Making data credible

Proposals formulated by members of a workshop held in February 1984 about the setting, achieving and monitoring of data standards with particular emphasis on standards for clinical activity data

Published by the King's Fund on behalf of the NHS/DHSS Health Services Information Steering Group

Making data credible is the fifth of a series of occasional papers produced by the NHS/DHSS Health Services Information Steering Group and published on its behalf by the King's Fund. The other titles in this series are

Converting data into information: proposals formulated by members of two workshops held in March 1982 about the management arrangements required for collecting valid clinical data and providing a district information service

Introducing IT in the district office: proposals arising from a study carried out in Southend Health District by Aslib Research and Consultancy in 1982

Developing a district IT policy: proposals formulated by members of a workshop held in June 1983 about the development of a district policy for the introduction of information technology with particular emphasis on the implementation of computerised departmental information systems

Piloting Körner: the views of senior administrators from the four districts who piloted the interm reports of Working Groups A to E from 1981 to 1983

Making data credible

Proposals formulated by members of a workshop held in February 1984 about the setting, achieving and monitoring of data standards with particular emphasis on standards for clinical activity data

Published by the King's Fund on behalf of the
NHS/DHSS Health Services Information Steering Group

Contents

Typeset by Prototype
Printed by G S Litho, London

Truth lies within a little and certain compass,
but error is immense.

Henry St John, Viscount Bolingbroke 1679-1751:
Reflections upon Exile

Preface

The cost of information systems in the NHS can only be justified on the ground of their usefulness. It is therefore important to bear in mind that the use that is or can be made of them by clinicians, managers and others depends crucially on the quality of the constituent data. If that quality does not measure up to the purposes for which the information is intended, the systems fall into desuetude and the investment of money and effort devoted to them is wasted.

A new awareness of the importance of information for decision makers at all levels has now brought the NHS to the threshold of a new era of information gathering, dissemination and application. The authors of the present publication correctly anticipate some of the problems about data standards which will have to be tackled if we are not to forfeit the harvest we hope to reap. The proposals made towards their solution will, I believe, be welcomed by all who will be implementing and maintaining the new systems.

We are grateful to all the contributors to the workshop for the time and effort they have put in. We are particularly indebted to Dr Michael Goldacre, Professor Anthony Hedley and Mr John Yates for so generously allowing us to use the results of their research.

Edith Körner *Chairman* Health Services Information Steering Group

Chapter 1 : Introduction

Information use

1.1 Although most national reports about the NHS contain recommendations to collect more data about health service activity, few district managers use information as an essential management tool in their day-to-day tasks. Despite the work of bodies like the NHS/DHSS Health Services Information Steering Group there is still, deep in the NHS management culture, a strong resistance to use even the simplest quantitative techniques.

1.2 Numerous excuses are given by managers to justify their failure to be informed. One of the most common is that the data cannot be trusted. Like most NHS excuses for management inaction, there is some truth in this allegation. There are intrinsic flaws in many of the national rules for data collection which impair the validity of the resulting information. Few district managers have devoted time and thought to the problems of data production.

1.3 However, even the currently collected data can be used to good purpose. With a knowledge of the deficiencies, the current statistics can greatly assist managers to allocate, review and plan the use of resources in their district. The performance of some services is so poor and so far below that of similar services in other districts that a failure to take action must call into doubt the competence of local managers.

1.4 The best way of ensuring that the data produced in a district are believable is to use information based on them locally. It is extremely unlikely that data which are only reviewed at the region or within the DHSS, and have thus been untouched by district

thought, will be credible. Only managers and clinicians, who know how services are organised locally, will spot the nonsenses and institute the appropriate remedial action.

Data quality

1.5 The better the quality of data produced in the district, the more likely it is that information derived from them will be used. The key characteristics of data which need to be considered in assessing data quality are:

a. Validity: the extent to which the data represent what they purport to represent.

b. Accuracy: the extent to which data items have been recorded correctly.

c. Completeness: the extent to which all eligible records and all data items on each record have been captured.

d. Timeliness: the extent to which information derived from the data is available on the timescale required.

