

Article

Three steps to data quality

Krish Thiru MPH

CISR Project Manager, St Mary's Hospital NHS Trust, London, UK

Simon de Lusignan MBBS MRCGP

Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London, UK

Frank Sullivan FRCGP

Professor of Primary Care Research and Development, Tayside Centre for General Practice, University of Dundee, Scotland, UK

Sarah Brew SRN

Practice Nurse

Alun Cooper MRCGP

General Practitioner

Bridge Medical Centre, Crawley, West Sussex, UK

ABSTRACT

Background The quality of data in general practice clinical information systems varies enormously. This variability jeopardises the proposed national strategy for an integrated care records service and the capacity of primary care organisations to respond coherently to the demands of clinical governance and the proposed quality-based general practice contract. This is apparent in the difficulty in automating the audit process and in comparing aggregated data from different practices. In an attempt to provide data of adequate quality to support such operational needs, increasing emphasis is being placed on the standardisation of data recording.

Objective To develop a conceptual framework to facilitate the recording of standardised data within primary care.

Method A multiprofessional group of primary care members from the South Thames Research

Network examined leading guidelines for best practice. Using the nominal group technique the group prioritised the information needs of primary care organisations for managing coronary heart disease according to current evidence.

Results Information needs identified were prioritised and stratified into a functional framework.

Conclusion It has been possible within the context of a primary care research network to produce a framework for standardising data collection. Motivation of front-line clinicians was achieved through the incorporation of their views into the synthesis of the dataset.

Keywords: computerised medical records, core dataset, coronary heart disease, data quality, general practice, primary care informatics, quality improvement

'The higher needs are a later phyletic or evolutionary development ...'

Abraham H Maslow 1970¹

Introduction

This paper outlines a multipractice initiative to clarify the information needs for evidence-based management

of coronary heart disease (CHD) in primary care. It provides a framework that can be used to achieve greater standardisation of computerised data.

The technological capability to automate parts of the clinical governance process and National Service Frameworks (NSFs) have existed for a number of years. However, the lack of consistent and accurate data continues to be the limiting factor preventing the widespread deployment of these technologies. This is because the implementation of evidence-based medicine

requires good communication of health information between carers and managers. Standardised use of codes is a prerequisite for such communication. Consequently, primary care is in need of strategies to promote standardised coding.

Background

Information is the raw material of quality improvement.² The building block of information is data. Computerisation of primary care in the UK has enabled clinicians to collect coded data on their practice clinical information systems (CIS). These systems use variants of the Read coding and classification scheme which is the national standard adopted by the NHS. In the future, Read codes are likely to be replaced by the collaborative initiative between the American Systematized Nomenclature for Medicine and British Clinical Terms (SNOMED-CT).^{3,4} Although CIS are increasingly being used for coding, the data collected are still not likely to meet the forecasted information needs for the implementation of the integrated care record service (ICRS).^{5,6}

Clinical governance and the promotion of evidence-based medicine predict the need for information technology to play a central role in care standardisation. These agendas in combination with the NSFs and integrated care plans represent increasingly visible pressures for change in the modern world of primary care with its public health viewpoints.⁷ In addition, the proposed new quality-based general practice contract is set to highlight the variation in data quality between practices further.⁸

Method

The NHS funds primary care research networks throughout the UK. The South Thames (primary care) Research Network (STaRNet) was created to encourage a research culture within primary care. One of the main objectives of STaRNet was to facilitate the uptake of evidence-based medicine in primary care.⁹ Six STaRNet practices made up the STaRNet Cholesterol Special Interest Group (CholSIG). The group comprised eight general practitioners (GPs), four nurses and one research assistant. They met five times for three-hour meetings to develop a core dataset. The aim of the meetings was to develop a guideline implementation and evaluation strategy directed by five pieces of 'best evidence' which could be facilitated by information and communication technologies (ICT).¹⁰⁻¹⁴ The guidelines were prioritised and discussed through qualitative consensus rather than

quantitative meta-analysis. Clinical management and information needs were defined, using a nominal group technique.¹⁵ Where there was ambiguity about the variables of interest, additional literature was sought through Medline. Where there was disagreement (for example, at what cholesterol level there should be an intervention) the group achieved compromise through discussion.

