Commentated case study

Info-tsunami: surviving the storm with data quality probes

Philip JB Brown MD MRCGP
Honorary Lecturer in Healthcare Informatics, School of Information Systems, University of East Anglia, Norwich, and Humbleyard Practice, Hethersett, Norfolk, UK

Victoria Warmington RGN BSc
Research Associate, Humbleyard Practice, Hethersett Surgery, Norfolk, UK

ABSTRACT

As a result of the rapid expansion of electronically available clinical knowledge, clinicians are faced with potential information overload (info-tsunami). The use of data quality probes (DQPs) in primary care can encourage clinicians’ awareness of, and improvement in, data quality entry over time. DQPs can also highlight areas of potential error or omission as well as good practice, which can impact directly upon the quality of patient care. In this paper, five specific conditions have been subjected to the use of a series of DQPs over a five-year period in order to assess and measure the performance of different initiatives on the quality of data capture and patient care.

Keywords: continuous quality improvement, data quality probes, electronic health records

Introduction

This paper outlines the threat clinicians, and the patients in their care, face by being swamped with information (info-tsunami). There is a danger that the plethora of knowledge and guidelines may overwhelm an individual’s stamina to rationally apply best practice across a population in their care. Articulate patients may themselves be able to monitor their care to ensure that the appropriate review and tests are being done at the correct intervals and their treatments adjusted. However, in order to deliver care equitably to all, and particularly to those potentially at greatest risk who may not be capable of alerting clinicians when omissions may have occurred, methods are required to assist healthcare providers in evolving towards continuous quality improvement (CQI) in the health care of populations.

This study reports a method of monitoring and promoting data quality in a primary care setting that has been applied over the last four years. It utilises the principle of data quality probes (DQPs) to encourage clinicians to improve both data quality entry and the care of patients. It explores whether initiatives orientated towards improving the care of patients with ischaemic heart disease (IHD), diabetes and asthma have been successful, and compares this with short-term quality initiatives (recording patients with glaucoma) and conditions that have not been the focus of any quality initiative (breast cancer).

Info-tsunami

The rapid expansion in information technology has supported an escalation in published knowledge in health care. For example, the number of publications relating to IHD has risen from 2930 in 1970 to 9098 in 2001, a trend mirrored in diabetes, breast cancer, asthma and glaucoma (see Figure 1).

This expansion of both researched and published knowledge has fostered the philosophy of evidence-based practice with an associated boom in guidelines and digital libraries. Many valuable approaches have been developed in improving access to knowledge, such as info-buttons and decision support services (for example, PRODIGY). However, despite these applications, the average clinician is in danger of being swamped with this ever-increasing tidal wave of information relentlessly gathering volume and pace – the information tsunami (info-tsunami). A key
challenge for informatics is to assist clinicians in converting knowledge into practice equitably across organisations and populations.

**Background**

**The encounter system**

The provision of health care during an encounter between a patient and a clinician involves a number of processes:

- **assessment** – eliciting and assessing a patient’s symptoms and signs
- **data entry** – recording of the data into a patient’s record (which may be a complex of electronic data, written records, letters, results, etc)
- **data retrieval** – extracting data for interpretation
- **knowledge and guidelines** – being available and aware of their existence
- **information interpretation** – governance of interpretation of individual patient data using existing knowledge and guidelines
- **action** – performing a therapeutic or investigative action on a patient in response to that information.

These interrelated processes can be seen as part of an overall **encounter system** in which the clinician–patient encounter operates (see Figure 2). In order to provide quality health care, all these processes and their interconnections need to be performed flawlessly. When examining the encounter system, an error at any of the above process steps, such as data entry error, lack of knowledge or failure of information processing, could have the consequence of sub-optimal care. In order for clinicians to take advantage of electronically assisted healthcare provision, the data within the electronic record has to be accurate. Whilst this is acknowledged, little has been published on methods of assessing accuracy and completeness of electronic health records.