1.6 The introduction of the minimum data sets recommended by the NHS/DHSS Health Services Information Steering Group will greatly improve the validity of data collected, particularly those concerning clinical activity. A guiding principle of the Steering Group's work is that the new data sets should reflect the volume of clinical work done and attribute it appropriately. Thus the Steering Group recommends that the SH3 data collection system be abolished, because it has ceased to relate to modern methods of hospital care.

Producing information

1.7 Accuracy, completeness and timeliness of data will be improved

when management thought, time and effort are devoted to developing a district capability for producing information. The process of information production entails:

a. the recording of data,
b. the collation of data,
c. the abstracting and coding of data,
d. data processing,
e. the conversion of data into information, and
f. the presentation of information.

1.8 At each stage of the process, consideration must be given to:

a. the possibilities of introducing information technology (IT);

b. the training and education requirements of the staff involved; and

c. the need for appropriate management arrangements.

1.9 Previous publications in this series have contained proposals which will assist districts to harness the exciting potential offered by recent developments in information technology. The factors to be considered in a district strategy are discussed in *Developing a district IT policy*, while office technology is considered in *Introducing IT in the district office.*

1.10 The key role that training and education must play in information production is stressed in *Piloting Körner*, a document containing the views of the administrators in the four districts which tested the recommendations of the working groups reviewing the data content of the NHS management information systems. National training tools using the media of videofilm and computer assisted learning are being developed by the Steering Group and some of these have already been released.

1.11 The management arrangements that need to be considered are:

a. systems for recording and collating data;

b. mechanisms for ensuring data quality;

c. the establishment of a district information service with the capability of converting data into information; and

d. arrangements for ensuring access to information and IT by all those in the district who require it, subject to confidentiality policies.

1.12 Some of these issues are considered in previous publications in this series, particularly *Converting data into information*, which contains detailed proposals for the setting up of a district information service and an outline of the management arrangements for ensuring the recording and collation of valid clinical data. Recommendations are made about the need for setting and reviewing standards of data quality but no guidance is given about how best these tasks might be done.

Ensuring data quality

1.13 An essential component of the process of producing information is ensuring the quality of the data. This involves:

a. setting standards for data recording, abstracting, coding and processing;

b. achieving the standards set; and

c. monitoring to ensure that the standards set have been achieved.

1.14 It is rare in the management of the NHS for explicit standards of activity to be set. Yet few techniques have as great a potential for improving patient care and motivating staff as a clear description of

the aims and objectives of a service. The production of information is a process which is amenable to the setting of quantitative and qualitative standards and ways of doing this are discussed in Chapter 2.

1.15 In Chapter 3 various methods and techniques which will help achieve data standards are discussed. Although information technology will be used increasingly to validate data, there will be a continuing need to train and motivate staff to record, collate, abstract and code data to the required standards. Essential to the motivation of data collectors is the knowledge that information based on their data is actually being used by senior managers.

1.16 The setting of standards allows the possibility of checking to ensure that they have been complied with, and Chapter 4 contains proposals about both the internal and external audit of health service statistics. Financial data have been audited formally since the inception of the NHS, but there are few arrangements for the regular review of the quality of activity and manpower data, despite the fact that these form the basis for financial budgets and accounts.

1.17 In all the Chapters, the general principles outlined are illustrated by examples drawn from the production of information about clinical activity. The provision of these data is beset by considerable organisational problems in that they are collected by a variety of staff, captured in many different locations in the district and are collated by different staff and at a variety of locations.

Chapter 2 : Setting standards

Introduction

2.1 The setting of data standards involves;

a. defining the characteristics of data for which standards should be set;

b. considering the relationships between the data characteristics;

c. setting standards for each characteristic appropriate to the use which will be made of the data; and

d. considering who should be responsible for setting standards.

Data characteristics

2.2 The key characteristics to be considered in assessing data quality are:

a. validity,
b. accuracy,
c. completeness, and
d. timeliness.

The minimum data sets recommended by the Steering Group will provide, for the first time in the NHS, standards of validity. The existence of prescribed data items with associated definitions and classifications will permit the setting up of arrangements to monitor that these standards are being met.

2.3 The accuracy of data needs to be related to the various stages of the process of producing information, namely:

a. Have data items been recorded accurately?

b. Have data items been abstracted and coded accurately?

c. Have data items already captured been put into a computer accurately?

