DESIGN OF A SURVEILLANCE SYSTEM FOR PREGNANCY AND ITS OUTCOMES IN RURAL NEPAL

Osrin D¹, Manandhar A S², Shrestha A², Mesko N¹
Tumbahangphe K M², Shrestha D², Manandhar D S², Costello A M¹

ABSTRACT

Introduction

Community trials in low-income countries require monitoring and evaluation systems. The requirements of a community surveillance system include coherent design, training, field supervision and reporting, as well as the need for a robust and flexible database.

Materials and methods

This paper describes a surveillance system for identification of pregnancy and its outcomes in a rural area of Nepal. Mother Infant Research Activities (MIRA), in collaboration with the Institute of Child Health, London, are presently conducting a study on the impact of a community-based participatory intervention to improve essential newborn care (ENC) in rural Nepal. The study is a cluster randomised controlled trial involving 12 pairs of Village Development Committees (VDCs) in Makwanpur District. The surveillance system covers approximately 28 000 households and 28 000 married women of reproductive age. It was designed to identify pregnancy, its outcome for mother and infant, and activities such as antenatal care and problem-related health care seeking behaviour.

Discussion

The paper describes the processes of mapping and enumeration, pregnancy identification, conduct of interviews, quality control and data management.

Key Words: *data quality, database, rural Nepal.*

1. Centre for International Child Health, Institute of Child Health, University of London.
2. Mother Infant Research Activities, Kathmandu, Nepal.

Address for correspondence : Dr. David Osrin
Mother Infant Research Activities (MIRA)
GPO Box: 921, Kathmandu, Nepal.
Tel: 266208, Email: dsm@healthnet.org.np

INTRODUCTION

The current emphasis on controlled trials for assessing the impact of health interventions has led to an increased interest in conducting trials at community level. In low-income countries, major decisions about health care occur in the home. Any attempt to improve health should address the decision chain if it is to be effective and sustainable.^{1,2} In order to evaluate potential interventions, therefore, it may be necessary to operate a monitoring system at community level. Limited published information exists to help in the design of such a community surveillance system, and this paper attempts to address some of the practicalities on the basis of recent experience in Nepal.

Mother Infant Research Activities (MIRA), in collaboration with the Institute of Child Health, London, are presently conducting a study on the impact of a community-based participatory intervention to improve essential newborn care (ENC) in rural Nepal. The study is a cluster randomised controlled trial involving 12 pairs of Village Development Committees (VDCs) in Makwanpur District.

SETTING

Makwanpur District occupies about 2500 sq km and lies to the south of Kathmandu in the Central Region, Narayani Zone. It has a population of 376 342 (1998 projection from 1991 census). The district has a Human Development Index of 0.309 and a Gender Sensitive Development Index of 0.231 and includes both hill and terai areas. The level of literacy is about 34% (all figures from ³). The ethnic composition of the district is mixed: District Development Committee records suggest that the largest group are Tamang (46%), followed by Brahmin/Chhetri (25%), Newar (8%), and Magar (5%). The major occupation is agriculture,

in which the great majority of residents are engaged. The district centre is Hetauda, the site of the MIRA field office. Hetauda is a growing town with good road connections and reasonably stable electricity and telephone provision.

DESIGN OF THE STUDY

The study involves five major areas of activity.

1. The intervention: this is a participatory intervention based on a model used in Bolivia.⁴ Briefly, 12 VDCs will receive the intervention over three years. 12 matched VDCs will serve as controls during this period, after which they will receive the intervention in a form that will have been modified on the basis of the findings of the initial study. A VDC Facilitator is placed in each intervention VDC with a brief to work with women's groups in an action research cycle to address problems in pregnancy and care of the newborn infant. The VDC Facilitator is a local woman who will attempt to broker change in care practices. Facilitators are supported by a tier of Facilitation Supervisors, who are in turn supported by a Senior Facilitation Officer.
2. Health service strengthening: this is carried out across all 24 VDCs in the study area. It comprises audit, training of personnel and supply of essential equipment and medications for care of the newborn infant.
3. Administration: this team provides financial and logistic support for a study that involves over 300 personnel.
4. Surveillance: this team functions independently of the facilitation team. It conducts surveillance of married women of reproductive age (MWRA) for pregnancy and related outcomes across all 24 VDCs.
5. Data management: this team processes the information provided by the surveillance team, manages the database and provides feedback.

