to modify and extend the software.

When a user first logs onto the HRS, the following main menu screen appears (Figure 2):

Figure 2 about here

The options in the main menu of the HRS are:

- Data entry: Allows for the entry, deletion, and editing of the baseline and longitudinal data. Baseline household information includes the household location, individuals within the household, relationships between individuals, and familial social groups. Longitudinal information includes basic information related to pregnancy observations and outcomes, deaths, migrations in and out of the study area, marriages, and any other measures specified by the investigator.
- Validation: Checks the logical consistency of data for subsets of households and members.
- Reports and Output: Calculates and displays demographic rates and life tables. Age-specific and overall rates can be computed.
- Visit Register: Used to print the household registration book. The household registration book is used by the field workers to record information during household interviews.
- Utilities: An option that is primarily used by the system administrator. It includes capabilities for adding new user IDs, setting interview round information, and generating reconciliation reports to help track down unreported pregnancy outcomes and unmatched internal migrants.

Collectively, these functions form a part of every application generated from the HRS.

Most HRS user interaction occurs in the data entry screens. Two screens are fundamental to the data entry process: the baseline screen which allows data entry from the initial enumeration, and the update screen for the entry of longitudinal information. The data entry window for baseline information is presented in Figure 3:

Figure 3 about here

Much of the information displayed in this screen can be adjusted to suit project-specific needs. For example, the labels and many of the field values can be changed. Also, while the HRS requires that all locations and individuals have unique IDs, the format of the ID (character, numeric) can be changed from project to project. The HRS requires that the gender of individuals be input to the system, but again, the format of this information (e.g. M/F, 1/2) can be changed. More details are provided in the HRS user documentation (http://www.popcouncil.org/hrs/hrs.html).

After entering the initial enumeration of locations, individuals, and relationships, the HRS is used to generate field books for entering demographic information collected during subsequent visits to the locations. After every visit round, the field worker will hand the books to the computer center and the changes in the demographic status of the household can be entered using the update screen (Figure 4):

Figure 4 about here

The data entry clerk enters the information at the top of the screen (location ID, ...). Once this information is provided, the grid of individuals currently resident at the location is displayed. If an event occurred to an individual (for example, a birth to Ajua Adugbire), then the user would scroll down to the relevant individual and click on the button representing the event (in our example, the Preg Outcome button would need to be pressed). Clicking on buttons causes an event-specific entry screen to appear that collects additional information about the event (an example of an event screen is provided below).

The above screen can be adapted to accommodate new types of "events" by adding additional event buttons. For example, suppose that information about malaria fever episodes occurring to children under five years old is required. A "Malaria fever" button could be added and checks could be put in place to restrict entries to the appropriate subset of individuals (in this example, under-five children). The process of adding new fields is described in section 4.6.

The event forms collect additional information about the particular event. One of the more complex events is a pregnancy outcome. After selecting Ajua Adugbire and clicking on the Preg Outcome button, the data entry clerk would see the following screen [Figure 5]:

Figure 5 about here

Certain information is automatically entered by the system (field worker, mother ID, observation number, and previous birth history). A unique Event ID is automatically generated to refer to this birth. The clerk fills in the other fields, including type (live vs. still birth), date, and father ID. Since this is a live birth, the system generates a new Individual ID number for Akalou Adugbire. Her father's ID is that of Akumdaare Adugbire, the household head [see Figure 4], thus her "Relation to Head" is recorded as "CHD" (child). Finally, the Status of Data field reflects the consistency status of the record. It initially contains the value "P" (pending); when the record has passed all validation checks, it is set to "V" (valid).

Information can be added to this core data about a birth. For example, a birth attendant field could be added to the information at the top of the screen and a birth-weight field could be added to the child-level fields in the grid.

The other event forms are similar in structure to (although in most cases, less complicated than) the Pregnancy Outcome form.

4.5 Output

Once data has been collected over a time interval, the HRS provides capabilities for generating key demographic rates, such as fertility, mortality, and migration. The HRS can also produce life tables and a population age distribution. The various rates can be calculated for a user-specified time interval and for different geographic subsets of the population.

4.6 Internal Logic and Customization of the HRS System

Typically, a study will require amending the core HRS specification to include new systems of variables on household attributes, individual characteristics, or events. The HRS is built from the form (data entry screen) menu and database builders of the Microsoft Visual FoxPro System (currently developed in Version 6.0). FoxPro is not needed to run the HRS; however, it is required to make any modifications to the programming. The FoxPro tools encourage and facilitate a modular, object-oriented software development approach. In the HRS, most of the objects represent variables from the HRS core tables and these objects can appear on the data entry, reporting, and rate generation forms. Small "code snippets" are segments of code that can be "attached" to these objects. Some code snippets control when data can be entered for a variable (i.e. at baseline or in updates); others enforce rules for legal variable values and legal relationships with other variables.

Collectively, the database, form, and menu specifications along with their associated code snippets define the HRS core. When changes to the core are required, a programmer locates the database table, menu, or form object where changes are needed and then works with the code snippets attached to the object. Since there are only a few code snippets attached to an object and these code snippets are usually short, the process of modifying and extending the HRS is significantly easier than changing code in a thousand-line program.

This method of modifying and extending software is very different from the process used only a few years ago. Formerly, software development required a programmer to manage both the control logic of a program (for example, the sequence in which fields in a data entry screen are visited) as well as the semantic logic (for example, logical checks). When changes were required, the relevant code section had to be manually located in the many thousands of lines of code. Seemingly straightforward modifications such as altering the format of the individual ID would require changing code in many different places. Then the programmer had to ensure that those changes did not cause problems with the control code in other portions of the program. Only experienced programmers could work with large software systems written in this way. In contrast, the HRS specifies the format of the individual ID in only one place (the class directory). When the format of the ID is modified, the change immediately cascades through the entire program, updating the format of the ID in every place that it appears.

The process of code modification is mechanized in the HRS. The new object-oriented, visual style of programming is much easier for both beginning and experienced programmers. Point-and-click specifications combined with a reasonable default behavior allow the beginning programmer to develop adequate software systems with relative ease. In addition, this new form of software development allows an experienced programmer to develop objects that can be dropped into an application. These objects can then be used by beginning programmers to develop new applications. The HRS contains many predesigned objects that are intended to facilitate demographic surveillance system development. For instance, template forms exist for each of the three different types of data (invariants, episodes, and events – see the paper detailing the Reference Data Model under "Publications" at http://www.indepth-network.org/). In starting a project to study malaria incidence among children under 5, the programmer would first open a new "Event" form which would come equipped with all of the navigational buttons along the bottom. Data fields are other examples of predesigned objects. The programmer can cut-and-paste the Individual ID, Location ID, Date, and Field Worker boxes from another event form, inheriting all of the associated logic.