2.4 The completeness of data should also be related to the different stages of information production and be assessed in terms of both individual data items and of records as follows:

a. Has a record been made of all the events that need to be recorded?

b. Have all the data items required on a record been recorded?

c. Have all the records been collated so that they can be processed?

d. Have all the requisite data items been abstracted and coded?

e. Have all the records to be put into a computer been entered?

f. Have all the data items on each record to be put into a computer been entered?

2.5 The timeliness with which data are required depends on the purposes for which they are to be used. Different items of data are required with varying urgency at different levels of the service. Information about bed use may, for example, be required at:

a. unit level within days of the end of the collection period,

b. district level within a few weeks, and

c. region and the DHSS within months.

Relationships between data characteristics

2.6 In setting standards for any of the individual characteristics of

data, it must be recognised that there are important relationships of:

a. Accuracy and completeness with timeliness.

b. Accuracy, completeness and timeliness with cost.

c. Accuracy, completeness and timeliness with the purpose for which information is required.

2.7 Particularly in the case of a batch process system, the more complete the data, the longer it is likely to be before information is available. For example, the table (below) shows at intervals in 1983 the level of completeness of the 1982 Hospital Activity Analysis data base for two districts. The information user has to choose between using data which are relatively incomplete or waiting until the middle of the year when they are more complete. The better the methods to ensure accuracy the longer may be the delays in obtaining information.

Table Hospital Activity Analysis data for 1982

Completeness of records and clinical details at intervals in 1983

Date of file status report	**Completeness of records %*		**Completeness of clinical details %*	
1983	*District I*	*District II*	*District I*	*District II*
February	99.1	93.6	93.1	86.7
March	99.6	96.0	98.3	87.7
May	99.9	96.7	98.4	94.8

Notes: *Completeness of records containing administrative data expressed as a percentage of the number of discharges recorded on the SH3 return.

**Completeness of clinical details on received HAA records.

Source: HAA file status monitor Oxford Regional Health Authority.

2.8 Ensuring data completeness, accuracy and timeliness is expensive. Omissions and inaccuracies can be reduced but always at some cost. The abstraction and coding of records can be double-checked, key punching can be verified, edit checks and corrections can be made. All these tasks, however, cost money.

2.9 Data which are not sufficiently complete, accurate or timely for a given purpose will not be considered relevant to the problem. If they are not considered relevant, they will not be used and if they are not used, then it is less likely that data will be submitted complete, accurate and promptly in the future.

2.10 On the other hand, some management tasks do not require data which are very accurate, complete or even timely. Long-term planning and the retrospective review of services over long periods of time can be carried out with data which would be totally unsuitable, for example, for the allocation of hospital resources between clinical specialties, or a review of sickness absence among the nursing staff.

Setting standards

2.11 If a system is instituted which requires a record to be raised and a specified list of data items to be collected about particular events, there is obviously an intention to capture all the records and complete accurately all the data items. Those involved in the process from recording the data to putting them into the computer should all be motivated to try and achieve 100 per cent accuracy and completeness and the appropriate timeliness. To set data collectors, coders and others lesser targets will soon destroy the credibility of the data produced.

2.12 To achieve completeness and accuracy may require procedures which are expensive. As in most activities, an inordinate amount of

time and effort may be required to achieve the last 5 per cent. Not all the data items on a record are of equal importance. Some are essential for operational purposes such as the patient number and address, others are less vital for the organisation of patient care but important for management purposes. Although the aim of any routine data collection must be to try and achieve 100 per cent accuracy and completeness, it is *proposed* that, for each data item being collected, intervention levels are set so that, when accuracy and completeness fall below a certain level, this is brought to the attention of those responsible for the system.

2.13 When standards have fallen below the intervention levels set, managers have to decide whether to make existing procedures work better or to change them. This will entail comparing the cost of introducing system improvements with the disadvantages and costs of maintaining the system as it is. The cost of maintaining data that are inaccurate and incomplete are rarely considered in deliberations about improving information systems. If, for example, a district computerised master patient index has a significant number of addresses on it which are incorrect, considerable costs will accrue due to:

a. missed outpatient appointments and hospital admissions;

b. abortive ambulance journeys to the wrong address;

c. multiple registrations and multiple sets of medical records; and

d. correspondence sent to wrong addresses.