Following the initial meeting CholSIG members searched their practice databases using in-house clinical query tools for patients with CHD Read codes (diagnosis: codes G4 – 4-byte Read Code Version 1 and G3 – 5-byte Read Code Version 2). Patients were identified through electronic searches and clinicians' knowledge of their practice population.¹⁶ The numbers of patients identified was substantially below that expected. The group decided that a preliminary part of their task was to ensure that future computer searches were productive and that all known cases of ischaemic heart disease (IHD) should be clearly defined and coded. Over the remaining meetings, through discussion and compromise, the group synthesised a core dataset (based on information needs) required for the definition and identification of patients, implementation of evidence-based medicine (EBM) and the evaluation of the implementation process (see Box 1).

Due to a lack of detailed electronic data, information for the actual guidelines implementation and evaluation phase of the study were collected from both computer notes and handwritten medical notes. Since practices had different clinical systems, they elected to code the CHD population with the appropriate practice-specific 'defining' codes. It was optional for practices to update their computer records beyond IHD defining codes with data from the manual records. A reminder card ('trigger') to facilitate identification of patients through their clinical notes was used by two practices. System-specific data collection templates were used to facilitate data collection by one of the practices. All new IHD patients prospectively identified were also to be coded appropriately.

Results

The participating practices were of varied size and distributed across south London and southeast England. They used different clinical systems and Read coding strategies (see Table 1). Age-standardised morbidity levels were substantially below expectation, making the data inadequate to enable computerised guideline implementation and evaluation. Data quality needed to be improved and maintained if automation was to be achieved. The different clinical systems required practices to code as much of their

Box 1 Information collected on patients with confirmed IHD¹⁷

- Demographic information
- Smoking status
- Family history
- Body mass index
- Blood pressure
- Initial cholesterol level
- Latest cholesterol level
- Treatment with aspirin
- Past history of revascularisation treatment
- Health promotion advice
- Treatment with lipid lowering drugs
- Co-morbidity

Table 1 Characteristics of the six practices that took part in this study

Site code	Number of partners	Number of patients identified	SMR %	SE	Population size	Location	Clinical system
A	6	61	15.6	2.0	10500	Twickenham	IPS
B	6	110	26.6	2.5	11100	Guildford	EMIS
C	5	48	17.6	2.5	7300	Battersea	EMIS
D	3	53	28.8	3.6	5500	Crawley	Torex
G	4	84	25.0	2.7	9000	Folkestone	EMIS
H	7	79	16.0	1.8	13200	Thamesmead	Torex

SMR: Standardised morbidity rate (standardisation population from General Practice Research Database – GPRD).¹⁹ SE: Standard error; EMIS: Egton Medical Information System; IPS: In Practice Systems.

core dataset as possible under the correct Read code version. It would be beyond the scope of this paper to clarify all the core codes for the various systems. However, examples are given to facilitate understanding of the proposed framework and show the importance of ensuring appropriate coding (see Table 2).

A data standardisation framework: the three steps

CholSIG generated a list of all information required for the management of established CHD; these are listed in Box 1.

Step 1: 'Netting' the patients – defining the prevalence

The initial focus of the core dataset was to capture general demographic information and set the

inclusion and exclusion criteria for the population under study. This 'netting step' represents the basic requirements in a hierarchy of data needs for guideline implementation. It satisfies the basic need to identify the patients of interest and establish a register.

CholSIG defined the IHD study population objectively as those patients either with confirmed myocardial infarction, a positive resting or exercise electrocardiograph, evidence of cardiac ischaemia on thallium scanning or an angiogram showing coronary atheroma or having received revascularisation surgery. These diagnostic procedures and results were to be Read-coded as in Table 2. Accurate data on patients' sex and date of birth were also associated with this step for registration purposes.

A baseline age–sex profile for the practice was also recorded, so that the prevalence could be calculated for the target population and compared with expected levels and other participating practices.

