

Principles of health information systems in developing countries

Klaus Krickeberg

Abstract

The setting of this article is an all-embracing Health Information System (HIS)* of the type that exists mainly in developing and transition countries. It was inspired by work in Vietnam and other places. The article discusses the basic principles on which a well-functioning HIS needs to rest regardless of the technical means employed (paper, electronic devices). Eleven principles for designing or reforming a HIS are identified, including: explicit description of the underlying units (target population) and variables; no list of indicators to be fixed in advance; only one register per target population; technical coordination between registers and reports; correction algorithms; local use of data and indicators; autonomy of health institutions regarding the information that concerns them; and novel use of registers for various studies. Apart from their technical role, these principles give shape to the philosophy that underlies this article, and make clear that a HIS is not only a tool for collecting indicators; it is intimately tied to clinical and preventive practice, as well as to health management and health economy. In fact, it permeates the entire health system. It can potentially play a much more extended, varied, and useful role than simply that of providing health services to a community.

Key Words (MeSH):

Health Information Systems; Information Management; Community Networks; Developing Countries; Vietnam

* Note: Throughout this paper the abbreviation HIS refers to 'Health Information System' although it also frequently stands for 'Health Information Service' in other publications.

Introduction

One evening on a holiday in the late eighties, while strolling in Hanoi, I passed by a small Health Station and decided to enter. The two women on duty received me very kindly, but when I told them that I was working on Health Information Systems (HISs) and would like to have a look at their registers and reports they hesitated and intimated that it was all very messy. Instead they brought forward with obvious pride the 'CDD (Control of Diarrhoeal Diseases) Book' that I had designed for UNICEF (which they did not know), and said they liked it because it was clear and easy to use.

The present article is founded on the idea that the needs of the health personnel and health institutions where data and indicators of a HIS originate ought to be taken into account just as much as those of the users on a higher level in

the health hierarchy. The best way to achieve this when designing a new, or reforming an existing, HIS is to give it a clear logical structure and to follow a few basic principles. Although these principles may appear in part fairly theoretical and technical, they are intimately tied to the daily practice of the system. Observing them will motivate and enlighten the health workers handling the system, increase its reliability, enlarge its scope, improve its usefulness on all levels, and reduce costs. This can be expressed in our first principle:

PRINCIPLE 1:

A health information system needs to have a logical and transparent structure.

There exists an abundant literature on HISs. A large part of it is devoted to the description of particular experiences. Another part treats

special features including non-technical obstacles like a bad adaptation to administrative structures, jealousy between institutions, and 'donor attractiveness'. Lippeveld, Sauerborn and Bodart (2000) provide a good survey both on these issues themselves and on the literature up to the late nineties. The World Health Organization (2000) produced a document, *Health Information Systems Development and Strengthening* which, although mainly devoted to the assessment of HISs, may also serve to complement Lippeveld et al. (2000), and includes useful definitions. Bredehorst (in press) discusses many general problems in particular settings.

There is, however, apparently very little to be found on the 'general principles' aspect in the literature. The technical documents by this author starting in 1987 and listed in Krickeberg (1999) and also in Krickeberg (1994, 2003 and 2005b) take several steps in this direction, whereas the present paper goes further and attempts to give a synthesis of these data.

Setting

Health information systems have normally been built in order to provide routine information to higher level health administrations. Most of this information is of a statistical nature, that is, in the form of indicators. The data from which these indicators are calculated arise in basic health facilities such as Commune Health Stations (CHS), polyclinics, and hospital wards. The information is being transmitted via regular reports, for example monthly or quarterly. Most HISs are all-embracing in the sense that their information concerns clinical and epidemiological facts as well as economic and management aspects. The classical applications of the indicators provided by a HIS are twofold:

- planning and managing within the health system, in particular drawing up yearly budgets
- publishing information on the state of health of the nation and on its health system, for example in the form of a Health Statistics Yearbook and of reports to WHO.

Epidemic surveillance is an additional activity which rests not only on indicators but also on reported data about individual cases.

Most existing HISs have grown over a long period, adapting themselves to various and

sometimes conflicting needs, and to many users. As a result, in a given country there are usually several distinct administrative structures that house their own particular HIS. Even different departments of a Ministry of Health may run separate HISs. In addition, specialised health networks that deal with issues like hygiene, reproductive health, malaria and others find it necessary to rely on their own HIS, and so do most of the so-called vertical programs that pursue particular health strategies such as CDD, ARI (Acute Respiratory Infections), EPI (Extended Program of Immunisation), or HIV/AIDS. These numerous HISs are rarely compatible with each other. They often use different definitions of basic variables and indicators and different reporting periods. Above all, keeping an enormous amount of registers of different layout, and having to file reports all the time in many directions, places an unbearable burden on health workers even when they can do it electronically. Hence our second principle:

PRINCIPLE 2:

Attempt by all means to create an integrated HIS that serves all users.