**Data quality probes**

The principle of DQPs involves the posing of a query in a clinical information system (which operates within an encounter system) where the result can be used as a measure of the performance of that system. The specific metrics used also share the feature that clinical knowledge would determine that the associations between two (or more) data items are perfect (complete union in set theory). In other words there is strict concordance of the association between one data item and another. For example, all patients with diabetes would be expected to have estimation of HbA1c performed at least annually; in an ideal system, a DQP to find all cases of diabetes with no record of an HbA1c would retrieve no cases. The reasons for these DQP errors can be due to a clinical error where the knowledge should have been applied but was not; or an error of entry where the data was either not entered or recorded incorrectly.
The serial application of a panel of DQPs can be used to feed back information to clinicians on their performance on a regular basis. This can be done by circulating to users of a system a list of cases that have been identified by the DQP. These lists can be focused to complete organisations, parts or even individuals. For example, the non-recording of PEFR (peak expiratory flow rate) during an asthma review performed in the previous six months can be broken down to individual clinicians. This method has been employed in the Humbleyard Practice to focus clinicians on specific quality initiatives to improve both data quality and patient care. Usually clinicians are aware of the facts but need assistance in putting them into practice consistently and completely.

Method

The Humbleyard Practice provides primary care to a population of approximately 16 000 patients in a mixed environment of rural, semi-rural and town with 11 doctors, seven nurses and accompanying administrative and managerial staff.

The prevalences of five conditions (IHD, diabetes, asthma, glaucoma and breast cancer) were identified from the clinical information system (Healthy Software) in a semi-rural primary care practice in Norfolk, United Kingdom (UK), for five consecutive years (1998–2002). The former three conditions had been subject to CQI initiatives using a variety of techniques, including DQP feedback, educational meetings and regional financial incentive schemes. Glaucoma had been subject to a ‘blitz’ initiative when resources had been invested over a short period of time to improve data quality; breast cancer had not been subject to any quality initiative during the time period. A series of sequential DQPs were performed relating to these five conditions as proxy measurements of data quality and care over five consecutive years (see Box 1).

The results of the prevalence of each condition and the applicable DQP are expressed as a ratio of the expected prevalence using published national statistics adjusting for changes of population size over the time frame.

Results

The observed prevalence of IHD, diabetes and asthma over the time period increased (see Figures 3, 4 and 5). The figures present the observed prevalence as a ratio of that expected for the population (in an average population the prevalence would be 1). The observed prevalence of IHD remains below 1 compared to that of diabetes which is around 1 and asthma which has exceeded the predicted prevalence; these findings are in keeping with previous clinical studies suggesting the Norfolk region of the UK has a below-average prevalence of IHD, average prevalence of diabetes and an above-average prevalence of asthma. The DQP concerning the recording of the PEFR (see Figure 5) has shown a continued improvement over the period with a gradual tendency towards a ratio of 1 (perfect); cases outstanding are a mixture of clinical omissions and patients managed predominantly in secondary care. A similar finding is apparent in relation to diabetes (see Figure 4), with a residual subgroup with no HbA1c recorded being predominantly type 1 diabetics under secondary care review. Patients with IHD (see Figure 3) show a fairly constant sub-optimal performance of the DQP which on further investigation relates the opinion of some clinicians (not unreasonably) that annual cholesterol levels in patients on a statin with stable levels is unnecessary and their interval of review can be extended.

The recording of glaucoma patients was subject to a ‘blitz’ quality initiative during Spring 2002; all patients receiving certain medication were reviewed and, where they were found to have glaucoma, this was then recorded. Figure 6 illustrates the sudden increase in observed prevalence and reduction of DQP to 1.

<table>
<thead>
<tr>
<th>Box 1 Data quality probes applied each year using the prevalence of recorded cases in that year as a denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>IHD</td>
</tr>
<tr>
<td>Diabetes</td>
</tr>
<tr>
<td>Asthma</td>
</tr>
<tr>
<td>Glaucoma</td>
</tr>
<tr>
<td>Breast cancer</td>
</tr>
</tbody>
</table>
Following this initiative there have been no further reminders or feedback of DQP results circulated, resulting in an upwards drift in the DQP ratio, indicating a gradual deterioration of data (leading to an under-recording of new cases).