Table 2 The three-step core data framework required by the three commonest clinical systems in the United Kingdom

Read code version	Clinical information systems and Read codes used	
	EMIS and IPS (5-byte)	Torex (4-byte)
Step 1: Patient identification (netting) data		
Date of birth		
Sex		
Confirmed myocardial infarction (diagnostic codes)	G3...	G4..
Resting or exercise electrocardiograph showing ischaemic changes	3213., 322., 323..	3213, 322., 323.
Evidence of cardiac ischaemia on thallium scanning	5744.	5744
Angiogram showing coronary atheroma	5533.	G451, 5533,
Angioplasty	7928.	771.
Coronary artery bypass graft	792..	7732
Step 2: Patient management data		
Lipid screening results	44O..	44O.
Body mass index	22K..	22K.
Blood pressure status	246..	246.
CHD medication		
Nitrates	bl1., bl2., bl3..	bl1., bl2, bl3.
Lipid lowering drug	bx...	bx..
Antihypertensive drugs	be..., bd...	be., bd..
Salicylate prophylaxis	8B63.	8B63
Aspirin contraindications	8I24., TJ53.	bu23, 8I24,
Anticoagulation prophylaxis	88A5., 14P1..	88A5, 14P1,
Health promotion interventions		
Smoking advice	6791., 9OO..	6791, 9OO.
Dietary advice and referral	6799., 13B3.	8B5., 8B57
Exercise advice and referral	138., 8CA5.	6798, 8H..
Step 3: Disease in a wider context		
Smoking status	137..	137.
Alcohol consumption	136..	136.
Family history of IHD	12C..	12C.
Personal or family diabetic history	C10., 1252.	1228, 1252
Hypertensive disease status	G2..., 14A2.	G3., 14A2
Cerebrovascular disease status	G513., G64..	G7., G74.
Aortic or cerebral aneurysm present	G341., G71..	G82., G452
Peripheral vascular disease	G73..	G86.

Step 2: Managing the patients

The next step set up the data requirements for effective management of the selected population. This considered the patient in the multifactorial environment of the disease under consideration and took account of numerous risk factors and related health interventions. This study paid particular attention to hyperlipidaemia screening results, requiring the patients to be coded as diagnosed with increased lipid levels or by

having direct pathology results recorded so that the 'cut off' points could be assigned during data analysis. The group felt that these data were essential for the study population in light of current evidence. Coded data and, where possible, diagnostic results were also required on body mass index and blood pressure. Due to the importance of lifestyle risk factors for this study group, data on health promotion intervention such as smoking, dietary and exercise advice, either directly

by the clinician or through referral, were included as core management data.

The management step also gave appropriate attention to all therapeutic interventions including nitrate, antihypertensives, lipid lowering drugs, antiplatelet and anticoagulant status.

Step 3: Disease in the wider context; managing the co-morbidity

The apex in our core data hierarchy was the co-morbidity step. This represented a higher evolutionary data requirement. It set the disease under consideration in the wider context of the patient and society. Data relating to personal and familial history of CHD, diabetes, hypertensive disease, cerebrovascular disease, aneurysm and peripheral vascular disease were also collected as they were recognised to be essential for targeting care.

The rationale for the three-stepped approach is set out in Box 2.

This three-step data collection strategy was best appreciated when viewed as part of a clinical governance spiral. Here the evolving information needs of an automated audit spiral are paralleled by improvements in data quality. It was expected that collection of simple prevalence data would happen quickly where data quality was poor, but slow down as the completeness of records increased. Hence the later cycles of the data quality spiral become more tightly coiled as data quality improves (see Figure 1).

Discussion

A multiprofessional group of primary care staff can, in the presence of evidence, collaboratively define and prioritise their information needs for evidence-based care of CHD patients. The proposed conceptual framework for data prioritisation can facilitate improvements in data quality standards within multipractice, multisystems organisations by providing a basis for data prioritisation and standards setting.