We will see in the sequel what this may mean in concrete technical terms, and in particular why one does not need a separate CDD-book! It also means that managing health insurance should be included in the first application of an integrated HIS mentioned above, and not rest on a separate information system. This is already being done in some countries, though not quite consistently.

Technical concepts

It is a common experience to meet professionals involved in designing or exploiting a HIS who either do not realise that they need a clear definition of the concepts they are handling, or disagree among themselves about their meaning. In order to avoid ambiguities, the use of statistical terminology in relation to HISs is clarified as follows.

A *variable* concerns a specific *target population*. The members of the target population are called *units*; they may be of any nature, and not necessarily human beings! A variable attributes a numerical value to each unit. Example: the target population consists of all consultations carried out in a given CHS during the month of May 2007, and the variable 'age' attributes to each

unit 'consultation' the age in years of the patient examined. The values of the variables under consideration are the *data*. A *register* is a concrete and explicit representation of data, or in other words a device for recording them. This may be on paper, in a computer memory, or on any other support. In the classical layout of a register on paper, the lines correspond to the units, the columns to the variables, and in the boxes we find the data. The first principle stated above means in particular that a clear idea about the target populations and variables underlying a given HIS is indispensable not only for its planners but also its users.

An *indicator*, in contrast to individual data, does not concern a single unit but an entire group, or set, of units; thus it is a statistical concept. It depends on one or several variables defined in relation to the same target population. In the example above, the average age of the patients is an indicator that depends on the variable 'age' alone, whereas the number of consultations leading to the diagnosis 'acute respiratory infection' in the age group 1 to 5 years is an indicator determined by the two variables 'age' and 'diagnosis'.

Indicators that are based on several variables are prominent in epidemiology because they describe *relations* between variables. They appear in classical health statistics, too, for example in the form of a measles incidence by sex or age. There, all variables involved in addition to the main variable 'diagnosis' are of the type 'sex', 'age', 'date', 'place of residence', and 'profession'. It is obvious that in order to calculate from registers an indicator that depends on several variables, these variables must have been recorded in the same register.

For the same reason, it would be disastrous not to register *health* data like 'diagnosis' and 'treatment' in the same book as *economic* data such as 'payment' and 'health insurance status', and to create separate 'epidemiological' and 'business' HISs as it has sometimes been advocated. It would in particular annihilate any effort to base the management, monitoring, and continuous improvement of health insurance consistently on the existing HISs. Nevertheless, health administrations have sometimes attempted

to 'split' a register in order to obtain smaller books.

We therefore state our third principle:

PRINCIPLE 3:

In a given health institution, for each target population, that is for each type of unit, there shall be only one register.

There exists another stringent reason why one needs to abide by this principle. It is a more practical one, which has already led us to Principle 2, namely that several registers for the same target population put an unnecessary burden on the health worker. For example, in order to satisfy the demands of different specialised health networks and vertical programs, a health worker is often obliged to record the data concerning a consultation into different books or computer files depending on the diagnosis; it may be in the 'general' register, or more specific ones, for example for malaria, tuberculosis, mother and child care.

The question arises, of course, is there enough space for all the variables we need in a single register? We are going to look at this problem in the following section.

In the theory and practice of HISs, the question of *links* between different target populations, and thus between different registers, is basic. For example, consultations (or episodes) are *subunits* of cases, which are subunits of patients and so on. Still another type of target population is a *cohort* of people to be followed from initial moments that may depend on the individual person. A correct EPI-register should in principle be based on a target population of this kind although in practice one uses a register of vaccinations; the link is provided by a Lexis-Diagram as found in any textbook on demography. A natural unit for a MCH (Mother and Child Health)-register would be a *pregnancy* where data about antenatal care, delivery, postnatal care and follow-up of the child up to a certain age, for example 12 months, are all recorded together. By including data on vaccinations for children under one year among the latter, we see that MCH- and EPI-registers concern in fact the same target population. Hence Principle 3 implies that separate EPI-register can be abolished, which is indeed quite feasible in practice. These questions will not be discussed

in this paper in any detail, however; we will mainly illustrate concepts by the simple, and most important, example of *consultations* as the underlying target population.

Choice of variables

It has been stressed many times that the shape of a HIS must be mainly determined by its objectives, and that it must not contain superfluous elements or redundancies. In practice, one does not always adhere to this laudable principle. Sometimes it is also interpreted in a naïve way that does not take into account the structure and functioning of a HIS. In particular it was stated that when designing a new HIS, one should start with a list of all indicators to be routinely reported. For example, recently in Vietnam when building the partial HIS known as the Health Management Information System, a list of roughly 400 such indicators was drawn up, of which about 180 alone concern basic health facilities like CHSs. It is unlikely that they are being analysed, let alone used, regularly. Such a rigorous and fixed list is also incompatible with the fact that the needs regarding indicators change and cannot be foreseen over a somewhat longer period. In other words, it contradicts our fourth principle:

PRINCIPLE 4:

A HIS must be flexible in order to adapt itself to changes of all kind such as evolving sociologic and economic conditions, changes of the epidemiological situation and the state of health of the population, scientific progress in public health and medicine, and changes in information technology.