The FoxPro form builder allows objects in a form specification to be visually manipulated. When a programmer wants to change the form layout or variable labels, the appropriate object is selected and then moved, edited, or deleted. It may be necessary to modify the core specification for a number of reasons. This may involve adding a variable, changing data prompts to a new language, changing logical constraints, or adding controls governing when data can be entered. For example, the update form specification appears to the programmer as follows (Figure 6):

Figure 6 about here

Obviously, the specification and the data entry form seen by users are visually very similar. If there is a problem with a data field, such as the Location Number, then a programmer can mouse-click on the corresponding location field, display the list of properties associated with that field, and make any corrections. Control buttons at the bottom of the screen also serve a useful function. These buttons contain a substantial amount of logic to control user interaction with the form. A novice programmer who needs to create a new data entry form need not understand all of the detailed logic associated with a control button, but can simply paste the control buttons onto the new data entry form and inherit all the embedded logic. This is an extremely powerful utility for code reuse. If the reuseable objects are sufficiently powerful and versatile, then programmers need only insert the right objects in the right places to invoke the requisite logic and operation.

Asking for the properties of any of the objects embedded in the form will reveal more of the details of that object's behavior and characteristics. For example, a few of the properties associated with the location number are shown in Figure 7:

Figure 7 about here

The object has tabs to reflect the various categories of attributes and behaviors. The Data tab allows the programmer to associate the object with a particular field in a table. Methods allows the programmer to attach computer code to particular events that occur during the running of the program. In the above example, the “When” code snippet returns ”true” if a user can enter data into the field and “false” otherwise (in this particular case, the code in the lower box stipulates that this field cannot be edited after the data has been entered). The Layout attributes affect the appearance of the object on the form, and the Other tab contains a few miscellaneous attributes.

Logic about legal values and relationships for variables can also be associated with the objects embedded in a form. But often a better place to put this logic is in the database table itself. This logic can be associated with a field in a table or with the entire row in the table. If data is inserted into the table, then the logical checks are applied to the data before it is inserted; if there are any errors, the data cannot be inserted. Logical checks involve both ensuring that the entry is a legal value for this variable and that it is consistent with past information on that individual or household. This capability is one of the most significant and powerful features for the HRS programmer. It allows changes to be made to some of the HRS logical checks (for example, changing the allowable codes for relationship to household head) as well as adding new project-specific logic to HRS fields or new fields. The following figure shows some of the fields for the individual table as well as more detailed information, including a logical rule, for the gender field (Figure 8):

Figure 8 about here

The Field Comment gives the programmer directions on how to change the legal values for the variable. In this case, to change the coding scheme of Gender from M/F to 1/2, the programmer would simply change “M” and “F” to “1” and “2” in the Rule (and adjust the Message accordingly). Imposing consistency checks with past data (for example, to allow a pregnancy observation to be recorded only for a woman) is a significantly more sophisticated task which requires attaching row-level rules to the database tables themselves.

We anticipate that most of the consistency logic for an application can reside in the table definition. This means that data entry screens that use these fields will only need to specify when the data should be entered as well as the position and layout of the data field.

5. Conclusion

Longitudinal community health studies are crucial to understanding the determinants of survival and establishing feasible health interventions to control major causes of mortality and morbidity. The software needed for such studies is complex to develop, manage, and modify, however. As a consequence, studies are restricted to a few settings where major resources are available and international staff commitments can be made. Our research demonstrates that recent developments in object-oriented programming and automated program generation can be used to meet the requirements of a diverse array of longitudinal household studies with low-cost and simple-to-use technology that facilitates the development and modification of database system code. The capability of the HRS to structure the generation of robust and extensible household monitoring software fosters low-cost longitudinal surveys. Once developed, a core system can be readily linked with concomitant specialized longitudinal studies, such as cohort studies and panel surveys. Such capacities can expand the scope of health services research since research on most practical policy questions concerns the morbidity and mortality effects which cannot be assessed by cross-sectional surveys. The ease of maintaining and extending the HRS has fostered rapid dissemination and utilization of the HRS to several developing countries [Table 1]. HRS-based research is testing the efficacy of malaria vaccines, assessing HIV epidemiology, researching pneumococcal pneumonia epidemiology and prevention, examining cerebral spinal meningitis sequelae, testing family planning impact, and exploring options for health care reform.

The HRS FoxPro code is available on the Population Council website (http://www.popcouncil.org/hrs/hrs.html) which also offers HRS user and technical manuals, answers to frequently asked questions, and a link to INDEPTH (http://www.indepth-network.org), a technical exchange network of eighteen longitudinal health research groups from Africa and Asia, many of whose sites use HRS-based systems. The current version of the HRS is undergoing revision to improve HRS data manipulation tools, develop a system interface with commercial software packages, and upgrade system efficiency. These improvements seek to further our goal of facilitating longitudinal demographic research in developing country settings.

References

1. Andersson, N. 1993. “Sentinel Community Surveillance (SCS): An Option for Monitoring Progress Towards the Mid-Decade Goals in ‘Data Poor’ Countries.” New York: UNICEF Paper of the Month, unpublished.
2. Balyagati, D., D. Lubamba, S. Nuko, V. Nyonyo, and D. Schapink. 1995. “HIV / AIDS and STD health promotion in Tanzanian fishing villages.” AIDS / STD Health Promotion Exchange 1995(2):3-7.
3. Behar, M., N.S. Scrimshaw, M.A. Guzman, and J.E. Gordon. 1968. “Nutrition and Infection Field Study in Guatemalan Villages, 1959-1964. VIII. An Epidemiological Appraisal of its Wisdom and Errors.” Archives of Environmental Health 17:814-827.
4. Billewicz, W.Z. and I.A. McGregor. 1981. “The Demography of Two West African (Gambian) Villages, 1951-1975.” Journal of Biosocial Science 13(2):219-240.
5. Binka, F., M. Gyapong, F.A. Kubaje, J.F. Phillips, and B. MacLeod. 1995a. “Instructions to Interviewers for the Demographic Data Component of the Navrongo Demographic Surveillance System.” Documentation Note Number 2. Navrongo, Ghana: Navrongo Health Research Centre.
6. Binka, F.N., P. Ngom, J.F. Phillips, K. Adazu, and B.B. MacLeod. 1999. “Assessing Population Dynamics in a Rural African Society: The Navrongo Demographic Surveillance System.” Journal of Biosocial Science 31: 373-391.
7. Byass, P. 1986. “The Design of Computer Forms for Tropical Medical Research.” Methods of Information in Medicine 25:229–232.
8. _______. 1987. “Computers in Africa: Appropriate technology?” Computer Bulletin 3(17).
9. Byass, P., P. Hanlon, L. Marsh, and B.M. Greenwood. 1988. “Microcomputer management of a vaccine trial.” Computers in Biology and Medicine 18(3).
10. Chen, L. 1981. “Child survival: Levels, trends and determinants,” in Determinants of Fertility in Developing Countries: A Summary of Knowledge. Panel on Fertility Determinants, Committee on Population and Demography, National Academy of Sciences.
11. Chen, L., E. Huq, and S. D’Souza. 1981. “Sex Bias in the Family Allocation of Food and Health Care in Rural Bangladesh.” Population and Development Review 7:55–70.

12. Chen, L., M. Rahman, S. D’Souza, J. Chakraborty, A.M. Sardar, and M.D. Yunus. 1983. “Mortality Impact of an MCH-FP program in Matlab, Bangladesh.” Studies in Family Planning 14:199–209.