Finally, in the case of breast cancer (see Figure 7), both the observed prevalence and the DQP have changed little over the last five years (no tendency to achieving a ratio of 1). This suggests that the quality of data in this case is poor.

**Discussion**

There is a rapid expansion of clinical knowledge and this is now becoming available electronically in the form of digital libraries and electronic protocols and guidelines. Mechanisms of linking these potential knowledge sources to clinical records and decision making are possible in support of evidence-based medicine. Although these sources are available to users, some recent studies have questioned the efficacy of such availability in practice. There are many reasons for this, including awareness, knowledge and aptitude of users as well as the usability of implementations. Many clinicians have been numbed by the info-tsunami and are having difficulty in keeping their heads above water. DQPs have been shown to identify potential data quality issues and like other initiatives such as PRIMIS, have been used to encourage migration towards a culture of data accuracy and consistency that supports care rather than another
The application of serial DQPs described here has indicated that this is a valuable method in:

- focusing clinicians on quality targets that need to be achieved
- providing timely and regular feedback to clinicians on their performance
- encouraging continued conformance to quality standards.

Thus in the domain of IHD, diabetes and asthma, the measures indicate gradual improvement in both coverage (approaching average prevalence) and consistency in achieving low levels of DQP errors. These improvements have been sustained by regular feedback of DQP queries. This contrasts with glaucoma data which, having been ‘perfect’ during a concerted effort, is now beginning to degrade without further feedback of performance. This suggests that clinicians are task-orientated and require ongoing prompts to consistently improve the quality of data recording and health care.

Conclusion

Info-tsunami is a problem. Without incentives and encouragement, clinicians will forget or lack motivation in recording and maintaining data. DQPs are useful in illustrating the trends in quality of data over time. If used sequentially they are valuable in monitoring data and clinical care quality; with the results fed back to clinicians they act as valuable prompts in promoting quality and maintaining improvements.

ACKNOWLEDGEMENTS

The authors would like to thank the clinicians and staff of the Humbleyard Practice, in particular the Clinical Governance Task Force (Cyril Lindsey, Siobhan Rowe, Jos Harwood, Pam Brantigan, Sharon Dack and the administrative staff of the Humbleyard Practice) for continued efforts in responding to data quality initiatives.

REFERENCES

3 PRODIGY: www.prodigy.nhs.uk
7 Healthy Software: www.healthysoft.com
9 Primary Care Information Services (PRIMIS): www.primis.nhs.uk

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Dr Philip JB Brown
School of Information Systems
University of East Anglia
Norwich NR4 7TJ
UK
Tel: +44 (0)1603 810250
Fax: +44 (0)1603 812402
Email: phil@hicomm.demon.co.uk

Accepted October 2003
Commentary

Simon de Lusignan
Senior Lecturer, Primary Care Informatics,
Department of Community Health Sciences,
St George’s Hospital Medical School, London, UK

Sheila Teasdale
Service Director, Primary Care Information
Services (PRIMIS), University of Nottingham,
Nottingham, UK

What is a DQP and how is it used?

A data quality probe (DQP) is a useful tool in the armamentarium of those seeking to raise data quality standards and the quality of clinical care. The core of the method is the development of a series of two or more clinical recordings which should or should not be found together, for example prescribing penicillin to a person with penicillin allergy. Table 1 shows how two-item DQPs might be constructed.

Brown and Warmington’s paper examines five DQPs. Four relate to therapy that should be accompanied by a diagnosis or clinical measurement, and one to a clinical measurement that should always be associated with a diagnosis. The DQP can be used in areas where there is a good evidence base for treatment and where therapy or clinical measurement is only used for that condition. The hypothesis proposed is that feedback of the DQP has resulted in the reported change in data quality.