The following discussion covers the activities of the surveillance and data management teams.

OBJECTIVES OF THE SURVEILLANCE SYSTEM

The effects of the intervention will be judged against a range of outputs; the central outputs are perinatal and neonatal mortality rates, on which the sample size has been based ⁵. The number of households in the 24 VDCs is about 28 000. The number of MWRA within these households is similar, with an average of one per household. The estimated number of pregnancies to be enrolled is about 6000 per year ⁶.

The requirements of the surveillance system are as follows:

- ? To locate by number all households in the study area.
- ? To identify by number all MWRA in these households.
- ? To enrol new pregnancies as they arise.
- ? To administer interview tools to women during pregnancy.
- ? To determine the outcome of pregnancy for both mother and infant.
- ? To administer interview tools or verbal autopsies in the postnatal period.

Logistically, this list of requirements gives rise to a secondary list of activities:

- ? To map and number all households in the study area.
- ? To identify and allocate a number to each MWRA in the study area.
- ? To visit each MWRA regularly to assess the possibility of pregnancy.
- ? To record pregnancy as it arises and communicate this to the appropriate team.
- ? To achieve a contact with each MWRA at seven months of pregnancy and administer a questionnaire.

? To achieve a contact with each MWRA at one month post-partum and administer a questionnaire.

? To monitor all information for accuracy.

? To enter all information into an electronic database.

? To store all information in a retrievable form.

SURVEILLANCE ACTIVITIES

Mapping and baseline information

Many of the procedures that underpin the surveillance system were developed after consultation with the Nepal Nutrition Intervention Project, Sarlahi (NNIPS). This group have been carrying out high quality surveillance in the Terai for over a decade, and have unsurpassed experience in the relevant areas.⁷ The surveillance covers 24 VDCs, each of which is divided into nine wards. Each ward is further divided (for study purposes only) into 4 sectors. The study area was mapped ward by ward by a team of 91 local staff, managed by nine Surveillance Supervisors and one Senior Surveillance Officer. During mapping, each household was allocated a unique identification number. The number was a concatenation of VDC, ward, sector and household identifiers. At the same time, a baseline questionnaire was administered to the head of each household (or proxy if not present). The questionnaire contained a range of questions that covered ethnicity, occupational and socioeconomic status, designed to help in (a) describing the backgrounds of MWRA within a household, and (b) assessing baseline comparability of matched pairs of VDCs. All women residing in the household were noted and subsequently classified to identify MWRA. A household was defined as a group of individuals sharing one kitchen; a MWRA as a married woman aged between 15 and 49 completed years, whose husband was alive on enrollment and whose husband either

lived with her or made visits that allowed for the possibility of conception within the study period.

After the mapping and enumeration visit, a second baseline contact was made. At this contact, MWRA were interviewed individually to document their educational status, work activities and details of previous pregnancies. The information provided at this stage would make it possible to determine

retrospective infant mortality rates, patterns of antenatal care, rates of morbidity during pregnancy and postpartum and patterns of health care seeking for maternal and neonatal problems.

Pregnancy surveillance

The structure of the pregnancy surveillance system is summarised in Figure 1. After error checking

and data entry, a list of all MWRA in the study area was generated, broken down by VDC, ward and sector. A VDC-specific section of this list was given to each of 24 VDC Interviewers. Each VDC Interviewer is in turn responsible for managing nine Ward Enumerators, one for each ward of each VDC. Each Ward Enumerator is a local woman whose role is solely to monitor menstrual status in MWRA in her ward. Every week, a Ward Enumerator visits all participants in one sector of her nominated ward. Each sector (and thus each MWRA) is therefore visited once per month. At each visit, a participant is asked about her menstrual status. The Ward Enumerator records the response on a preprinted chart on which the names and numbers of her allotted MWRA appear. The chart is then scrutinised by the VDC Interviewer at a weekly VDC-level meeting. Unless there are other known reasons for amenorrhoea, a participant is classified as pregnant when she has not menstruated for three months. A memo to this effect is written on a standardised form, which is in turn transmitted to the central office. Since the first interview is carried out at about seven months' gestation, there remain four months in which pregnancies that are missed can be identified before an interview is necessary. A second corollary is that resumption of menstruation after a three month gap is classified as a miscarriage.