13. Comstock, G.W. 1964. “Community Research in Tuberculosis, Muscogee County, Georgia.” Public Health Report 79:1045-1056.

14. Defo, B.K. 1992. “Mortality and Attrition Processes in Longitudinal Studies in Africa: An Appraisal of the IFORD Surveys.” Population Studies 46:327–348.

15. Densen, P.M. 1940. “Family Studies in the Eastern Health District, II. The Accuracy of Statements of Age on Census Records.” American Journal of Hygiene 32:1-38.

16. D’Souza, S. 1981. “A Population Laboratory for Studying Disease Processes and Mortality -- the Demographic Surveillance System, Matlab Comilla, Bangladesh.” Rural Demography 8(1):29-51.

17. _______. 1982. “Mortality Case Study: Matlab, Bangladesh.” Paper prepared for the United Nations/World Health Organization, Third Project Collaborators’ Meeting on Case Studies of Determinants of Mortality Change and Differentials, Geneva.

18. _______. 1984. “Population Laboratories for Studying Disease Processes and Mortality. The Demographic Surveillance System, Matlab,” in Methodologies for the Collection and Analysis of Mortality Data, edited by J. Vallin, J.H. Pollard, and L. Heligman. Proceedings of a Seminar at Dakar, Senegal, July 17-10. Liège: Ordina Editions.

19. D’Souza, S. and L. Chen. 1980. “Sex Differentials in Mortality in Rural Bangladesh.” Population and Development Review 8:753–769.

20. D’Souza, S. and A. Bhuiya. 1982. “Socioeconomic Mortality Differentials in a Rural Area of Bangladesh.” Population and Development Review 6:257–270.

21. Fauveau, V. 1991. “Behavioral Aspects of Health Intervention Programmes: A Case-Study in Matlab, Bangladesh.” Paper presented at the International Union for the Scientific Study of Population (IUSSP) Committee on Anthropological Demography and ORSTOM Seminar on Socio-Cultural Determinants of Morbidity and Mortality in Developing Countries: The Role of Longitudinal Studies, Saly Portudal, Senegal, October 7-11.

22. Ferebee, S.H. and F.W. Mount. 1962. “Tuberculosis Morbidity in a Controlled Trial of the Prophylactic Use of Isoniazid Among Household Contacts.” American Review of Respiratory Diseases 85:495-521.

23. Goldberger, J., G.A. Wheeler, and E. Sydenstricker. 1920a. “A Study of the Relation of Diet to Pellagra Incidence in Seven Textile Mill Communities of South Carolina in 1916.” Public Health Reports 35:648-713.

24. _______. 1920b. “Pellagra Incidence in Relation to Sex, Age, Season, Occupation, and Disablig Sickness in Seven Cotton-Mill Villages of South Carolina in 1916.” Public Health Reports 35:1650-1664.

25. _______. 1920c. “A Study of the Relation of Factors of a Sanitary Character to Pellagra Incidence in Seven Textile Mill Communities of South Carolina during 1916.” Public Health Reports 35:1701-1714.

26. _______. 1920d. “A Study of the Relation of Family Income and Other Economic Factors to Pellagra Incidence in Seven Cotton-Mill Villages of South Carolina during 1916.” Public Health Reports 35:2673-2714.

27. Gordon, J.E., W. Ascoli, L.J. Mata, M.A. Guzman, and N.S. Scrimshaw. 1968. “Nutrition and Infection Field Study in Guatemalan Villages, 1959-1964.” Archives of Environmental Health 16:424-437.

28. Guzman, M.A., N.S. Scrimshaw, H.A. Bruch, and J.E. Gordon. 1968. “Nutrition and Infection Field Study in Guatemalan Villages, 1959-1964. VII. Physical Growth and Development of Preschool Children.” Archives of Environmental Health 17:107-118.

29. Indome, F., B. Macleod, J. Phillips, F.K. Adazu, and F. Binka. 1995. “Computational Procedures for the Navrongo Demographic Surveillance System Rates Program.” Documentation Note Number 26 (May).

30. Jarrett, R.J., M.J. Shipley, and G. Rose. “Weight and mortality in the Whitehall study.” British Medical Journal 285(6341): 535-537.

31. Kielman, A.A., C.E. Taylor, C. DeSweemer, R.L. Parker, D. Chernichovsky, W.A. Reinke, I.S. Uberoi, D.N. Kakar, N. Masih, and R.S.S. Sarma. 1983. Child and Maternal Health Services in Rural India: The Narangwal Experiment. Vol. 2: Integrated Nutrition and Health Care. Baltimore: The Johns Hopkins University Press.

32. Leon, D. 1986a. “SRS Version 1.1 User's Manual: The Microcomputer Software Component of the Sample Registration System of the MCH-FP Extension Project.” Dhaka: International Centre for Diarrhoeal Disease Research, Bangladesh; Bangkok: Bangkok Regional Office for South and East Asia, The Population Council.

33. _______. 1986b. “SRS Version 1.1 Technical Manual: The Microcomputer Software Component of the Sample Registration System of the MCH-FP Extension Project.” Dhaka: International Centre for Diarrhoeal Disease Research, Bangladesh; Bangkok: Regional Office for South and East Asia, The Population Council.

34. _______. 1987. “The Microcomputer Software Concept of the Sample Registration System: Demonstration Manual.” Unpublished. Bangkok: The Population Council Regional Office for South and East Asia.

35. Linder, F.E. 1971. “The Concept and the Program of the Laboratories for Population Statistics.” Laboratories for Population Statistics Scientific Series No. 1, March.

36. Luykx, H.M.C. 1947. “Family Studies in the Eastern Health District. IV. Permanence of Residence with Respect to Various Family Characteristics.” Human Biology 19(1).

37. MacLeod, B., D. Leon, and J.F. Phillips. 1992. "The Household Registration System: A Database Software Generator for Longitudinal Studies of Health Status and Demographic Dynamics," Computing in the Social Sciences 10:310-328.

38. MacLeod, B.B., J. Phillips, and F. Binka. 1996. “Sustainable Software Technology Transfer: The Household Registration System.” Pp. 302-310 in Encyclopedia of Library and Information Science, Vol. 58, Supplement 21, edited by K. Allen. New York: Marcel Dekker, Inc.

39. Marmot M. “Social differences in mortality: The Whitehall studies.” In Lopez, A.D., G. Caselli, and T. Valkonen, Eds., Adult Mortality in Developed Countries: From Description to Explanation. Clarendon Press: Oxford, England.

40. Menken, J. and J.F. Phillips. 1990. “Population Change in a Rural Area of Bangladesh: 1967-1987.” The Annals of the American Academy of Political and Social Science 510:87–101.

41. Mosley, W.H. 1988. "Community Based Health Research." Paper presented at the seminar "Community Based Research in Mexico: A Proposal for the Initial Stage," Institute for Public Health, Mexico.

42. Mozumder, K.A. 1983. "Coding Plan for Vital Event and Other SRS Event Reporting Forms." MCH-FP Extension Project Documentation Note (14). Dhaka: International Centre for Diarhhoeal Disease Research, Bangladesh.