Where does the DQP fit in the data quality literature?

The paper rightly embraces many of the features of an effective data quality intervention reported elsewhere in the literature.

Focus on reliability and validity

The literature

Thiru et al. performed a systematic review of the data quality literature relating to primary care. They reported that there should be data quality reference standards for computerised records. These should give an index for reliability (such as comparison of prevalence of disease with national figures) and that objective items (such as prescriptions and test results) should be used for their positive predictive value and sensitivity.

This paper

The authors explain their recorded prevalence in terms of regional trends (reliability) and use the DQP in a way that should have both a high sensitivity and positive predictive value as to whether a patient has the condition in question (for instance, patients with a heart disease diagnosis who are taking a statin should have a high chance of being patients with heart disease, and it is unlikely that they have another diagnosis).

Table 1 Examples of two-item DQPs

<table>
<thead>
<tr>
<th>Data item B</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Present</strong></td>
<td></td>
</tr>
<tr>
<td>A + B must be present, e.g. patients with hypertension must have had a blood pressure (BP) measure within a time period</td>
<td>A is present + B is absent, e.g. patients with penicillin allergy should not be prescribed penicillin</td>
</tr>
<tr>
<td>Probe: patients with hypertension and no BP measure within the time period</td>
<td>Probe: patients prescribed penicillin who are allergic</td>
</tr>
<tr>
<td><strong>Absent</strong></td>
<td></td>
</tr>
<tr>
<td>A is absent + B is present, e.g. patients without heart disease should not be taking long-acting nitrates</td>
<td>A and B are mutually exclusive, e.g. men recorded as having had a hysterectomy</td>
</tr>
<tr>
<td>Probe: patients on long-acting nitrates with no record of heart disease</td>
<td>Probe: non-females with a recorded hysterectomy</td>
</tr>
</tbody>
</table>
Practice-based learning and national comparative analysis service

The literature
The PRIMIS service (www.primis.nhs.uk), developed from research carried out in the late 1990s, offers in-depth learning about data quality and information management in practices, with one local information facilitator working with around 20 practices. Its emphasis is on teaching facilitators to help practices improve their information management skills in order to raise the quality of patient care. This approach includes DQPs, though within PRIMIS these are described as ‘data quality queries’. These data quality queries include looking at prescriptions usually associated with a condition but where the diagnosis is not recorded (for example, insulin without a diagnosis of diabetes mellitus), and age- and gender-related data quality problems (children with Alzheimer’s, men with hysterectomies, etc). These have many of the features of data quality probes and are fed back graphically, but in a different way from that represented in this study.

This paper
It would appear that the authors, or a member of their team, took on the role of ‘PRIMIS facilitator’ across the surgeries, creating an appropriate non-judgemental learning environment within which practitioners were happy to share data and receive feedback. It would be useful to know the facilitator time needed to achieve and sustain change. It is of interest that they report that when active feedback ceases, the quality of the data degrades, suggesting that data quality work needs to be ongoing rather than a ‘one-off’ process.

Audit-based education

The literature
The Primary Care Data Quality (PCDQ) Programme focuses on audit-based education at the level of the primary care organisation. Practitioners are encouraged to understand inter-practice variation in their data using age–sex profiles and other data. It has been described as ‘action-epidemiology’. The study only feeds back a small volume of data at a time, waiting until it ‘saturates’ before moving on.

This paper
The authors mention that feedback is used as part of postgraduate education, and provided to colleagues within their workplace.

Completeness and currency score, and phronesis

The literature
Williams has reported on the importance of ‘completeness and currency’ of computer data. He observed a degree of consensus as to what clinicians feel should be reported in diabetes. Experienced clinicians tend not to blindly follow rules that they don’t see as important. The concept of ‘phronesis’—knowledge born of practical experience— is a concept that was first taught by Aristotle. It is less valued in today’s evidence-based world where ‘episteme’ (science) rather than ‘techne’ (art) predominates.