All practices involved in this study used clinical information systems but were unable easily to identify the patient group of interest, as required by the NSF. In consequence, complex queries and costly note searches were required for case identification.¹⁸ Identifying and validating patient populations is a prerequisite before progress in data and care quality can be made. The initial assessment of data quality was based on both a comparison of actual IHD code prevalence with the expected number of IHD cases as stipulated by the General Practice Research Database.¹⁹ Although recognised as crude, this can act as a good reference point in assessing coding practice and data quality. However this is more of an indicator of data reliability than of data validity.⁸

Practices and primary care organisations (PCOs) are now required to develop their use of the available information and communication technologies to overcome data quality inadequacies. There has been little non-prescriptive guidance to tackle the data quality issue other than the facilitative approach used by PRIMIS (Primary Care Information Services²⁰) and the I3PC project in Wales delivered by the same team. Recommended coding strategies as suggested by the Sowerby Centre for Health Informatics are central to the national data standardisation initiatives.²¹ A similar solution is being proposed by SCIMP in Scotland (www.show.scot.nhs.uk/). There has been little input from agencies other than national or regional data quality schemes. More proactive development of practice information policies through the synthesis of a core dataset as suggested by Cholsig can empower practices to view data less mechanistically and arrive at a virtually identical core dataset. The approach we describe was helpful for our group; others may benefit from undertaking a similar process of dialogue and engagement in their own context.²² Although GPs and nurses dominate the services provided, the proposed framework is open to input from all professionals within primary care. However, it must be appreciated that clinicians are the prime data gatherers and as such, data being requested from them are more likely to be gathered when directly useful to them.

Data quality improvements should no longer be seen as simply standard setting and performance

Box 2 Guidelines for core dataset synthesis

- Step 1: Patient identification 'netting' data – clearly define the inclusion and exclusion criteria: preferably with diagnostic test results or procedure codes.
- Step 2: Patient management data – focus on the recording of management data. Look at disease management in multifactorial context. Identify which factors are of greatest significance to the patient, to you as a practice and as a PCG member. Attempt to code all data gathered.
- Step 3: Disease in a wider context – set the disease under consideration in the context of a patient as a whole. Give attention to family history, co-morbidity data and poly-pharmacy issues.