2.14 In the Steering Group's First Report to the Secretary of State a minimum data set is recommended to be collected on all patients using a hospital bed. This system provides a useful example to illustrate some of the principles behind setting intervention levels. The data set can be divided into items about:

a. An individual which allow him to be identified, such as date of birth, sex, address and district number.

b. An individual, not necessary for identification purposes, such as marital status and legal status.

c. Referral for admission to and discharge from hospital care.

d. The use of wards and the consultants responsible for care.

e. The diagnoses made and operative procedures performed.

2.15 It is *proposed* that anything less than 100 per cent of records being made available for putting into the computer is unsatisfactory and should be the cause of urgent management action. The fact that an individual has used a hospital bed is not only important for operational purposes and for the patient's future care but will also contribute to the information about how beds are used. Among many uses, information about bed utilisation is included in the RAWP formula and, therefore, missing records may directly affect the revenue allocation to a district. One district in which it was estimated that only 76 per cent of patients had a completed Hospital Activity Analysis form may have been losing 10 per cent of its revenue target annually.

2.16 Personal details which identify the individual are vital to organising the care of a patient as well as being the lynch-pin permitting the linking of all the data held. It is *proposed* that anything less than 100 per cent completeness and accuracy for data items such as date of birth, address, sex and district number is unsatisfactory. These data which are being kept with increasing frequency on a computerised master patient index comprise one of the most important management tools possessed by a district and a failure to maintain their completeness and accuracy can rapidly disrupt the smooth organisation of patient care.

2.17 Personal details not required for identification or operational purposes need not be kept to the same levels of accuracy and completeness. Individual districts must set intervention levels which are appropriate to the purposes to which the data will be used. Particular problems are often found in obtaining data about marital or civil status and an intervention level triggering management action when less than 95 per cent of records have this item completed is not unreasonable.

2.18 Details about admission to and discharge from hospital comprise essential data items such as the dates of admission and discharge and less important items such as the source and method of admission, destination and method of discharge. It is *proposed* that admission and discharge dates be maintained 100 per cent complete and accurate but intervention levels lower than this can be set for the other data items.

2.19 The minimum data set contains a requirement to record all the wards used and the consultants responsible for care during a stay in hospital with the dates of any changes. These data are important for operational purposes as well as for providing essential management information about the use of beds and for specialty costing. It is *proposed* that less than 100 per cent completeness and accuracy for data about wards and consultants is unsatisfactory and requires management action.

2.20 The most difficult items to obtain are diagnoses and operative procedures. Unlike the rest of the data set, the recording is done by doctors in the clinical notes. The task of abstracting and coding the data is skilled and current coding procedures, such as the Ninth Revision of the International Classification of Diseases and the Coding of Operative Procedures produced by the Office of Population Censuses and Surveys, do not necessarily reflect the way clinical work is carried out and recorded.

2.21 It is the often poor quality of the data about diagnoses and operative procedures currently collected in the regional Hospital Activity Analysis that has severely impaired the credibility of this system with clinicians. The procedures to maintain high levels of accuracy and completeness of these data items may be expensive and impose an unwelcome discipline on the doctors who record them. The setting of intervention levels is thus a complex task in which the cost and inconvenience of procedures must be weighed against how the data will be used. An important factor to be considered is the use, if any, which individual clinicians in the district will make of these data in planning and evaluating patient care.

Responsibility for standard setting

2.22 The greatest need for accurate, complete and timely data is within the district. Thus it is *proposed* that each district set intervention levels appropriate to the use which will be made of the information derived from the data. These intervention levels should be publicly stated and be an integral part of a district's information policy.

2.23 Although the district management team and authority will formally endorse the standards set, it is unlikely that they will be involved in the detailed discussion between information users and data producers about what they should be. In a previous publication in this series, *Converting data into information*, the distinction is drawn between staff responsible for producing data and those involved in converting the data collected into management information. The latter tasks should be carried out within a district information service in the charge of an information officer.

2.24 District information service staff are the vital link between information users and data producers. They should know, on one

hand, the information needs of managers and, on the other, the ability of data producers to provide the information required. Occupying, as they do, this central role in the organisation, it is *proposed* that staff of the district information service carry out the detailed work of setting data standards.