43. Mozumder, K.A., U. Rob, and J.F. Phillips. 1983b. "English Language Revised Field Instructions to Interviewers for the Demographic Data Component of the Sample Registration

System." MCH-FP Extension Project Documentation Note (10). Dhaka: International Centre for Diarhhoeal Disease Research, Bangladesh.

44. Mozumder, K.A., M.A. Koenig, J.F. Phillips, and S. Murad. 1990. "The Sample Registration System: An Innovative System for Monitoring Demographic Dynamics." Asia-Pacific Journal 5: 63–72.

45. Muller, A.S. and J.K. van Ginneken. 1991. “Morbidity and Mortality in Machakos, Kenya 1974-81.” In Disease and Mortality in Sub-Saharan Africa, edited by R.G. Feachem and D.T. Jamison. Washington, DC: The World Bank and Oxford University Press.

46. Napier, J.A., H.L. Metzner, and B.C. Johnson. 1970. “Limitations of Morbidity and Mortality Data Obtained from Family Histories: A Report form the Tecumseh Community Health Study.” Paper presented at the 98th Annual Meeting of the American Public Health Association, Houston, October.

47. Notestein, F. 1963. “India: The Singur Study.” Studies in Family Planning 1(1):1-9.

48. Nuko, S. and R. Pool. 1997. “Sexual discourse in the context of AIDS: Dominant themes on adolescent sexuality among primary school pupils in Mago District, Tanzania.” Health Transition Review 7 Suppl. 3: 85-90.

49. Palmer, C.E., L.W. Shaw, and G.W. Comstock. 1958. “Community Trials of BCG Vaccination.” American Review of Tuberculosis 77:877-907.

50. Phillips, J.F., R. Simmons, M.A. Koenig, and J. Chakraborty. 1988. “The Determinants of Reproductive Change in a Traditional Society: Evidence from Matlab, Bangladesh,” Studies in Family Planning 19:313–334.

51. Rahman, M. and S. D'Souza. 1981. "A Review of Findings on the Impact of Health Intervention Programmes in Two Rural Areas of Bangladesh," in United Nations Economic Commission for Asia and the Pacific and World Health Organization, Mortality in South and East Asia: A Review of Changing Trends and Patterns 1050-75; A Report and Selected Papers, presented at a joint WHO/ESCAP meeting, Manila (December 1-5). Geneva, World Health Organization.

52. Rao, M.N. and K.K. Mathen. 1970. Rural Field Study of Population Control, Singur, 1957-1969. Calcutta, India: All-India Institute of Hygiene and Public Health.

53. Rider, R.V. and G.F. Badger. 1943. “Family Studies in the Eastern Health District. III. A consideration of Issues Involved in Determining Migration Rates for Families.” Human Biology 15:12.

54. Scrimshaw, N.S. 1970. Ecological Assessment of the Nutritional Status of a Guatemalan Finca Population. Unpublished INCAP study.

55. Scrimshaw, N.S., M.A. Guzman, M. Flores, and J.E. Gordon. 1968. "Nutrition and Infection Field Study in Guatemalan Villages, 1959-1964. V. Disease Incidence Among Preschool Children Under Natural Village Conditions with Improved Diet and With Medical and Public Health Services." Archives of Environmental Health 16(2):233-234.

56. Sydenstricker, E. 1926. "A Study of Illness in a General Population Group: Hagerstown Morbidity Studies Number 1: The Method of Study and General Results." Public Health Reports 24:2069-2088.

57. Tanner, M. 1988. "District Level Data Collection and Use." Paper prepared for the Independent Commission on Health Research for Development, Harvard University School of Public Health. Cambridge: MA.

58. Tanner, M. and D. de Savigny. 1987. "Monitoring of Community Health Status: Experience from a Case Study in Tanzania." Acta Tropica 44:261–270.

59. Taylor, C.E., R.S.,S. Sarma, R.L. Parker, W.A. Reinke, and R. Faruqee. 1983. Child and Maternal Health Services in Rural India: The Narangwal Experiment. Vol. 2: Integrated Family Planning and Health Care. Baltimore: The Johns Hopkins University Press.

60. Terris, M. (ed.) 1964. Goldberger on Pellagra. Baton Rouge, Louisiana: Baton Rouge University Press

61. Utomo, B., C. Costello, J.F. Phillips, T. Budiono, and G. DasVarma. 1990. "A General Description of the Indramayu Health and Family Planning Prospective Study." Unpublished report. West Java: Center for Child Survival, University of Indonesia.

62. van Ginnneken, J.K. and A.S. Muller. 1984. Maternal and Child Health in Rural Kenya: An Epidemiological Study. London: Croon Helm.

63. Wyon, J.B. and J.E. Gordon. 1971. The Khana Study: Population Problems in the Rural Punjab. Cambridge, Massachusetts: Harvard University Press.

Notes

1. Systematic research on this computing problem is lacking. Few papers on the computer science challenges of longitudinal demographic studies exist in the scientific literature despite frequent references to computer problems that have posed a major constraint on research [45,62]. More typically, however, research designs and collection designs are published without detailed discussion of issues in computer systems design [57,58]. Some analysts have provided useful presentations of problems [7-9] although tractable solutions remain the subject of investigation.
2. Most field stations utilize some form of demographic surveillance. Demographic surveillance, however, is sometimes used in cluster samples when longitudinal studies require dispersed national or regional samples [1,44].
3. See the review by Terris [60] of the classical studies by Goldberger, Wheeler, and Sydenstricker [23-26].
4. There are several examples of this type of epidemiological investigation in the early public health literature, although the most thoroughly documented are the Hagerstown study and the East Health District investigation [15,36,53,56]. Field stations have played an important role in research on the epidemiology of chronic disease [46].
5. For example, the Muskogee County study in Georgia was a longitudinal investigation of tuberculosis that was eventually used to research the efficacy of BCG vaccine and other tuberculosis control modalities [13,22,49].
6. A summary of the Singur project is in Notestein [47] and in Rao and Mathen [52]. Singur was the first factorial demonstration of the impact of family planning service delivery on fertility. To date, only three projects have demonstrated demographic impact: Singur, Matlab, and Danfa. Both the Khana and Singur initiatives involved family planning only, and were constrained to offering foam tablets and information about the rhythm method. The Singur project was the first experiment to demonstrate that family planning service delivery can have an impact on fertility. The Khana experiment failed to confirm this impact, however. In 1968, a study was launched in 26 villages of Narangwal, Punjab to test the efficacy of integrating health services with family planning. Results were subject to considerable debate, and fertility impact was not definitively demonstrated [31,59].
7. Over a thousand papers are based on Matlab investigations and trials. Useful overviews of the Matlab story include [10-12,17-21,40,50].

8. A useful review of strengths and pitfalls of longitudinal community and household cohort studies appears in Mosley [41]. See also Defo [14].

9. Notable exceptions exist of cohort studies that serve multiple research protocols and whose past data contribute to new studies (for example, the Whitehall and TANESA studies [2,30,39,48]).