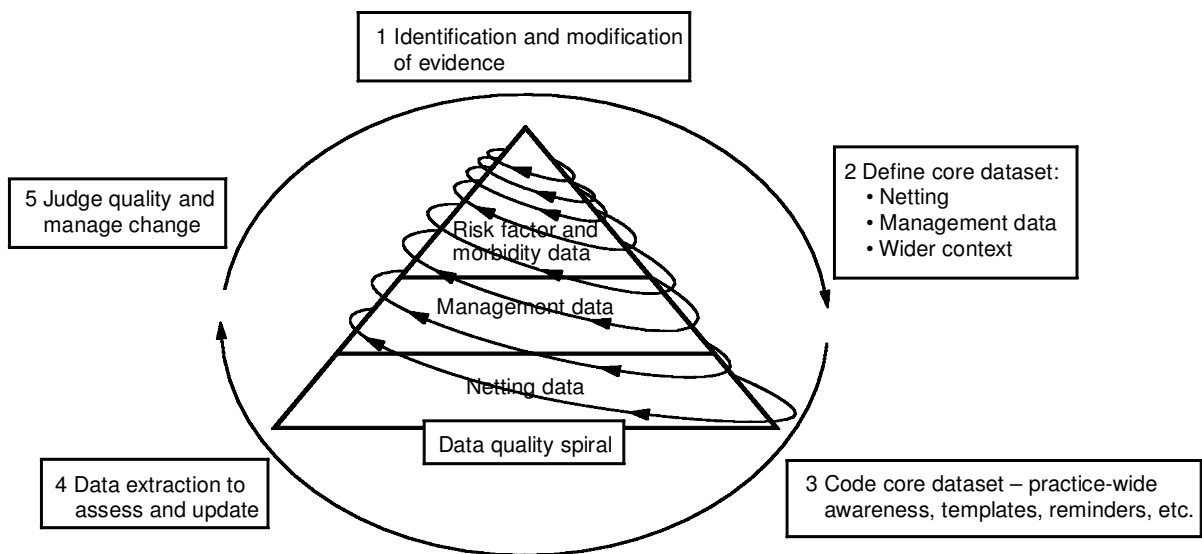


Figure 1 Data quality improvement spiral

measurement.²³ An informatics strategy founded on ruthless standardisation may undermine itself and should be viewed cautiously by informaticians.²⁴ A more holistic approach is needed, one based on a motivation (a recognition of shared information for better healthcare delivery), an opportunity (group work for core dataset development, training, support) and weapons (ICT, quality management tools).²⁵ The role of a collaborative approach to information needs assessment and subsequent data collection requires those concerned with data collection to understand and appreciate the cost benefits of core dataset development. The suggested framework facilitates the balancing of the cost (time and money) for data gathering against data needs and benefits. The audit cycle presented in Figure 1 interlaces with the data quality spiral to act as a continuous (care and data) quality improvement tool with improvements ceasing when costs outweigh the benefits.

The recommended synthesis process requires a select multiprofessional group to work (directed by current evidence) under some guiding principles of data quality management. These are:

- a *commitment* to change
- a recognition of the *cultural* requirements
- to *communicate* effectively
- to work *co-operatively*
- to be mindful of *confidentiality*.

The group can then establish what data (preferably coded data) are required to meet the information needs of the collective. Under quality management philosophies, these five Cs are central principles that integrate the technical, human and cultural elements of an

organisation that affect quality management. Here data quality management is seen as a continuous process (see Figure 1).²⁶ In this context, quality assurance gives us a succinct but practical definition of quality as ‘fitness for purpose’.²⁷ The purpose of computerised primary care data is rooted in the information needs of PCO members and their commitment to clinical governance. Nevertheless, for the data to be shared, aggregated or communicated it needs an acceptable level of standardisation.

Data collection templates, coding reminders and ‘pull down’ code menus have been found to be partially useful in standardising data collection.²⁸ More importantly there should be a recognition of the reasons for the standardisation of data collection and adherence to standards. Finally it must be noted that all core datasets, though structured, must continue to respond to new medical evidence and societal pressures. It is envisaged that the proposed three-step framework is transferable to other chronic disease areas where there is a requirement for an agreed dataset that satisfies the needs of several different groups of data users.

Conclusion

The only way that the NHS will know whether it has delivered its clinical ambitions, as set out in *The NHS Plan* and its supporting NSFes and *Cancer Plan*, is if it can rely on the quality of data collected in primary care.^{29,30} Although national strategies are undoubtedly effective in increasing awareness and improving data

quality, motivation to collect data cannot be imposed from above. The collaborative generation of disease registers with a hierarchically synthesised core dataset can provide a synergistic 'bottom up' approach to becoming data conscious. The suggested framework can empower practices to tackle the issue of data quality to a greater extent for themselves. Improving data quality and the quality of clinical care can both then become part of what Maslow would have described as the 'self-actualisation' of clinical information systems.¹

ACKNOWLEDGEMENTS

Thanks to the staff of the six STaRNet general practices that took part in this study for their help in collecting the data: Mrs S Brew, Dr B Christie, Dr A Cooper, Dr D Goodwin, Dr J Gray, Dr C Kroll, Mrs K Masters, Prof J Majeed, Dr J Oxenbury, Mrs J Sawyer, Mrs S Wells and Dr D Wheeler. K Thiru is supported by Fisher Medical Centre Research Unit (FMC-RU) under Northern and Yorkshire NHS Executive R&D funds. Simon de Lusignan is Director of KSSnet (Kent Surrey and Sussex Primary Care Research Network – one of the successors to STaRNet – supported by NHS DHSC South). A Cooper and S Brew are KSSnet Foundation Members.