First interview

The first formal study interview is carried out by the VDC Interviewer at as near to seven months gestation as possible. At seven months, there are three possible outcomes for a pregnant woman: she may be progressing through pregnancy, she may have had a miscarriage or she may have died. Clearly, the first of these outcomes is by far the most likely. Since the Ward Enumerator continues to visit each woman on a monthly basis, it is

possible to know in advance (in the majority of cases) which of these three outcomes is relevant. The Ward Enumerator appraises the VDC Interviewer of the situation, and the first interview is taken. There is a specific questionnaire for each of the three possible outcomes.

Second interview

The second interview is carried out by the VDC Interviewer at as near to one month post-partum as possible. At one month post-partum, there are two possible outcomes for a woman: she may have delivered her baby and progressed to one month, or she may have died. There are likewise three possible outcomes for her infant: progress to one month of age, stillbirth, or neonatal death. The Ward Enumerator again appraises the VDC Interviewer of the situation, and the second interview is taken. There is a specific questionnaire for each of the two possible maternal and each of the three possible infant outcomes.

By the end of the process, the following questionnaires have been completed: a household questionnaire that provides information on socioeconomic status; a MWRA questionnaire that provides information on education, work, previous pregnancies, previous antenatal care, breastfeeding and health care practices, and retrospective infant mortality; a pregnancy questionnaire (or miscarriage questionnaire or maternal verbal autopsy) that provides information on illness in the first two trimesters, antenatal care and health care; a maternal post-partum questionnaire (or maternal verbal autopsy) that provides information on illness in later pregnancy, the process of delivery, puerperal illness and health care; an infant one-month questionnaire (or neonatal verbal autopsy) that provides information on essential newborn care, health at birth, early morbidity and health care.

Questionnaire processing

After completion, each questionnaire passes through a checking process, a schematic view of which is presented in Figure 2. The completed questionnaire is reviewed by a Surveillance Supervisor, who checks it on site and arranges for any corrections to be made immediately. The Surveillance Supervisor also observes at least five percent of interviews on site. The questionnaire is then transferred to the central office, where it is received by a Record Keeper. Questionnaires are counted, logged, filed and then manually checked for missing data and errors by a Data Auditor. If errors are correctable at this point, they are logged and corrected on site. Otherwise, the Data Auditor appraises the field team (in the person of either the Senior Surveillance Officer or a Surveillance Supervisor) of the error and the questionnaire is transmitted back to the field for correction. Error-free questionnaires are transferred to the data entry office. Data entry is carried out by

two Data Feeders. In the event of a hard copy error being noted during data entry, the Database Administrator is informed and the error marked. If the error can be corrected in the office, it is corrected and returned to the Data Feeder. If this is not appropriate, the error is logged and the questionnaire transmitted to the Data Auditor for discussion with the field team. Within the hierarchy of checks, the specific point of error detection and amendment is identifiable: the VDC Interviewer, Surveillance Supervisor, Data Auditor and Database Administrator each use a different coloured ink for annotation of questionnaires.

Questionnaire design

The questionnaires were designed in tandem with the data management system. Key variables required by the study were identified, posed as questions within the questionnaires, and defined as fields within the database. In cases where the range of

Figure 2

possible answers was unknown, the initial questions were left open and the answers were post-coded. Questionnaires were piloted in cycles, during which responses were observed and recorded, modifications suggested by the field team, and reformulation carried out. Questionnaires were formatted in Nepali, back-translated into English, piloted by the field team and amended by group consensus.

ELECTRONIC DATABASE

The MIRA Makwanpur study is large and the complexity of the data has been challenging. Previously, the study team would have entered the data into a non-relational ("flat-file") database. The first challenge, however, is the quantity of information. Some applications would not be able to handle serial data on 28 000 participants. The second challenge is complexity. Each participant receives a combination of questionnaires that depends on the outcome of her pregnancy (normal, miscarriage, maternal death, stillbirth, neonatal death). In a non-relational database, large sections would be empty: a woman with a normal pregnancy does not need to be asked questions about cause of death; a woman who has died before delivery can provide no information about newborn care.