The basic structural elements of a HIS are the *variables* that appear in it, not the indicators to be calculated. The shape of the registers and the type of the reports depend on the variables, and therefore these cannot be altered without affecting the entire HIS, be it paper- or computer-based. Hence a good choice of the set of variables to be included in the system is one of the fundamental decisions to be taken right in the beginning. Indicators, on the contrary, can be calculated from a given and fixed set of variables in many ways dictated by needs that may change with time. They may even include certain indicators needed only once.

In order to determine the variables that are required in view of deriving indicators, it suffices to have a general idea of the kind of indicators that are deemed necessary at present or in a foreseeable future. A careful analysis will usually reveal that such a list of variables can be relatively short even if we take into account the demands of many users in view of the Principle 2. In the following section, we will also mention techniques for reducing the space needed for recording variables in a register on paper.

Above all, calculating indicators for the first two ‘classical’ uses listed above is not the only purpose of a HIS. There are several other uses of equal importance. They will be taken up in the sections ‘Local use of Health Information Systems’ and ‘Novel uses of a Health Information System’ below, and they, too, influence the choice of variables to be in a HIS.

Registers, reports, flow of information

Registers form the root of a HIS. They start with the village health worker who employs small notebooks or sheets. They have to be designed carefully in view of their many uses. It is important to designate the units and the variables clearly. Simple standardised rules on how to write the data by appropriate coding are helpful for the health worker handling them. It might be expedient to denote the most frequent diagnoses made in a CHS by abbreviations of their names in local language, for example in Vietnamese LYA for amoebic dysentery, CUM for influenza, IAC for diarrhoea. Such a coding must be compatible with the International Classification of Diseases in the sense that every such abbreviation corresponds to a well-defined disease or group of diseases of the International Classification, but it would be impractical to use this classification itself in lower level health institutions. The notebooks of village health workers, too, should be standardised, include clear rules for their use, and be coordinated with the registers of their CHSs.

To enhance the logical structure of a register, the data for a given variable have to be recorded in a single ‘column’ of the book or computer file that serves as its support. It is, for instance, bad practice to use *two* contiguous columns for the variable ‘age’ by writing the age of a man in the left and that of a woman in the right column, thus making it difficult to separate the variables ‘age’

and 'sex' when calculating compound indicators that imply one or both of them. In certain situations one can, however, save space by recording several variables in the same column provided that they are clearly distinguished from each other in view of the evaluation of the data. For example, for some data it is important to point out that they pertain to a particular vertical program. This can be indicated in the same box by a suitable code like D for CDD or R for ARI. With the help of good coding the column 'Remarks' may also be employed in many more ways than usual.

The hardest and most time-consuming tasks for health personnel charged with handling a paper-based HIS are calculating indicators, transferring them onto reporting forms, and dispatching the reports in various, and often many, directions. The main reason is that reporting forms are almost always designed in order to report indicators considered as having already been calculated before. They do not take into account the process of calculation and the form of the registers that contain the original data. Calculating indicators is a separate activity, executed, for example once a month, on separate supports like tally sheets or blackboards. Much work, time and paper can be saved by adhering to the following principle:

PRINCIPLE 5:

Coordinate registers and reporting forms between them by their layout and by a clear designation of corresponding variables. Calculate indicators as much as possible as part of the daily work routine.

The CDD-information system mentioned in the Introduction was built around these ideas. Principle 5 implies in particular that instructions for health workers on how to calculate and report a specific indicator should start by an explicit identification of the variables on whom it depends, mimicking thus what the software of a partially or entirely computerised HIS does. In a HIS which follows this principle, it is possible to calculate *all* required indicators that concern a given target population and a given period by running through the register only once, and not anew for every indicator as is usually being done. If tally sheets are being used, they must

of course also be coordinated with the registers and reporting forms. In the normal case where a report does not involve too many indicators, one can even incorporate tally sheets into the forms as 'tally areas'.

Information may flow horizontally, that is, between institutions of the same level; upwards, that is, to higher-level institutions; or downwards. In a paper-based HIS, routine reports are either being mailed or handed over at the occasion of meetings between the staff of different health facilities. Transmitting the same or similar reports to several health offices of the same or a higher level places again an unnecessary burden on the health workers involved, regardless of the means employed. It also still happens that an institution is asked to report information that the receiver can deduce himself from information already reported before, for example from bi-annual consolidated incidences of monthly incidences. Hence the sixth principle:

PRINCIPLE 6:

A health institution sends a report on paper only to the higher-level institution that needs it most or most urgently. It is then up to the latter to distribute it horizontally to those who require it. A higher-level office never requires 'summary' reports from the lower level.

If a health facility is connected electronically to higher- or equal-level institutions by email or in a special network, sending several copies of a report in several directions is easier, but one should not use this fact as an excuse for multiplying reports indiscriminately: this would simply shift the burden from the sender to the receiver!