REFERENCES

- Maslow AH. *Motivation and Personality*. San Francisco: Harper Collins, 1970.
- Moss F. Spreading the word: information for quality. *Quality in Health Care* 1994;3(Supp):46–50.
- Sowerby Centre for Health Informatics at Newcastle and University of Newcastle. *Prodigy Project: computerisation in general practices (1996 survey)*. Newcastle: Sowerby Centre for Health Informatics, 1998.
- O'Neil M, Payne C and Read J. Read codes Version 3: a user led terminology. *Methods of Information in Medicine* 1995;34:187–92.
- Gilley J and Majeed A. Meeting the information and budgetary requirements of primary care groups, a commentary: accurate information may be difficult to produce. *British Medical Journal* 1999;318:168–70.
- Pringle M, Ward P and Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. *British Journal of General Practice* 1995;45:537–41.
- Department of Health. *National Service Frameworks*. London: Department of Health, 2002. www.doh.gov.uk/nsf/nsfhome.htm (accessed 01.01.03).
- Thiru K, Hassey A and Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *British Medical Journal* 2003;326:1070.
- Kendrick T. *South Thames Research Network Project Plan – STaRNet Academic Team internal document*. London: St George's Hospital Medical School, Department of General Practice and Primary Care, 1999.
- Shepherd J, Cobbe SM, Ford I *et al*. Prevention of coronary heart disease with pravastatin in men with hypercholesterolaemia. *New England Journal of Medicine* 1995;333:1301–7.
- ASPIRE Steering Group. A British Cardiac Society survey of the potential for the secondary prevention of coronary disease: ASPIRE (action on secondary prevention through intervention to reduce events). *Heart* 1996;75:334–42.
- MeReC Briefing. Hyperlipidaemia. *MeReC Briefing 10*. Liverpool: National Prescribing Centre, 1996.
- Which? Ltd. Management of hyperlipidaemia. *Drug and Therapeutics Bulletin* 1996;34:89–93.
- Department of Health Standing Medical Advisory Committee. *The Use of Statins*. London: Department of Health, 1997.
- Jones J and Hunter D. Qualitative research: consensus methods for medical and health services research. *British Medical Journal* 1995;311:376–80.
- Moher M, Yudkin P, Turner R, Schofield T and Mant D. An assessment of morbidity registers for coronary heart disease in primary care. ASSIST (ASSESSment of Implementation STRategy) trial collaborative group. *British Journal of General Practice* 2000;50:706–9.
- Thiru K, Gray J and Majeed A. Management of ischaemic heart disease in primary care: towards better practice. *Journal of Public Health Medicine* 1999;21:179–84.
- Gray J, Majeed A, Kerry S and Rowlands G. Identifying patients with ischaemic heart disease in general practice: cross sectional study of paper and computerised medical records. *British Medical Journal* 2000;321:548–50.
- Office of National Statistics. *Key Health Statistics from General Practice 1998: analysis of morbidity and treatment data, including time trends, England and Wales*. Series MB6 No 2. 2000. London: Office of National Statistics, 2000.
- Primary Care Information Services (PRIMIS): www.primis.nhs.uk/ (accessed 01.01.03).
- Cowley C. *Developing the CHD Knowledge Base: Part 1 The Read Clinical Terms and Virtual Registers*. Newcastle: Sowerby Centre for Health Informatics, 2001.
- Wenger E. *Communities of Practice: learning, meaning, and identity (learning in doing)*. Cambridge: Cambridge University Press, 2000.
- Taylor FW. In: Caulcutt R (ed) *Achieving Quality Improvement: a practical guide*. London: Chapman and Hall, 1995.
- Department of Health. *Delivering 21st Century IT support for the NHS: national specification for integrated care records service – consultation draft*. London: Department of Health, 2002.
- Øvretveit J. What is quality in health services? *Health Service Management* 1990;86(3):132–3.
- Oakland J. *Total Quality Management, the Route to Improving Performance*. Oxford: Butterworth Heinemann, 1993.
- Ellis B. Identification of the critical success factors involved in the implementation of clinical governance arrangements within primary care. *Informatics in Primary Care* 2002;10:19–23.
- Tai SS, Nazareth I, Donegan C and Haines A. Evaluation of general practice computer templates. Lessons from a pilot

randomised controlled trial. *Methods of Information in Medicine* 1999;38:177–81.

29 Department of Health. *The NHS Plan: a plan for investment, a plan for reform*. London: Department of Health, 2000.

30 Department of Health. *The NHS Cancer Plan*. London: Department of Health, 2000.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Mr K Thiru
St Mary's Hospital NHS Trust
IM&T Department
Praed Street
London W2 1NY
UK
Tel: +44 (0)20 7886 2283
Email: krish.thiru@st-marys.nhs.uk

Accepted July 2003