One solution was to create a database in which questionnaires themselves are not the primary units of data storage. Instead, each questionnaire is divided into coherent blocks of information, and each block is entered into its own table. For example, information about antenatal care is stored in an antenatal care table. The information itself may come from a number of possible questionnaires depending on the outcome of pregnancy, one of which provides it for a given participant. The tables are then interlinked to a master table for the participant. Each participant then has a master table which is related to a variable number of subsidiary tables. This is an example of a relational database management system

(RDBMS). There are three major advantages to this system: decreased redundancy, variable sources of similar data and ease of extraction.

Decreased redundancy

We have already seen that, in a given case, certain information will be lacking. Women who do not attend for antenatal care can provide no information about it; women who have not had previous pregnancies can provide no information about them; women who do not access health care can provide no further information about it. If each block of data occupies its own table in the database, empty fields (the database equivalent of questions that are left blank because to fill them is inappropriate) are avoided. For example, if a woman does not attend for antenatal care, her master identification table has no related antenatal care table; *if a woman has had no previous pregnancy, her master identification table has no related retrospective pregnancy tables. This lack of redundancy has two primary effects: it makes the database more stable and it saves disc space. Note that there is an important difference between information that is missing on purpose because an event has not occurred, and information that is missing by accident because a question was left blank. The RDBMS deals with these in different ways. If data are missing on purpose, it checks that the relationships between tables are plausible. If data are missing by accident, it is programmed to query the gap, and often will not accept it.*

A further advantage of the linkage of multiple tables is the possibility of repeated tables. In one household, for example, more than one MWRA may reside. In a non-relational database, household details would have to be entered separately for each MWRA. In a relational system, however, one household table can be related to any number of MWRA master tables. This also applies when one

participant has more than one pregnancy during the study: only one master identification table is needed, but this single table is related to both sets of pregnancy tables. *This type of multiple linkage (often called a 1-to-many relation, as opposed to a 1-to-1) further reduces redundancy by removing unnecessarily repeated data from the database.*

Variable sources of similar data

Once the unit of data storage has been shifted away from an individual questionnaire, the opportunity for variable sources of data arises. Information about a certain subject may be gathered from one of several questionnaires. For example, if a woman's pregnancy is continuing at seven months of gestation, she will receive a standard pregnancy questionnaire. Within this questionnaire is a section about antenatal care. If, however, she has been unfortunate enough to have a miscarriage, she receives a miscarriage questionnaire: this also contains a section about antenatal care. Although the language in which the relevant questions are asked differs depending on the questionnaire, the fields (the answers) are the same. This means that the antenatal care table in the database can be filled from either of two sources (actually, from three, since the same questions are present in the verbal autopsy questionnaire). Overall, the result is that the dataset for each woman ends up being comparable to the dataset for all other women, even though the sources of the data (the questionnaires) are different.

Figure 3 is a schematic diagram of the database. Within each named table are a number of fields which are not illustrated for reasons of clarity. A number of other subtables have also been omitted, including some that are necessary for database administration. For each participant, the tables are filled in by a combination of questionnaires. The key table is the master table, which identifies the participant. The master table is then related to a

number of other tables. If a participant has had a previous pregnancy, the master table is related to each previous pregnancy (there can have been more than one) by a one-to-many join. Likewise, new pregnancies are related to the master table by a one-to-many join, and new infant tables are in turn related to pregnancies by a one-to-many join (this is commonest in the case of twins).

Ease of extraction

A database is designed for data analysis. The challenges posed by the MIRA Makwanpur database have had the secondary benefit of making the design team think "backwards" from the required outputs of the study. The issues discussed above – where they were explained in terms of storage space – have still larger effects on ease of analysis. Firstly, in a study of this size it is unlikely that any analyst will require the complete dataset. It is much more likely that questions will be asked of specific subsections of the data. The flexibility of the system allows retrieval of any combination of tables, or indeed of fields within those tables. Simpler still, however, is the process of analysing the gross outputs. To see how many women have attended antenatal care, it is only necessary to count the number of antenatal care tables. Likewise, the number of neonatal mortality tables is equivalent to the number of neonatal deaths.

Many of the outputs of the study can therefore be generated within the database management system, minimising the need to export data to statistical software applications, which reduces the file size required. Export of subcollections of data is, however, straightforward: all the data are numerically coded and exportable in a range of standard formats. After quantifying neonatal mortality, for example, it is a simple task to extract from the database verbal autopsies, aspects of health seeking behaviour, even socioeconomic status.