Different periodicities of reports requested by different administrations are often cited as obstacles to the integration of their own HISs. For example, a department of a Ministry of Health may ask for quarterly reports and a particular vertical program may insist on monthly ones for the same indicators. This problem can practically always be solved by renewed planning in order to eliminate unnecessarily frequent reports, and in the worst case by suitable horizontal reporting between the offices in question.

The flow of information as a whole should of course be minimal, and dictated by the uses one makes of the HIS. We will therefore come back

to it in the two subsequent sections, including downward flows.

Local use of health information systems

Although rarely mentioned in descriptions, instruction manuals, and other documents surrounding a HIS, there exist in fact many other actual and potential uses in addition to the first three applications quoted above. To start with, recall that most data that enter the system arise from clinical work. Although also being used to derive indicators, they had originally been recorded by the health personnel during consultations as a support for case-management. A line in the Register of Consultations of a CHS corresponds to the classical clinical record in a medical practice or hospital ward. Clear and precise recording of symptoms is indeed essential for a good clinical diagnosis and the choice of the treatment. This may go as far as applying an elementary but still precise and rigorous formal clinical decision process such as in the programs CDD and ARI. If several consultations are involved for a case, mainly for chronic diseases, the entries in the register of consultations, together with other data, are indispensable for the follow-up. Similar remarks apply to case-management in polyclinics and hospitals. Hence the fourth objective of a HIS:

To support and improve the daily clinical work of the basic health institutions by using the recorded data directly.

The dual use of data from basic health facilities for case-management on the one side, for calculating indicators on the other, is reflected by techniques to coordinate and combine them as discussed in the section 'Registers, reports, flow of information'. These two functions should be viewed together. For instance, reports on forms that incorporate tally areas for counting consultations for the most frequent causes provide the health personnel with an immediate and vivid view of the health situation in their community.

More generally, many indicators that are derived in a basic institution in order to be reported are also highly relevant to the work of the institution itself. This is fairly obvious for simple indicators on management like 'number of consultations' or 'number of patients without health insurance' in a given month, or indicators on the logistics of drugs. They are being used

routinely in running a CHS. To exploit epidemiological indicators locally is less common although it had once been advocated by WHO under the name of 'epidemiology at the basis'. Almost every health worker has of course a general idea of the disease spectrum in his CHS but to calculate and write down, or even to represent graphically, the main incidences is much more helpful. Figures on case fatality, too, may be revealing to him or her. The planning of preventive measures by a CHS can profit from the knowledge of indicators that link morbidities and mortalities to some of the variables recorded in the register of consultations like 'date' (to study seasonal fluctuations and trends), 'age' or 'sex' (to describe the distribution of cases and deaths by age and sex, respectively), or 'address' (to obtain the geographical distribution of morbidities and mortalities in the commune, for example in order to link certain diarrhoeic diseases to possible environmental factors). Interpreting such indicators is also interesting and motivating for the health worker. Moreover, it allows a self-evaluation and self-control of the work of the CHS. In short, the fifth objective of a HIS is:

To enable basic health institutions to calculate useful indicators about their own situation and work, and to exploit them locally.

Finally, beyond local use of data and indicators, a clearly and logically built HIS in the sense of Principle 1 fulfils an important pedagogic function on all levels, up to the Ministry of Health. Its structure reflects important aspects of the structure of the health system itself, and can even suggest improvements of the latter. In particular, a well structured HIS indicates ways of how to integrate, or not to integrate, various parts of the health system. For instance a simple look at the way of handling information on CDD and ARI would have shown that these programs should never have been separated, and that ARI should never have been combined with a tuberculosis program (Krickeberg 1989-90).

Autonomy regarding information

The interaction between local uses and reporting functions as sketched in the preceding section, which is supported by the technical Principle 5 above, is the main idea that governs a modern efficient HIS. Health administrations are not

always aware of it; for example, in a recent attempt to simplify the calculation of indicators arising in a CHS, the station was asked to abolish its Register of Consultations and to make instead for each consultation a separate report on a slip of paper. These slips were then brought to its District Health Centre (DHC) and used there to calculate the requested indicators on a computer. The DHC sent the indicators back to the CHS in addition to forwarding them to the Province Health Department (PHD). The reformers forgot what the personnel of the CHS knew very well, namely that the data in the Register of Consultations are also needed, and used, in the clinical work of the station. This register was their indispensable 'data bank', and therefore the CHS kept and used it as before *in addition* to handling the slips, which had not been foreseen. Thus the reform caused much more work to the health workers of the station instead of alleviating their burden. Regarding the DHC, it was flooded with information on all single cases treated in the CHS for which it had no individual use.

Recall from the section 'Registers, reports, flow of information' above that local calculation of indicators can be made quite easy if well organised. The fact that in the 'reformed' system the indicators pertaining to their own domain of the CHS were 'handed down' from the district level instead of being obtained locally, also amounted to a certain alienation. Its staff now only obtained a belated and much less direct view of the epidemiological, economic and management aspects of its activities.

In short, it is vital that CHSs and similar basic health facilities remain autonomous masters of the information that originates in them and concerns them, and that district level institutions do not receive information they do not need.