2.25 Regions and the DHSS have legitimate reasons for requiring districts to produce data. If, as is unlikely, the standards set by a district are not sufficiently high to meet regional information needs, a region may require such a district to improve its performance. Regional and district staff may wish to agree standards to which all districts in the region should aspire so as to maintain high quality comparative data within the region.

Chapter 3 : Achieving standards

Introduction

3.1 The achievement of high standards of data quality requires time and thought being given to the process of data production. The key tasks are:

a. data recording,

b. data collation,

c. data abstraction and coding, and

d. data processing.

For each of these tasks, consideration needs to be given to the training and motivation of staff, the use of information technology and the organisational arrangements.

Data recording

3.2 Management data are recorded by hundreds of staff in any one district. Clinical activity data, for example, may be recorded by administrative and clerical staff, doctors, nurses and other health professionals. Factors which will help achieve high standards of recording are:

a. ensuring that there are adequate staff and suitable conditions for data recording;

b. motivating and training data recorders;

c. using information derived from the data for operational purposes;

d. using the patient as a data recorder; and

e. recording directly onto a computer with the assistance of computer prompts.

3.3 Because few district managers have, as yet, realised the crucial link between the quality of data recording and the credibility of the management information derived from them, all too often there are inadequate staff for the job, which is frequently carried out in unsuitable conditions. The importance of remedying these defects is highlighted in a previous publication in this series, *Piloting Körner.*

3.4 High standards of data recording are achieved when:

a. Staff know why they are recording data, and are aware of the results of their labours. Few staff are solely concerned with data collection. If the purpose of data recording is not apparent, it will have a low priority with staff who have other work to do.

b. Staff know how to collect data. Few districts currently provide any formal training for clerical staff let alone nurses and other health professionals. The accurate recording of data requires a detailed knowledge of the items to be collected and their definitions and classifications.

3.5 Departmental managers are responsible for the production of valid data about the activity of their department. Thus the responsibility for ensuring that staff involved in recording data know how and why they are doing it falls to the departmental head, be it ward sister, district physiotherapist or operating theatre manager. National training tools are being developed to facilitate the training of data recorders. It is *proposed* that departmental managers ensure that staff recording data about the activity of their department be given formal training for this task.

3.6 Departmental managers are also the major motivating force for the collection of high quality data. If it is known that a departmental head considers data recording a key task and not an irrelevant chore, and informs staff about the value of the data they have collected, data recorders will be encouraged to achieve high standards of completeness and accuracy.

3.7 The more data are used operationally, particularly for the purposes of direct patient care, the more likely that they will be recorded accurately and completely. It is helpful to establish clerical jobs which combine both data recording and the use of the data collected. If staff have no opportunity to see how important data are for the effective organisation of clinical care, then standards of recording are liable to fall.

3.8 Clinical data, such as diagnoses and operative procedures, are recorded in medical notes by doctors. The prime use of such data is for the continuing care of the patient. The accurate and complete recording of these details is thus an essential component of good clinical practice, yet all too often these data items are missing or hidden amidst a wealth of data in an unstructured medical record. Diagnostic data are now used for management purposes. As they are a source of data for the resource allocation formula, districts may lose financially if records are incomplete. It is *proposed* that efforts continue to be made to teach and motivate doctors to keep accurate, structured records for the benefit of their patients, and that this be supplemented by informing doctors about the effect poor record keeping may have on the district's financial allocation and the setting of clinical budgets.

3.9 Many of the data required for clinical care can be recorded by the patient. A number of districts send out questionnaires to patients which are completed before they attend the hospital. This is an excellent method for obtaining data about personal details, the past

medical history, drugs being taken, family and social circumstances. There have been a few experiments in which patients directly put data into a computer. As the population becomes more used to this type of activity, such techniques should be more widely used. Patients should be encouraged not only to act as data recorders but also to assist in checking data already collected. Whenever a patient has contact with the hospital, basic data which might change, such as address, should be checked with the patient.

3.10 Information technology can also assist the clerical recording of patient data. When clerks put data directly into a computer at the time of interview, prompts and edit checks can be built into the software program to assist the data recorder. However, experience in Nottingham has shown that the use of VDU prompts, when data were input after the patient interview, was not as effective as the use of a patient questionnaire in improving the completeness of patient records.