This paper
The clinicians in the study practice appear be making a statement about the completeness and currency of the cholesterol measurement. Their unwillingness to make what they thought was an unnecessary additional measurement of lipid profiles is an example of them exercising phronesis.

Financial incentives

The literature
There were considerably improved levels of data recording when targeted financial rewards were introduced into UK general practice. Primary care is now preparing to meet more quality targets set out in its new contract. Progress towards these targets will be measured by extracting structured data from general practice computer systems, and financially rewarded.

This paper
The authors report how financial incentives were also part of the picture.

Limitations of the study

It is not clear which factors contributed to the change reported

Inevitably this intervention, like the other contributions to data quality research— including those by the authors of this commentary— is multifactorial. How important each element is cannot be concluded from these data.
It is known that some of the factors, that form components of this programme, are more or less likely to contribute towards the positive outcomes shown. Passive dissemination of information is known not to change practice, and education is an appropriate change agent in effective data quality work. A detailed study of feedback to practices showed that fewer than half the items fed back could be shown to change data quality faster than background change.

Usability where there are not ‘signature’ medications or clinical measurements

DQPs clearly can be applied in areas where the medical intervention is a good proxy, such as the use of insulin for the treatment of diabetes mellitus. However, there are many more areas where such DQPs will be much more difficult to construct, for example beta-blockers are a class of drugs that can be used in a wide range of conditions.

It is unclear whether DQPs would be useful at all levels of data quality

The practice reported has high data quality standards and the DQPs yielded useful information in this context. If DQPs were used in a practice with very little coded data recorded, it is possible they would achieve ‘better’ results. There is a danger that if they were the only means of quality assessment that the use of codes like IHD might be restricted to patients who were definitely on a statin and had their cholesterol recorded.

Positive predictive values and sensitivities could be calculated for each DQP

Gray et al. have shown that it is possible to calculate for a simple search strategy the positive predictive value, sensitivity and yield rates. If these were supplied it would be possible for practitioners to pick simple or complex probes to use, according to their current perceptions of what was required.

Knowledge management may be a better approach for the tidal wave of knowledge (info-tsunami)

There may be better strategies for coping with knowledge overload. A knowledge management strategy may be a better approach than focusing on DQPs.

Conclusions

The DQP is a useful tool for improving data quality in primary care. More tools are needed because even in straightforward areas of medicine, the gap between what is on the computer records and what is needed in practice is massive.

The paper as a whole characterises the features of an effective data quality programme:

1. Informatics is a discipline whose principles and practice can be deployed to enable quality improvement.
2. Data quality programmes need to motivate primary care professionals to value the structured Read-coded data recorded on their computer system.
3. Clinicians need to be shown that it is feasible and realistic to use their data to monitor progress towards implementing evidence-based quality improvement programmes.
4. A data quality programme needs to respect the clinical judgement (‘phronesis’) of experienced clinicians.
5. Feedback using parameters likely to have a positive predictive value and high sensitivity are likely to be effective in improving data quality.
6. Educational interventions provided from trusted institutions or individuals independent of management form an effective change agent in primary care.
7. Alignment of data quality initiatives with local and national evidence-based quality improvement programmes increases their chance of success.
8. Individually tailored feedback from a skilled facilitator in the workplace increases its effectiveness.
9. A data quality initiative is more likely to succeed if led by a local champion.
10. Financial incentives are useful in helping to promote and sustain interest in data quality.

Further research is needed to establish the relative weighting required. Meanwhile, DQPs should be added to the data quality tool kit.

REFERENCES


5 Horsfield P. Trends in data recording by general practice teams: an analysis of data extracted from clinical computer systems by the PRIMIS project. *Informatics in Primary Care* 2002;10(4):227–34.


**CONFLICTS OF INTEREST**

Simon de Lusignan is Programme Director of the PCDQ (Primary Care Data Quality) Programme, and Sheila Teasdale is Service Director of PRIMIS (Primary Care Information Services).

Email: slusigna@sghms.ac.uk or sheila.teasdale@primis.nottingham.ac.uk