We are now going to generalise this idea to health institutions of any level. In a DHC, the indicators emanating from its communes are being consolidated to obtain indicators about the district as a whole, to be reported to its PHD. They also form at the DHC a valuable 'indicator bank' (indicator base) for its own needs in planning, managing, deciding on health strategies, and possibly for supervision of its CHSs, although the usefulness of the latter application is usually over-estimated. However, contrary

to the practice in certain countries, the DHC should *not* transfer these indicators about its individual communes to the PHD. Regardless of the technical means employed at the district and province level (paper, electronic), storing and handling all of these 'communal' indicators causes much superfluous work and expenses in the PHD. There, they are not being used routinely, and in fact hardly ever. If, for a rare special occasion like a particular publication, the PHD needs such and such communal indicator, it can easily go back to the relevant DHC and retrieve it from there.

Analogous remarks apply to the next higher level, that is, to reports from the PHD to national institutions such as the Ministry of Health. This can be summarised in the principle of autonomy regarding information:

PRINCIPLE 7:

A health institution needs to develop the use of information in its own work including the setting up of data or indicator banks. With a few exceptions, a basic health institution does not report data that concern only its individual acts (e.g. consultations) but transmits instead indicators about the institution as a whole. Similarly, a district or province level health office normally reports indicators about itself but not about entities on a lower level.

There have been frequent complaints in the literature about insufficient 'feed-back' of indicators and other information. However, as we have seen in the preceding discussion and as it results from the principle of autonomy, feeding back to a health institution indicators that concern its own domain is rarely called for. The institution obtains them directly in the course of its work. It makes sense, however, to feed back information about higher levels, for instance to give a CHS selected indicators about its district, its province, its region, or the entire nation.

Autonomy regarding information plays another fundamental role. It is in fact an absolutely necessary prerequisite for decentralising the health system without destroying the nation-wide links between its elements that are indispensable for good planning and efficient management.

Correction algorithms

Quality of data and indicators is a crucial issue. Whereas for single statistical studies many

methods for checking data at their source and for controlling errors in their further treatment exist, most of them are time-consuming and expensive and can hardly be put into practice when collecting and reporting data and indicators routinely. The most efficient way for reducing gaps and errors in data and indicators is no doubt to motivate the health staff by giving them a logical, transparent, and easy to use HIS.

Nevertheless, many gaps and errors of various kinds will remain. The normal practice of health administrations is to do nothing about them. For example, if a district counts 13 communes, but only seven CHSs have handed in their monthly report, and if there had been 18 consultations in these seven CHSs that lead to the diagnosis 'measles', then the District Health Centre will report 18 cases of measles to its Province Health Department, a very unlikely figure for the whole district. Worse, the DHC will not inform its PHD that some reports had been missing.

What is needed, and ought to be the subject of more research, are simple correction procedures that can be applied, for instance, to every monthly report of a DHC. They would necessarily be based on the same idea as the various sophisticated methods in use in single studies, namely to exploit additional information that we may have. Note that DHCs, in contrast to CHSs, frequently have a simple computer at their disposal for composing reports. Applying a correction algorithm already incorporated into the relevant software hardly demands any additional work from the health worker who makes the report. To illustrate the idea, let us look again at the example above. Instead of reporting 18 cases of measles to the PHD, the DHC would exploit the additional knowledge that there are 13 CHSs of which only seven have filed their report, and replace the incidence figure 18 by the *estimated* figure, rounded to the nearest integer, $18 \times (13/7) \approx 33$. In general, and in the long run, this procedure will furnish much better incidence figures than the old one.

At the present time there seems to be only one method in actual use for dealing with missing or erroneous information, namely *correcting factors*. An example: a household survey in Guatemala has furnished a maternal mortality 1.58 times higher than the figure reported for the same

period within the demographic information system. This fact is the additional information to be exploited. Hence the reported figures on maternal mortality in Guatemala are now systematically being multiplied by the factor 1.58 (Ministerio de Salud Pública 2002).

Methods like the preceding ones can of course be refined in many ways. *Bayesian algorithms* will be sketched in the following section. It goes without saying that the use of any correction mechanism must be clearly indicated when reporting or publishing the resulting figures.

In a sense, 'active search for cases' can be regarded as a method for improving incidence or prevalence indicators. For example, in the present cancer registries in Vietnam, the collaborators pay regular visits to the treatment facilities that are likely to come into contact with cases of cancer. In a well functioning HIS these should have been reported automatically but it turns out that the incidences obtained by active search are significantly higher.

To sum up in the form of another basic principle:

PRINCIPLE 8:

A HIS must comprise clear rules and methods for handling missing or erroneous data and indicators. Correction algorithms should be developed.