Data collation

3.11 Data collation entails bringing records together to a central point for processing. Most information systems in the NHS are still paper-based and for a system, such as Hospital Activity Analysis, many hundreds of bits of paper have to be collated each week. Unless systems are well designed and organised, documents go missing. The communications capability of modern information technology allows data to be electronically transmitted from the recording source to the collation point. The costs of such systems are decreasing and a district-wide experiment in electronic data collation is under way in the Hammersmith and Fulham District. It is *proposed* that further work be carried out to develop transferable systems permitting the electronic transfer of data from recording sources to collation points.

3.12 The completeness of collation is assisted when data are collected in stages. For example, in the Steering Group's First Report it is recommended that data about patients using a hospital bed be entered on a computer at admission, at each change of ward or consultant, immediately on discharge and when the clinical details have been completed. If any of these parts is missing, its absence will be identified when the various sections are linked in the computer.

Data abstraction and coding

3.13 Currently most data abstraction and coding are done manually. High standards of accuracy require an adequate number of trained staff to carry out the work. This is particularly important for the abstraction and coding of data from clinical notes such as diagnoses and operative procedures. For some years there have been national arrangements for the training of HAA coding clerks. Manual systems are, however, labour intensive and time consuming. Automatic coding of data in the computer will not only be more accurate but should be cost-effective as systems are developed and made more widely available.

3.14 The introduction of a computerised medical record provides the opportunity for both automatic abstraction and coding. At St Chad's Hospital, Birmingham, encounter forms, completed by clinicians, are put on the computer which produces a structured medical record. A data dictionary and a variety of coding structures will allow diagnostic data to be abstracted from the record and automatically coded.

3.15 In Glasgow, diagnoses abstracted from clinical records have been entered on a mainframe computer and automatically coded. Despite the success of this system it has not yet been transferred to other sites. Another time consuming task, the post-coding of

addresses, can also be done automatically but, as yet, few health authorities have used the systems available. It is *proposed* that further work be carried out to develop transferable systems permitting the automatic coding of data entered on the computer. In the processing of data about clinical activity, applications are particularly required for coding addresses, diagnoses and operative procedures.

Data processing

3.16 Two main techniques are used to achieve high standards of accuracy for data entry on a computer, namely:

a. key punch verification, and
b. validation by edit checks.

3.17 Currently most clinical activity data are batch processed on regional mainframe computers. Key punch verification, which entails the duplicate punching of all or selected items on a record, is used by most regions as an aid to maintaining data quality. Increasingly, however, clinical data will be entered into district patient administration systems. As these data are used for operational purposes, the need for key punch verification is considerably less than for batch processed data.

3.18 Edit checks may be used to validate:

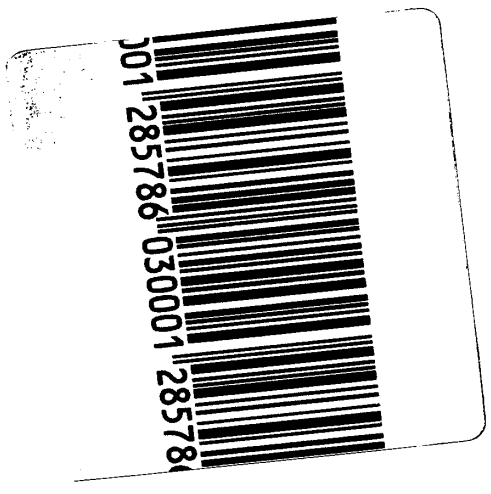
a. Format: for example, detecting an alphabetical character in a file in which only numerical values are possible.

b. Range: for example, checking that values fall within an acceptable range.

c. Logic: for example, rejecting the impossible (male hysterectomies) or warning of the improbable (abortions in women over 50).

Format and range validation are relatively straightforward but logical validation presents the major problem of identifying and accommodating all possible combinations of impossible or improbable events.

3.19 The more stringent the edit checks, the greater will be the resource implications for programming time, computer time, checking of warnings and costs of correction. There will also be delays in information production. When data are input in the district, transferred to region and then to a central government computer, the costs and time delays will be greatly increased if different edit checks are used at different levels. It is thus *proposed* that for data items which are nationally required, edit checks be agreed nationally so that data once put into the system at district level will not be rejected when transferred to region or central government.