Figure 3. Schematic diagram of the database

Short programmes have also been written (stored procedures) that execute basic analyses and output them in a form that provides feedback to the field team. These stored procedures help to locate outliers and trends in the data, and the results are discussed with Surveillance Supervisors regularly. Frequency tabulation, measures of central tendency and spread and breakdowns to any level of data collection can be performed rapidly.

Data entry

Recent advances in software sophistication have made it possible to dispense with the traditional intermediate step of data coding. More accurately, the answers to a given question are precoded internally and the code is invisible at the user interface. Likewise, the complex traffic of data from a point in a questionnaire to a field in a table – several tables being filled by one questionnaire, and several questionnaires filling one table – is invisible. What appears on the computer screen is a facsimile of the questionnaire: the Data Feeder simply points and clicks on the appropriate response or enters numbers into the appropriate box. Although the programming was carried out in English, all the interfaces for data entry are presented in Nepali. During data entry, the system runs checks of completeness, range and validation, consistency, logic, and data entry. All questionnaires are double entered. The patterns of relations between tables and fields that match with every possible combination of questionnaires have been set. The RDBMS does not permit data to be entered that does not fit with a specific pattern. For example, if a woman has a miscarriage, it is not possible to enter data about the neonatal period.

The electronic system communicates with the field operations through a system of memoranda. Weekly reports on the identification of pregnancy are sent

from the field to the central office. These reports are acted upon by the Database Administrator, who converts the state of relevant MWRA within the RDBMS to “pregnant”. This and other changes in state trigger the initiation of a scheduler, which times the pregnancy and generates lists of questionnaires that are due at a given date. This allows the data team to produce a backup list of overdue questionnaires that can be communicated to the field team. The state coding system allows the team to identify sections of the dataset that are incomplete (through omission or delay), to attempt to rectify the omissions, and to deal with the omissions during the analysis phase.

System specifications

The system hardware runs on a Dell P1300 series server with a Pentium III 700MHz processor, 256 MB RAM and a 9GB hard disc (Dell Computer Corporation). The server is networked with four workstations, running via a dedicated online 3KVA Uninterrupted Power Supply affording at least three hours of runtime in the event of outage (Guraya Stabiline, Nepal). The workstations are dedicated for data feeding, password protected and not externally networked. Administration and modification are restricted to the Database Administrator. Backup is performed daily onto a compartmentalised hard disc, weekly onto a 250MB Zip disc (Iomega Corporation), and monthly onto a Zip disc that is transferred to Kathmandu and copied onto an identical server.

The RDBMS was created in Microsoft SQL Server 7.0. The operating system is Windows NT 4.0 and the data entry interface was programmed in Visual Basic 6.0. Questionnaires were developed in Microsoft Word 98 (all the foregoing: Microsoft Corporation) and formatted for printing in Adobe PageMaker 6.5 (Adobe Systems Incorporated).

DISCUSSION

The design of a surveillance system for a large community study is necessarily iterative, and perhaps the most important requirement is sufficient time to go through the process thoroughly. Various activities – database and questionnaire design, personnel guidelines, training, reporting – should mesh into a coherent whole. There is no substitute for repeated meetings between different cadres, to raise hypothetical scenarios, achieve consistency and engender ownership: an interviewer will probably perform better if she has played a part in defining the questions she is asking. Box 1 summarises the activities that we have found

Box 1

Key issues in designing a community surveillance system

A local emphasis

- Involvement of local leaders.
- Recruitment of local personnel.
- Involvement of field personnel in the design of the system.
- Use of locally appropriate concepts and language.
- Management of audit and correction at the periphery.
- Communication and movement of personnel between centre and periphery.

Clarity of aims

- Early definition of outputs.
- Generation of process from output.
- Understanding of potential outputs by all cadres.
- Feedback of outputs to all cadres.

Clarity of support

- A clear job description for each cadre.
- Focused training for each cadre at each phase of the surveillance.
- Extensive piloting and feedback.
- A clear chain of supervision.
- A clear system of reporting.

to be beneficial from the outset. The use of locally appropriate terminology and concepts is often recommended in guidelines for quality research. However, if the process really does take place in the local setting, and if interviews really do evolve in the presence of the interviewers, it is hard to see how the product could not be locally appropriate.