10. Two versions of the HRS have been developed. HRS1 is a DOS-based system written in MicroSoft FoxPro 2.1. In January 1999 the HRS2 became available. This is a new MS-Windows version which improves on the HRS1 in a number of significant ways. The system is WINDOWS-95, -98, and -NT compatible. HRS2 has a more user-friendly data entry screen design than HRS1. The layout of the data entry screen resembles the paper forms used for data recording. Consistency logic can be attached to the database table allowing easier project-specific changes and limiting the possibility of inserting data without proper logical checks. Core relationship information now allows for extra-familial relationships, such as social networks, kinship structure, or economic associations. The HRS system permits entry of data on "externally defined" individuals. This, in turn, permits research on ways in which events to nonmembers of households affect risk among members. Also, new HRS2 "external" individual features permit periodic registration of individuals temporarily present in households. This information is sometimes required for prevalence studies. Procedures for recording pregnancy outcomes have been refined so that information about births is registered, not just information about mothers and pregnancies. This permits research requiring data on births (such as birth weight) or data on the characteristics of multiple births. HRS2 is designed to support a number of different ID formats with little to no changes. The over-arching aim of these system changes is to develop a single model of surveillance that can accommodate almost all longitudinal surveillance operations. By incorporating features of systems that enhance versatility, the HRS2 works for any specified interviewing interval, permits studies on social groups other than households, facilitates detailed specification of household relationships, and allows for complex extended family structures. These features permit health research on social characteristics of families, networks, kinship structures, and other complex relationships.

11. When the clerk types in the household ID, the first screen to appear is the list of household members. In order to assign an event to an individual, the clerk must select that person from the list of members. If the clerk miskeys the household number but the mistake corresponds to a valid ID for another household, that household’s member list will appear. At this point the clerk’s error should be apparent (the member list on the screen will not match the list on the paper register from which s/he is typing), and s/he can “revert” (back out of data entry mode for

that record) without leaving a permanent trace. Only after an individual has been selected and an event button has been clicked is a permanent visitation record associated with the household for that round.

12. Quite detailed field manuals have been developed for the Bangladesh Sample Registration System, a version of the HRS that operates in a dispersed cluster sample of rural households [32-34,42,43]. Using the Bangladesh system as a model, corresponding field manuals have been developed for the Indonesian system, located in Indramayu Regency of West Java [61]. The Navrongo field manual is documented in Binka et al. [5,6].

Tables

Table 1:
Applications of the HRS in developing countries

Protocol or project	Institution	Country	Substantive focus
PRAPASS	University of Heidelberg, Ministry of Health, Nouna Health Research Center	Burkina Faso	Acute respiratory infection, child survival
The Farafeni Project	British Medical Research Council and Harvard University	The Gambia	Reproductive health, malaria, nutrition, and immunizable diseases.
Community Health and Family Planning Project Perethritin Impregnated Bednet Trial Etaquine trial MuStDO-15 Vaccine Trial	Navrongo Health Research Centre	Ghana	Community medicine, family planning and reproductive health; randomized quasi-experiment. Randomized factorial trial Malaria prophylaxis; randomized trial Vaccine trial
The Gada Madja Child Survival Project	Gada Madja University	Indonesia	Health systems research
Kisumu	Kenya Medical Research Institute	Kenya	Malaria, HIV
The Nairobi Urban Health and Poverty Project*	Africa Population and Health Research Centre	Kenya	Urban poverty and health research.
The Health and Child Survival Fellows Program, Kolondieba	Centre d'Etudes et de Recherche sur la Population pour le Developpement	Mali	Child survival and reproductive health
The Spf66 Malaria Vaccine Trial	Universidad de Barcelona	Mozam-bique	Vaccine against malaria; randomized trial
Tanzania Essential Health Intervention Project	Ifakara Centre for Health Research and Development, Ministry of Health	Tanzania	Spf66 malaria vaccine trial, health systems research
The Rakai Project	Makerre University and Johns Hopkins University	Uganda	Reproductive health