Novel uses of a health information system

Much information on the structure and functioning of a health system, and most epidemiological knowledge, is being derived from sample surveys and related single studies. They are often being viewed as complementary elements to the output of a routine HIS. We will have a glance at this aspect in the following section. The present one is devoted to the converse problem, which we consider equally important, namely how can a routine HIS be put to good use in order to answer particular non-routine questions of the type that are normally the subject of special studies based on sampling. Such applications, including examples, have already been identified by Krickeberg (1994); and a systematic treatment outlined by Krickeberg (2005c). Both of these papers provide more detail; here we will only outline the main ideas.

We note first that an analogous trend prevails in other parts of statistics. For example, in some countries the classical census is being abolished and replaced by an investigation that is largely based on existing registers (interview with Eric Schulte Nordholt 2004). In fact, in a census as well as in a sample survey or epidemiological study, the most expensive and difficult part is not the later analysis but taking the sample and obtaining the data. In a register, these two are already there; hence it appears tempting to use them, although it is of course not obvious how and to what extent this could be done.

To have something concrete in mind, let us take the register of consultations in a basic health facility like a CHS. As observed in the section 'Autonomy regarding information', it is a gold mine of information. How can we exploit it? This question has four facets: what are the possible target populations; which variables may be involved; which categories of studies can be envisaged (cross-sectional, cohort etc.); and finally, intimately connected with the three preceding facets, what are the interesting subjects that can be treated in this way?

The target population is, by definition of the register at hand, the set of all consultations carried out during a certain period. Hence a study that exploits the register will be appropriate for any problem that concerns precisely this population. Prominent among such problems is the evaluation of the case-management routines of the CHS. Case-management generally consists of the anamnesis, diagnosis, treatment, payment, sometimes follow-up, and ideally good advice to the patient about treatment and prevention. A study may concern some aspects or the entire process. An important class of problems consists of the evaluation of decisions made exclusively on the basis of the anamnesis and clinical symptoms without using laboratory tests because the CHS may not be sufficiently equipped for this and there may be no time or money to call upon the services of an outside laboratory.

Let us illustrate the basic concepts and steps with the help of a few examples. First we look at the question of estimating the four characteristics (sensitivity, specificity, predictive values) of a purely clinical diagnosis of an acute disease like shigellosis. To start with, we have to decide about

the consultations to be included in the study. We may take all consultations made during the period in question, or only a sample drawn by a correct sampling plan. The latter option is known as 'sampling from records' and is quite useful but little applied. In either case, we will have a cross-sectional study.

Next, regarding the second facet of the general question of how to exploit the register we note that in order to estimate the four characteristics of the decision scheme in use at the CHS, all variables needed are already in the register except the 'gold standard', that is, the true state of health of the patient (having shigellosis or not). For the purpose of the study we now have to do what the CHS cannot do routinely, namely to provide suitable laboratory facilities in order to verify the gold standard by an analysis of the stool. Adding the variable 'gold standard' during the study to those of the register makes it a 'partially register-based study', which is in fact the key concept of the present section.

Having obtained estimates of the four characteristics at hand, we can apply them to setting up *Bayesian correction algorithms* in order to replace reported incidences in basic health facilities such as those of Shigellosis by estimated ones (Krickeberg 1994). The latter will generally come closer to the truth than those found in the usual reports. Similar ideas have been advocated in (Byass, Dao Lan Huong, & Hoang Van Minh 2003) in a slightly different context; they could also be applied in the present setting.

In our second example we want to evaluate the case-management of patients suspected of having malaria. In many CHSs, the decision scheme still consists in providing an initial drug treatment depending only on an initial clinical diagnosis before the correct one is known through the examination of a blood sample under a microscope or a more modern test, and before referral to, for example, a district hospital. However, in most malaria control schemes, suspected malaria patients are being followed up in contrast to those suspected of suffering from shigellosis, hence the correct diagnosis will *eventually* be known and recorded. Therefore, in order to estimate any kind of efficacy of the decision scheme from the registers, it is not necessary to add the 'gold standard' as a new variable for the purpose of the

study; it is already there, and the study will be entirely, and not only partially, register-based. It is both cross-sectional by the choice of the sample, and longitudinal in the sense of follow-up in time.

The most interesting feature of the two preceding examples is that the *same* study also allows to investigate the characteristics of *other* conceivable diagnostic schemes that are also based exclusively on clinical symptoms recorded in the registry, and thus to compare them, find an optimal one, and perhaps replace the one actually in use by a better one.

Our third example does not concern case-management but prevention. Sometimes it appears desirable to monitor the efficacy of a vaccination strategy that had already been applied for some time. Since this efficacy is a simple rational function of the vaccination coverage, which is supposed to be known in a good vaccination program, and the proportion of vaccinated persons among cases (Kleinbaum, Kupper & Morgenstern 1982), it suffices to estimate the latter by a partially register-based cross-sectional study where we add the variable 'vaccination status' to those already in the register of consultations. This estimation combined with that of the vaccination coverage amounts to a hybrid study but again we cannot enter into details.