Chapter 4 : Monitoring standards

Introduction

4.1 The setting of standards allows the possibility of checking to ensure that they have been complied with. However, the introduction of monitoring systems raises both organisational and technical issues. The two most important questions which need to be resolved are:

a. who is responsible for monitoring standards of data quality, and

b. how should it be done?

Responsibility for monitoring

4.2 Many of the information systems in the NHS involve staff from a variety of disciplines in capturing the data. This is particularly the case with clinical activity systems, but similar problems will occur as district manpower information systems are developed in which data are collected by line managers as well as personnel and finance staff. Even though the staff involved in data production may not be managed as a single entity, it is *proposed* that arrangements be set up to ensure that an individual is responsible for monitoring data standards for each information system in the district. When the data produced is outside the limits prescribed, such an individual will inform the appropriate departmental manager and ensure that advice and help are available to make the necessary improvements.

4.3 For clinical activity systems, district information staff are the vital link between information users and data producers. They should know on the one hand about the information needs of managers and, on the other, the ability of producers to provide the

data required. In Chapter 2 it is noted that they should be responsible for the detailed work of standard setting. It is thus *proposed* that the monitoring of data standards in clinical activity information systems be the responsibility of staff in the district information service. It is not necessary for the monitor of data standards to be the manager of the staff responsible for data production. In fact there may be considerable advantages in organisational arrangements such as those instituted in Brighton Health District, where the district information officer and the district medical records manager are both accountable to the deputy district administrator.

4.4 However good the intra-district arrangements for producing valid data, two possible sources of difficulty have to be recognised, namely:

a. the systematic error which is so ingrained in the data production system that nobody in the organisation identifies it; and

b. the deliberate falsifying of data.

4.5 To counter these problems and to ensure comparability between districts, it is *proposed* that external audit procedures be set up. The external audit should be organised on a regional basis and should involve experienced staff both from region and districts other than that being reviewed. Such an arrangement would also allow the cross-fertilisation of ideas and encourage good data production practices.

Methods of monitoring

4.6 Monitoring data standards involves attempting to identify and quantify errors. Errors may be systematic, in that they occur consistently, or sporadic. They may be due to:

a. Observer error. The fault lies with the individual recording, abstracting, coding or processing the data.

b. Error in the observed. Such errors occur when people give misleading information and are thus of importance in systems concerned with data about individual patients or staff.

c. Instrument error. The fault lies with the rules and processes of the system in which data are recorded, abstracted, coded or processed.

4.7 Two particular techniques are useful in monitoring data standards, namely:

a. regular computer listings identifying incomplete items on the records held; and

b. checks on sample records held on the computer by matching them against source documents.

Checks on sample records may be done as a continuous monitor or periodic survey. The advantage of this technique is that it helps to quantify the level of random error in an information system. However, unless the sample size is very large, it is unlikely to detect systematic errors, such as the consistent miscoding of a diagnosis.

4.8 Although formal methods of monitoring data standards are important, they are no substitute for the best way of verifying data, namely the use of information, derived from the data, for operational purposes and for health service decision making. It is only when information becomes an essential tool used by managers throughout the district that valid data will be produced and managers will be able to believe them.

Appendix

WORKSHOP HELD IN BIRMINGHAM 9-10 FEBRUARY 1984

Attenders

Dr J Clarke	Specialist in Community Medicine, Greater Glasgow Health Board
Mr K Cottrell	Director, Performance Review Team, Mersey RHA
Dr M Goldacre	Director, Oxford Record Linkage Study
Professor A Hedley	Community Medicine Department, University of Glasgow
Mr R Jones	Research Associate, Department of Community Health, University of Nottingham
Mr I Kidson	Consultant Surgeon, Whittington Hospital, London
Dr A Mason	Secretariat of Information Steering Group
Mr M Slattery	Regional Statistician, Wessex RHA
Miss F Stevens	District Medical Records Manager, Brighton DHA
Mrs L Wainwright	Secretariat of Information Steering Group
Mr C West	District Administrator, Portsmouth DHA
Dr G Winyard	District Medical Officer, Lewisham & North Southwark DHA
Mr J Yates	Research Associate, Health Services Management Centre, University of Birmingham
Dr D Young	Consultant Physician, St Chad's Hospital, Birmingham

£1·00