Data quality remains the prime concern. A summary of the error management system is provided in Box 2. Clearly, an error detection

Box 2

Summary of error checking procedures

- Observation of a designated proportion of interviews by a supervisor.
- Manual checking of all questionnaires by a supervisor.
- Sequencing of questionnaires and manual checking by an auditor.
- Omission checking by data feeders during data entry.
- Automatic electronic checking within software: checks of completeness, range and validation, consistency, logic, and data entry.
- Double entry, at a proportion depending on feasibility.
- Personal error logging for each data feeder.
- Raw output checking for outlier values and interfield relations.

system is only as good as the basic quality of the data within it. Misclassification and fictitious entries will not be picked up if data is self-consistent. To this end, it is important for data collectors to feel involvement in the study as a whole, and for there to be few incentives for lack of care. The commonest method to ensure this is to directly observe a proportion of the data collection interviews. This is only a technical requirement, however, and more powerful seems to be the creation of a network of responsibility for and understanding of the data collected. In this sense, a systematic introduction of the study to

community leaders and participants can pay dividends, particularly if they see themselves as primary stakeholders. The visit of an interviewer can be less of a negative experience if the interviewee knows of the reasons for the visit, and indeed has a personal connection with the interviewer, such as might result from the recruitment of local personnel.

The main issues of interest as regards the database system are flexibility and linkage with the field activities. Rapid advances in technology have made some of the more "traditional" aspects of health study data collection obsolete. Coding need no longer be an intermediate step in the handling of quantitative data, relational databases make information retrieval rapid and flexible, stored procedures make processing of raw data simple. The large size of the present study is something of a barrier to data handling, but this has been a stimulus for economical approaches. The use of multiple, linked tables allows redundancy to be minimised and also improves system reliability.

The most striking experience, moreover, has been the blending of the electronic process with the field activities. Widely available software allows questionnaires to be formatted and adapted within minutes of (and sometimes during) discussions. It allows results to be fed back to field personnel on a regular or ad hoc basis. It allows changes to be made at any time, reports to be generated and memorandum forms to be printed. What is most important, however, is that the clarity of thinking required to construct the system mandates a clear understanding of the required outputs. As a criterion for good scientific research, we are familiar with the aphorism "what is the question?". It can also be helpful to be asked by a machine, "what are all the possible answers, and which do you want me to show you?".

ACKNOWLEDGEMENTS

The MIRA Makwanpur Study receives funding from the Department for International Development (DFID), UK, from the World Health Organization (WHO), Geneva and from UNICEF, Nepal.

The surveillance system could not have been set up without the help of Mr JR Shrestha, Mr BP Shrestha, Mr S Tamang, Ms S Thapa, Mr B Thapa. The design was influenced by input from the Surveillance Supervisors: Mr D Adhikari, Mr B Aryal, Mr V Bhandari, Mr S Bhatta, Mr B Budathoki, Mr B Khadka, Mr SR Lama, Mr M Neupane and Mr R Neupane.

REFERENCES

1. Acharya L, Cleland J. Maternal and child health services in rural Nepal: does access or quality matter more? *Health Policy and Planning* 2000;223-9.
2. Sepehri A, Pettigrew J. Primary health care, community participation and community-financing: experiences of two middle hill villages in Nepal. *Health Policy and Planning* 1996;11:93-100.
3. Nepal Research Associates. Makwanpur District Profile: District Development Committee, 1999.
4. Howard-Grabman L, Seoane G, Davenport C. The Warmi Project: a participatory approach to improve maternal and neonatal health. An implementor's manual. Westport: John Snow International, Mothercare Project, Save the Children Fund, n.d.
5. Donner A. Sample size requirements for stratified cluster randomization designs. *Statistics in Medicine* 1992;11:743-50.
6. Ministry of Health (Nepal), New ERA, ORC Macro. Nepal demographic and health survey 2001. Calverton, Maryland: Family Health Division, Ministry of Health; New ERA; ORC Macro, 2002.
7. Pradhan E, Katz J, LeClerq S, West K. Data management for large community trials in Nepal. *Controlled Clinical Trials* 1994;15:220-34.