* Pilot trial scheduled for 2000

Figures

Figure 1:
HRS Data Collection and Management Cycle

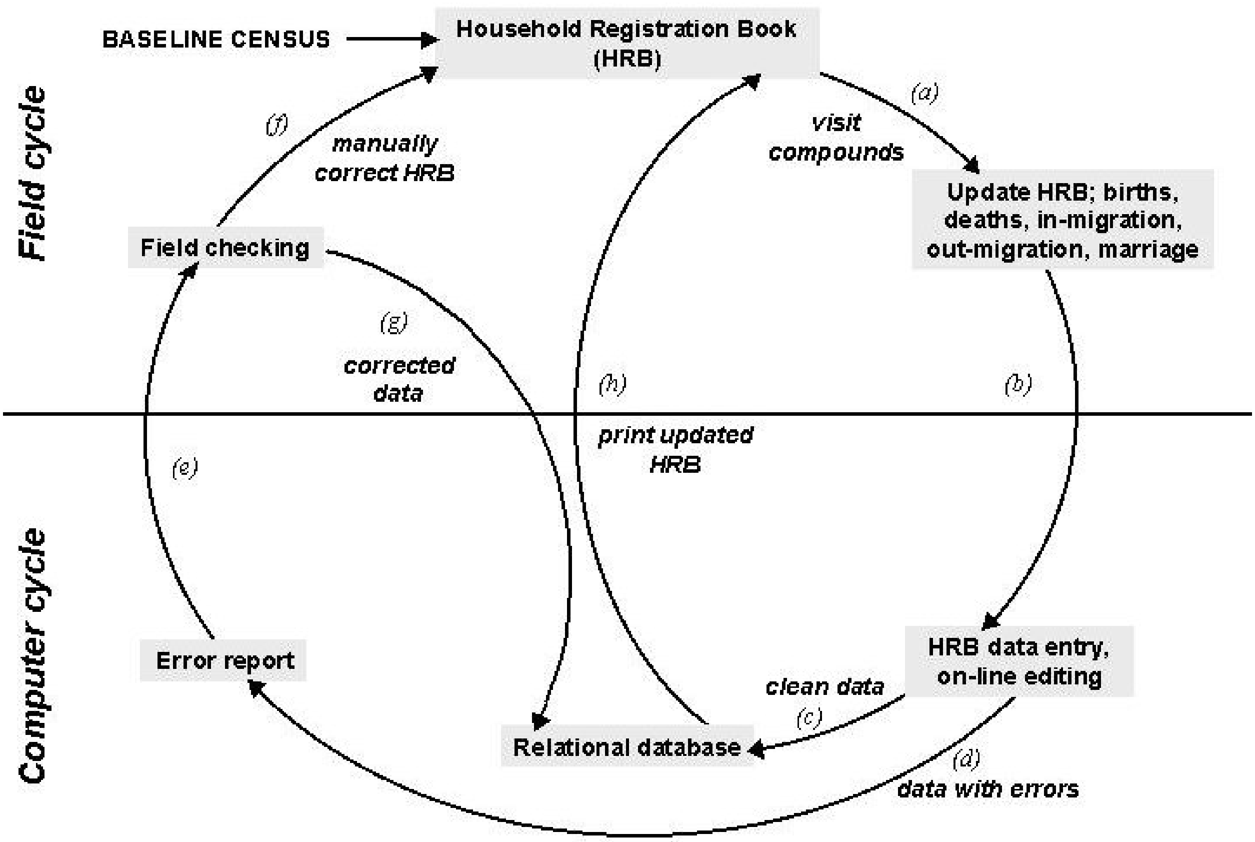


Figure 2:
Main Menu

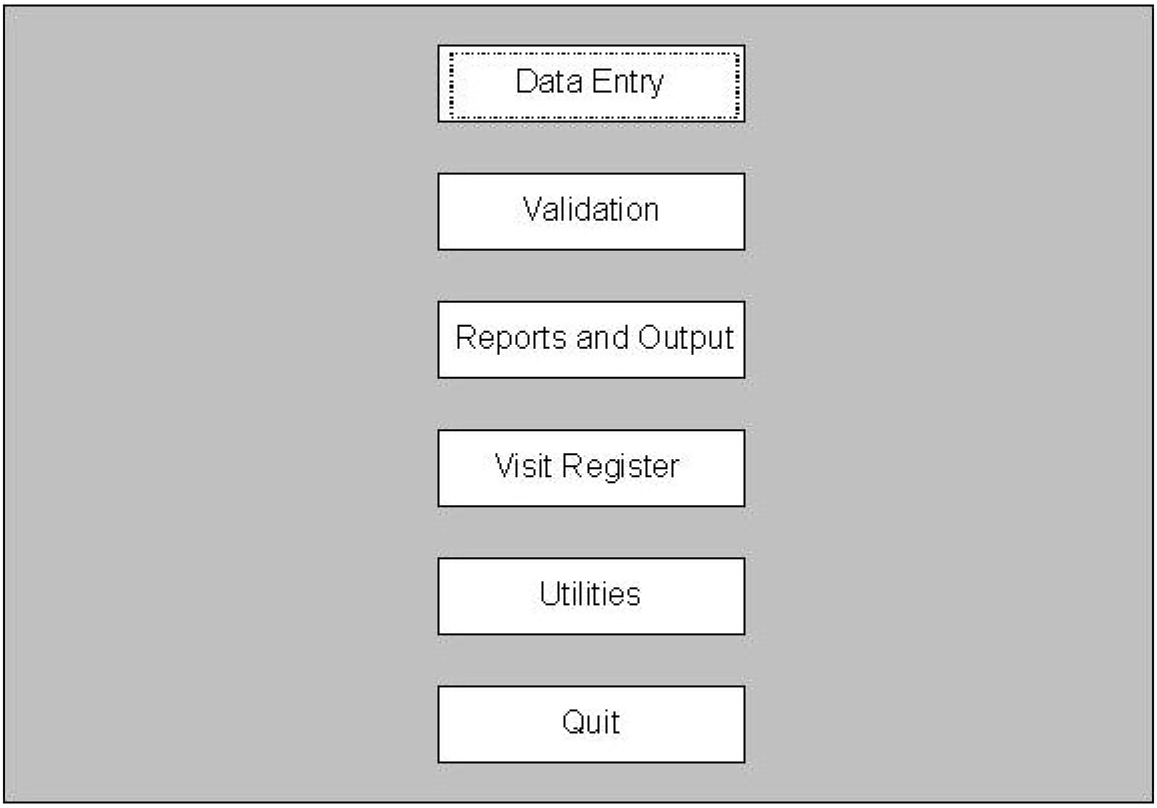
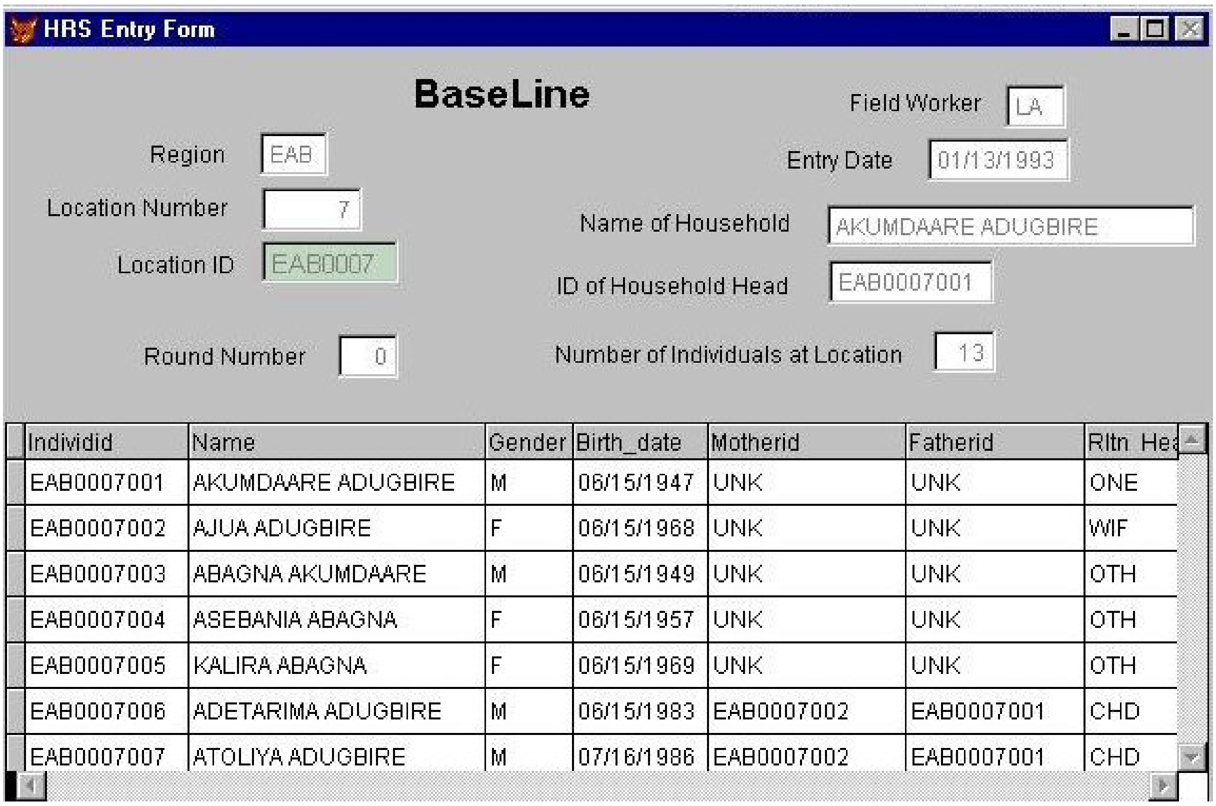


Figure 3:
Baseline



HRS Entry Form

BaseLine

Field Worker: LA
Region: EAB
Entry Date: 01/13/1993
Location Number: 7
Name of Household: AKUMDAARE ADUGBIRE
Location ID: EAB0007
ID of Household Head: EAB0007001
Round Number: 0
Number of Individuals at Location: 13