We can also envisage certain etiologic studies. Some elementary examples have already been mentioned at the end of the section 'Local use of Health Information Systems'. When trying to proceed along these lines on rigorous grounds, we are first confronted with the question of the right target population. The natural one will now be one like 'all inhabitants' or 'all children' in the commune. In a cross-sectional study that takes only recorded consultations into account, there will of course be a hospital bias, or more precisely a 'CHS-bias', but the population consisting of those persons who come occasionally to their CHS for consultation is usually more characteristic of the entire population than the one found in a hospital. It is also possible, and no doubt useful, to conduct fruitful case-control studies where the controls are taken from the population at large regardless of whether they had consulted their CHS or not. On the contrary, efficient cohort

studies making essential use of a register in a basic health facility are hard to conceive.

Regarding the risk factors to be investigated, some are already in the registers as noted earlier; others will have to be added during the study, making it a partially register-based one.

The studies mentioned until now concern health directly. Entirely or partially register-based investigations of the structure of the health system and of its management and economic aspects have apparently not been considered in a systematic and rigorous fashion, but there are obviously many profitable ways to conduct them. They can certainly play a very useful role in developing an efficient health insurance system.

To sum up, we state the ninth principle:

PRINCIPLE 9:

Explore all possibilities of exploiting a HIS for conducting, or facilitating, research studies.

Pursuing such studies would be the sixth objective of a HIS.

Non-routine methods

Surveys and similar studies on issues of health abound in developing and transition countries. Most of them are household- or family-based, others are person-based or community-based. Isolated sample surveys evaluated as cross-sectional studies represent the majority of these. In particular, vertical programs have usually launched a survey at their beginning in order to estimate 'base-line indicators', and repeated this later to monitor changes that might reflect the impact of the program. In addition, there are longitudinal observational studies, some of them on a permanent basis, conducted by a 'sentinel' or 'field' site like FilaBavi in Vietnam (*Scandinavian Journal of Public Health* 2003). Finally, epidemiological cohort and case-control studies are being organised by various bodies.

The various surveys done in a given country are rarely coordinated with each other. They may employ different definitions of basic concepts and variables, and often duplicate existing knowledge. With the exception of scientific studies published in the relevant journals, the results are accessible only with difficulty or not at all. (For a more detailed discussion, see Sections 5 and 6 of Krickeberg 2005a).

In the present section, we are interested in the interplay between special surveys on the one hand and the routine HIS on the other. Health administrations sometimes quote the usefulness of independent sample surveys for controlling the work of the HIS by comparing the resulting indicators. However, organising a sample survey with this sole objective is never justified. An 'internal' control by inspecting the registers, by talking to the staff, and by tracking the various steps that lead to the reports, is much cheaper, faster, and more efficient. It provides a direct and vivid vision of the functioning of the HIS.

More generally, there should never be a survey where results can be completely derived from the HIS as well. The money is better spent on strengthening the relevant components of the HIS.

There exist of course many well known and much discussed situations in which sample surveys are called for, and we will not take them up here. However, any one of them ought to be conceived of as a supplement to, or even part of, the general HIS. This means describing its objectives in relation to those of the routine part of the HIS, and coordinating concepts and notations.

Quite often, there is a *partial* overlap between the results of a survey and those that can be deduced from the routine HIS. In such a case, comparing both is almost always illuminating. Apparently, it is hardly ever done, though.

If some, but not all, variables in a sample survey appear also in a register of the routine HIS for the same, or a related, target population, it is instructive to 'link' the survey to the HIS, for example by trying to identify persons included in the sample also in the register. This allows answers to interesting questions on vaccinations, use of health facilities and the like, without much extra cost.

In short, our tenth principle reads:

PRINCIPLE 10:

Regard, and treat, sample surveys and related studies as components of a general HIS.

The Health Metrics Network launched by WHO a few years ago attempts to provide a standardised framework for obtaining statistical health information from *all* available sources. These include clinically-based data and other routine data like

those from vital event registries on the one hand, and surveys on the other. (For a short description, illustrated by the case of mortality indicators, see Mikkelsen 2007).

Use of electronic mean

The HISs of several countries have been the subject of numerous reports and recommendations for improvement made by various organisations, the most prominent among them being WHO. In the case of Vietnam, there exists, as described by Krickeberg (1999), a survey and analysis of *all* such reports up to 1999; it was updated by this author for WHO until early 2004. In these reports, the first proposal regarding the use of computers appeared apparently in 1987 and was followed by many more. They did not however, explore in depth matters tied to the general structure of the HIS. In the meantime, several components of the Vietnamese system have been computerised, mainly in order to facilitate the writing of reports and the calculation of indicators from the district level up, but without touching the structure proper. A lot of software was developed, especially for hospital information systems, often without coordination among them.

Similarly, in 1987 (see Krickeberg 1999), it was suggested that when designing or reforming a HIS, its logical structure should already be shaped with a view of later use of electronic means but that it would be useless to computerise a badly structured HIS. In fact, as already observed in the section 'Register, reports, flow of information', in an efficient HIS many operations on paper reflect what a computer does. These principles are still valid.