Individid	Name	Gender	Birth_date	Motherid	Fatherid	Rltn_Hea
EAB0007001	AKUMDAARE ADUGBIRE	M	06/15/1947	UNK	UNK	ONE
EAB0007002	AJUA ADUGBIRE	F	06/15/1968	UNK	UNK	WIF
EAB0007003	ABAGNA AKUMDAARE	M	06/15/1949	UNK	UNK	OTH
EAB0007004	ASEBANIA ABAGNA	F	06/15/1957	UNK	UNK	OTH
EAB0007005	KALIRA ABAGNA	F	06/15/1969	UNK	UNK	OTH
EAB0007006	ADETARIMA ADUGBIRE	M	06/15/1983	EAB0007002	EAB0007001	CHD
EAB0007007	ATOLIYA ADUGBIRE	M	07/16/1986	EAB0007002	EAB0007001	CHD

Figure 4:
Updates

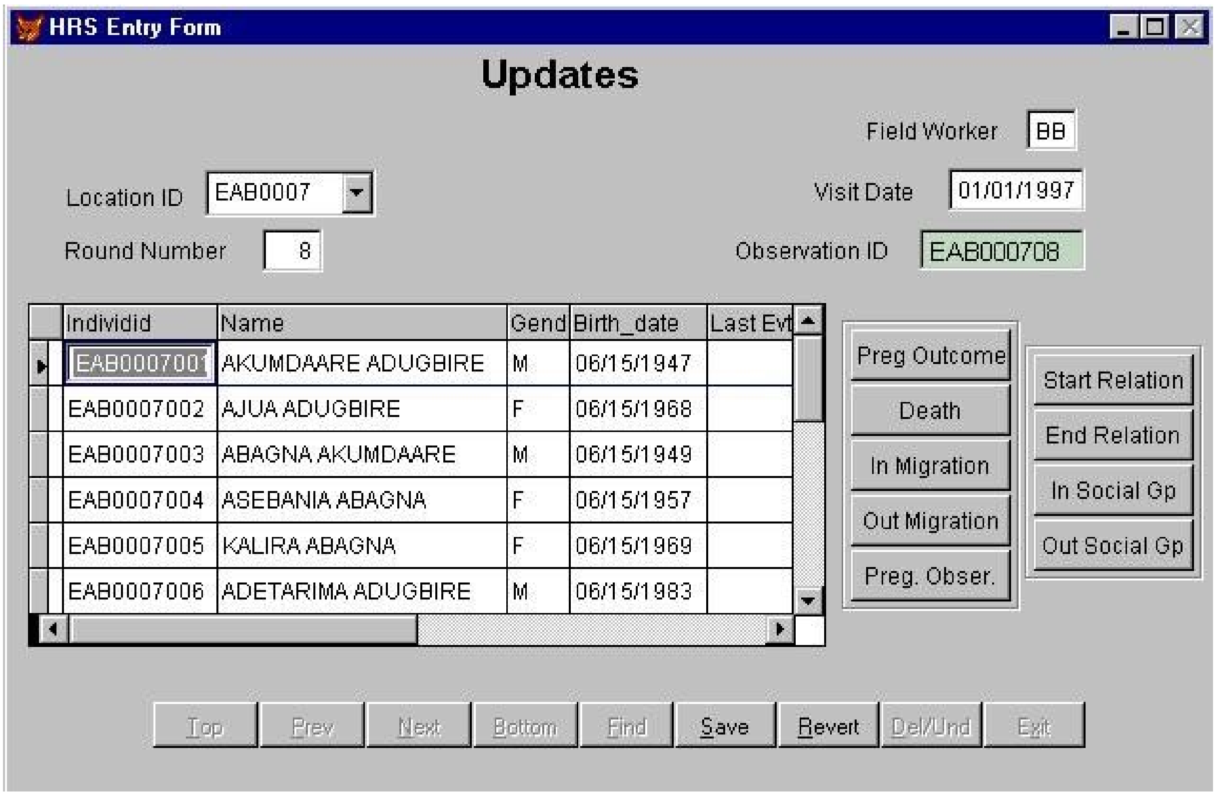


Figure 5:
Pregnancy Outcome

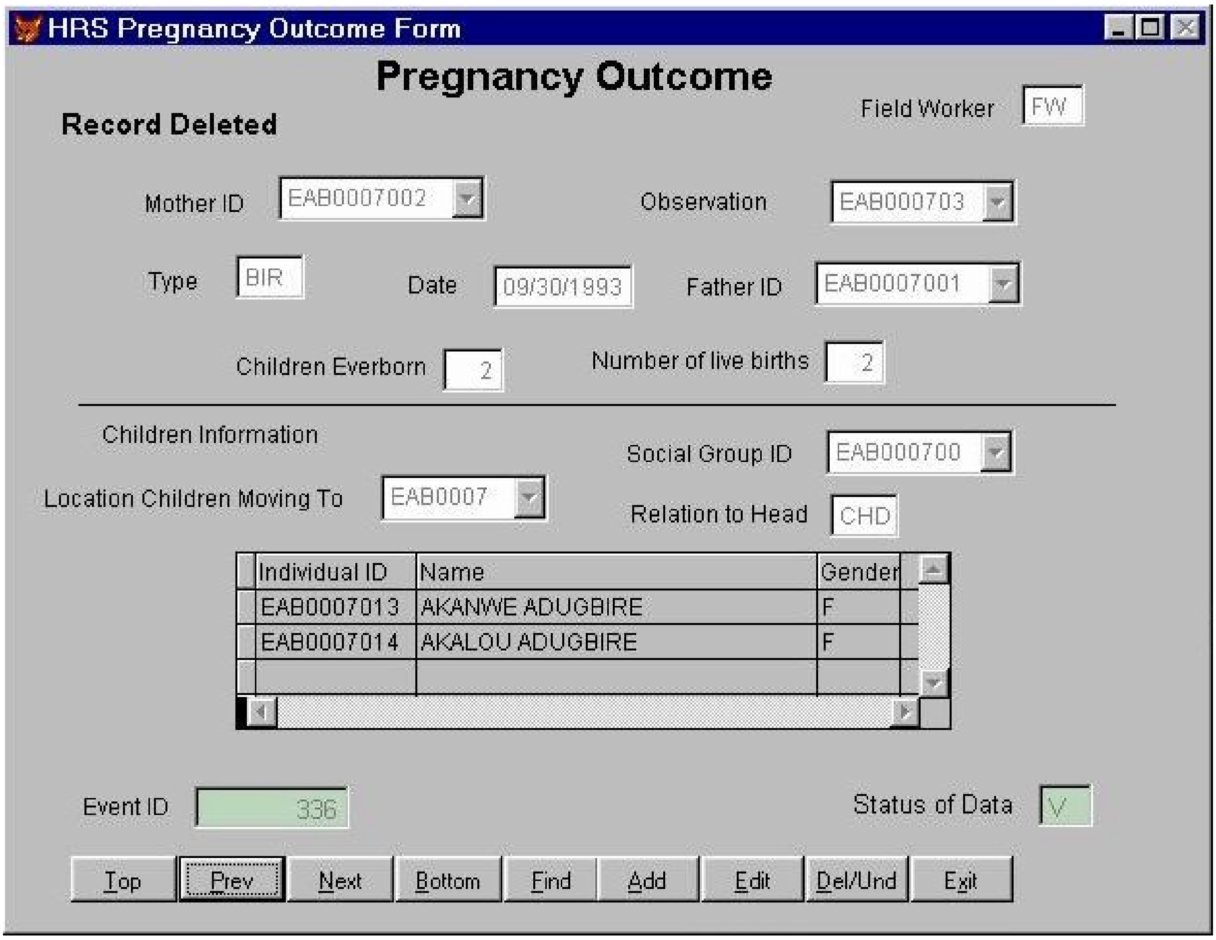


Figure 6:
Update Specification Form

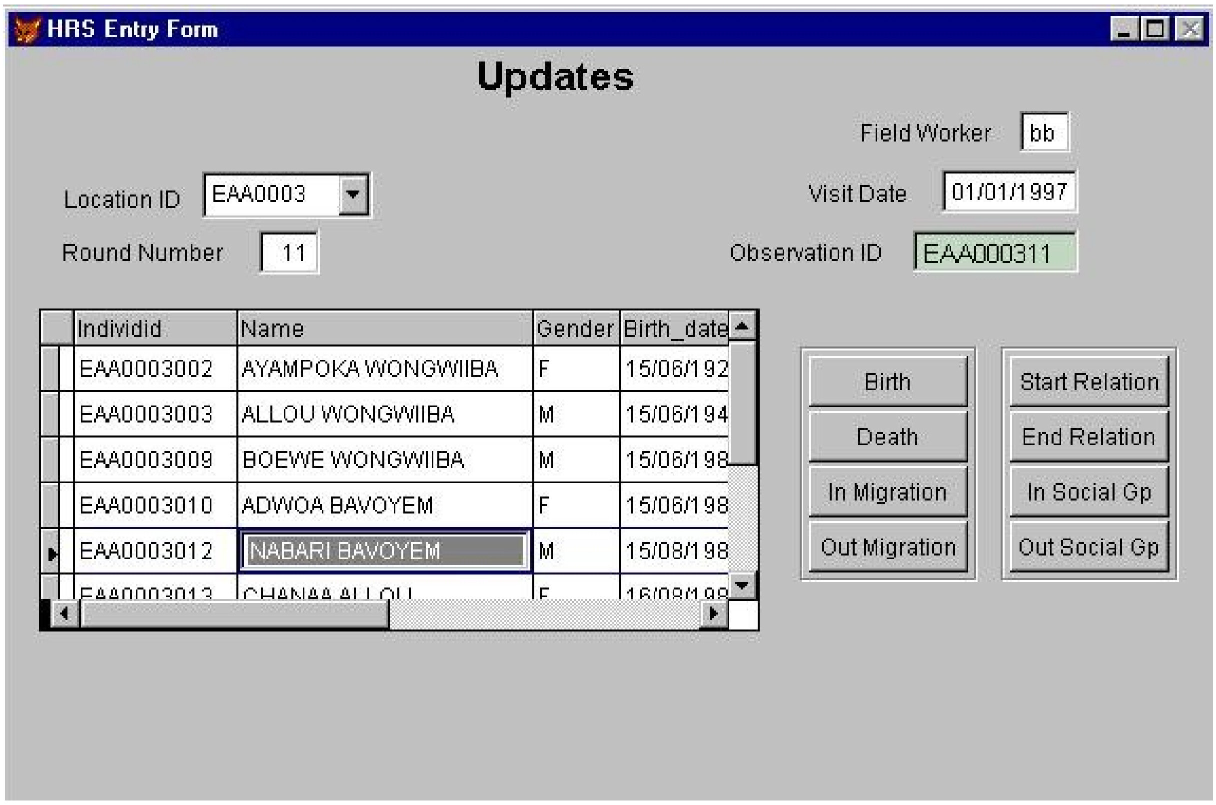


Figure 7:
Properties

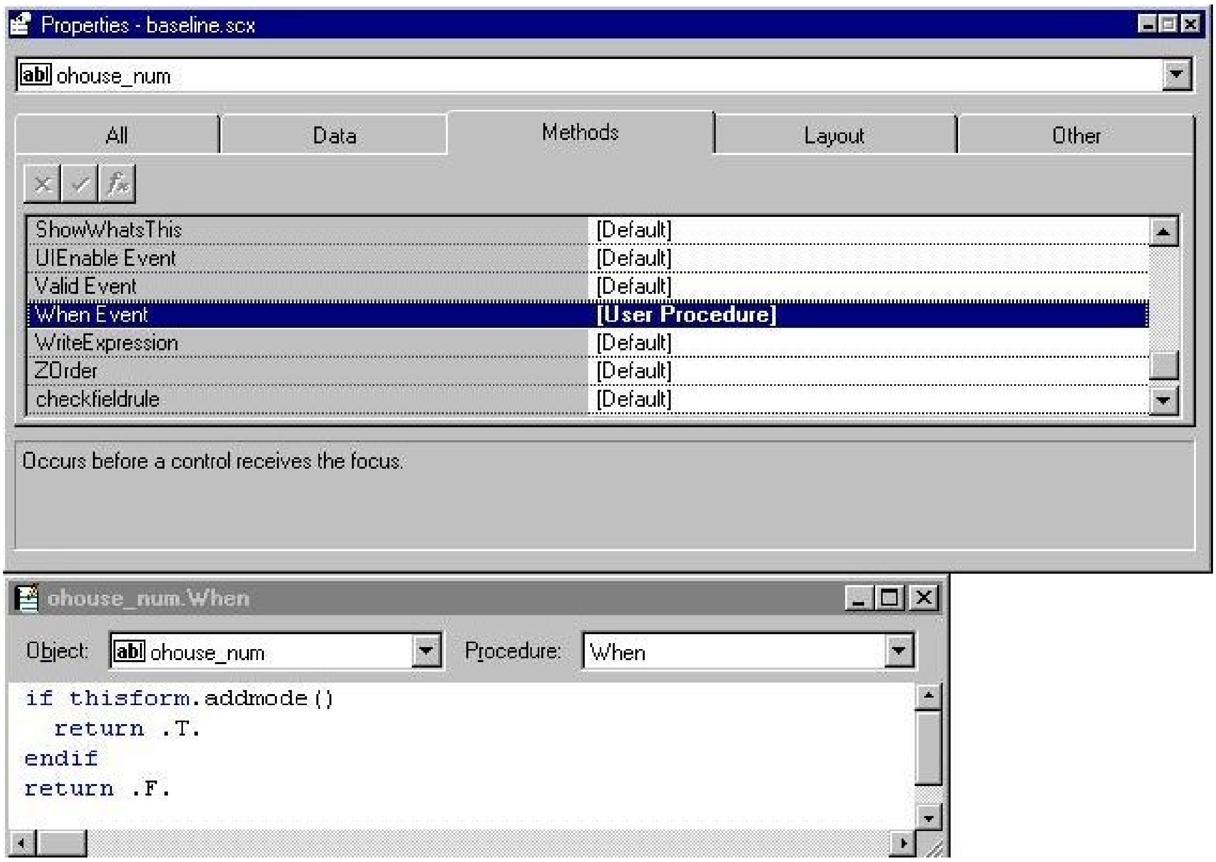


Figure 8:
Table Designer