Moreover, looking at the matter now in the light of the present situation in most developing countries, it becomes clear that the main problem is not whether to stick to a paper-based system or to switch to a completely electronic one, but rather how to build a streamlined, partially computerised HIS where paper-based components function together with computer-based ones. In general, work on paper will be concentrated at the lower level, for example in some CHSs, but one needs to leave open the possibility of different information carriers at the same level, or even paper at a higher and computers at a lower one. Such a mixed HIS, too, can only be efficient if its basic structure along the lines sketched in

the preceding sections had been defined previously. This holds equally well for projects like electronic patient records.

Thus, health administrations might consider the following principle:

PRINCIPLE 11:

Analyse and make apparent the structure of the HIS first, and then think about computerisation. All software within the HIS including that for hospital information systems needs to be developed and applied in a coherent and coordinated fashion.

Summary

The design or reform of a HIS should be based upon an understanding of six principle objectives or applications of the system:

- planning and managing the health system including health insurance
- publishing health-related information
- epidemic surveillance
- supporting and improving daily clinical work
- obtaining information for local use
- conducting studies

With these functions in mind, eleven structural principles necessary for this process have been identified:

- 1 Logical and transparent structure
- 2 Integration in order to serve all users
- 3 Only one register for a given target population
- 4 Flexibility – no rigid list of indicators fixed in advance
- 5 Coordination between registers and reporting forms
- 6 Minimal flow of information
- 7 Autonomy regarding information
- 8 Correction algorithms
- 9 Exploit the HIS for conducting studies
- 10 Sample surveys and related studies as components of a general HIS
- 11 Computerise only a HIS that has a clear and well understood structure. Develop software in a coherent way.

A HIS should not be regarded only as a tool for collecting indicators. It is intimately tied to clinical and preventive practice as well as to health management and health economy; in fact, it permeates the entire health system. It can potentially play a much more extended, varied,

and useful role than the one for which it is primarily intended.

Acknowledgement

I am much indebted to Dr. Björn Wenngren for discussions we had when working together in Vietnam in 2004 for the Vietnamese-European Health Systems Development Program. Dao Lan Huong, Phan Vu Diem Hang, Kampheng Phongluxa and Maren Bredehorst read a draft of the present paper and made pertinent remarks that helped to improve it and for which I am grateful. Last but not least, I thank the two referees for their useful comments.

References

Bredehorst, M. (in press). Information systems for the rehabilitation of landmine survivors. In: Razum, O., ed. *Challenges in Public Health*. Frankfurt am Main, Peter Lang.

Byass, P, Dao Lan Huong, and Hoang Van Minh (2003). A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. *Scandinavian Journal of Public Health* 31; Suppl. 62: 32-37.

Interview with Eric Schulte Nordholt. (2004). The Dutch virtual census of 2001. *International Statistical Institute Newsletter* 28: 3 (84).

Kleinbaum, D.G., Kupper, L.L., and Morgenstern, H. (1982). Section 9.1.2, *Epidemiologic Research*. Belmont, CA: Lifetime Learning Publications: 164-168.

Krickeberg, K. (1989-90). Intégration des soins de santé primaires. Lecture at the Faculty of Medicine of Phnom Penh, January 1989. *Séminaire de Statistique Médicale, Université de Paris V*: 45-69.

Krickeberg, K. (1994) Health information in developing countries. *Frontiers in Mathematical Biology, Lecture Notes in Biomathematics* 100: 550-568.

Krickeberg, K. (1999) . Chapter 6 of: *The health information system in Vietnam in 1999*. Assignment report for the joint programme of Vietnam and the European Committee on Health Systems Development. Available from the author.

Krickeberg, K. (2003). Health information systems in developing countries. *Bull International Statistical Institute, 54th Session Proceedings, Invited Paper Meeting* 54. Den Hague, ISI.

Krickeberg, K. (2005a). Epidemiology in developing countries. Chapter 6 of: *Handbook of Epidemiology*. Berlin-Heidelberg-New York, Springer.

- Krickeberg, K. (2005b). Health information systems. In: Becher, H., Kouyaté, B., editors. *Health research in developing countries*. Berlin-Heidelberg: Springer: 43-49.
- Krickeberg, K. (2005c). Neue mathematisch-statistische Anwendungen von Gesundheitsregistern. *Austrian Journal of Statistics* 34: 403-410.
- Lippeveld, T., Sauerborn, R. and Bodart, C., editors (2000). *Design and implementation of health information systems*. Geneva, World Health Organization.
- Mikkelsen, L.(2007). Health Metrics Network and mortality data. *Health Information Management Journal* 36: 49-53.
- Ministerio de Salud Pública y Asistencia Social (2002). *Línea basal de mortalidad materna para el año 2000* Guatemala. *Scandinavian Journal. of Public Health* (2003); 31; Suppl. 62.
- World Health Organization (2000). *Health Information Systems Development and Strengthening*. Geneva, World Health Organization.

Klaus Krickeberg *Dr.rer.nat*

Grosser Kamp 4

33619 Bielefeld

GERMANY

email: krik@ideenwelt.de ■