Data quality improvement in general practice

HJ Brouwer, PJE Bindels and HC Van Weert


Background. The importance of routine data generated by GPs has grown extensively in the last decade. These data have found many applications other than patient care. More attention has therefore been given to the issue of data quality. Several systematic reviews have detected ample space for improvement of data quality. A new review was conducted in order to find out which methods of improvement are effective.

Method. The Medline database was searched using an iteratively composed set of terms and MeSH (Medical Subject Headings) headings. Only papers that focused on explicit attempts at improving data quality of medical records in general practice were included.

Results. Twelve studies met the inclusion criteria. No study used patient-based comparison of records with external sources as the method to assess data quality improvement. Ten studies used internal indicators or markers of data quality instead. Attempts at data quality improvement often involve some sort of individualized feedback, and nearly all attempts seem to have some positive effect. Only one of the included studies fulfilled the basic methodological requirements of an intervention study. The most recent studies used a simple before–after design.

Conclusion. No intervention to improve data quality has been put to a rigorous enough test. We still lack empirical knowledge as to how improvement can be brought about.

Keywords. Family practice, general practice, primary health care, research.

Introduction

Groups of GPs recording medical data for research purposes have a long tradition. The aims and scopes vary, so do the number of practices involved and the methods used.1–5 The number of these networks has exploded in the last decade.2,3 Three factors have contributed towards this development: first, a growing awareness of the potential of primary health care data for public health policy and post-marketing surveillance6–8; second, the evolution of basic tools, such as widely accepted primary health care classification and coding systems9,10; and finally, the emergence of the computerized patient record (CPR) that has made it possible to record data more easily and continuously with little or no disturbance of daily practice.11–13

Data quality in general practice records has been studied regularly. Two recent systematic reviews summarize data quality in general practice computerized medical records.14,15 Many included studies are assessment studies and are of a snapshot nature; they present a momentary picture of data quality. From these reviews, it becomes clear that data quality of medical records assessed in terms of completeness and correctness varies greatly, depending on the item of focus, the diagnosis studied and the reference standard used for comparison. There seems to be ample space for improvement, although no criteria have been suggested or developed as to what are acceptable levels of data quality, and when is ‘good’ or ‘high’ ‘good enough’ or ‘high enough’.

Received 18 July 2006; Revised 19 May 2006; Accepted 28 June 2004.
Division of Clinical Information and Public Health, Department of General Practice, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. Correspondence to HJ Brouwer, Division of Clinical Information and Public Health, Department of General Practice, Academic Medical Center, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands; E-mail: h.j.brouwer@amc.uva.nl

© 2006 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Furthermore, within the framework of a group of GPs dedicated to continuously recording data for research purposes, assessment is only one (although necessary) element within a continuous cycle of quality assurance. Depicting the factors that influence data quality and designing quality improvement programs seem at least of equal importance. Hence, we asked ourselves the following questions. Within the context of general practice,

- which methods of data quality improvement have been used?
- which methods have been proven effective?

To answer these questions we performed a review of the literature.

Methods

Search strategy

Our aim was to identify and review the literature on the improvement of data quality in general practice. As there is no MeSH (Medical Subject Headings) heading for ‘data quality’, a search strategy was developed iteratively using keywords and MeSH headings of key papers on the subject. The result was a search strategy that could include all of the key papers. The following terms were used to search the Medline database until publication date 31 December 2005: (MH & TW: general pract* OR family pract* OR primary care) AND (MH & TW: medical records) AND (TW: improve* OR control* OR assur*). We did not confine our search to studies of CPRs because we were curious about the similarities or differences in comparison with quality assurance studies of paper patient records. References of collected papers were searched as well. The papers identified through the search were downloaded from Medline into a Reference Manager database.

Results of search strategy and selection of articles

A total of 1541 papers were identified as potentially relevant. Titles and abstracts were scanned by the first author in Reference Manager using four criteria: (i) empirical study in the general practice setting (ii) that made use of the medical record and (iii) in which an intervention was attempted to improve data quality of the medical record and (iv) where the results of at least two measurements of data quality in time were presented. On the basis of title and abstract 1464 papers could be excluded. The remaining 77 papers were retrieved and read by the first author to decide on inclusion. When in doubt, the second author was consulted and a joint decision was made. Excluded were non-empirical discussion papers, quality of care improvement studies using medical records, single assessment studies without intervention and studies on the quality of information and communication systems (hardware and software) in general practice (see Fig. 1). Thus, in total 12 studies were included in the review. These studies are described in Table 1.

Results

Methodology of included studies

Only one of the 12 included studies fulfilled the basic methodological requirements of an intervention study: control group, randomized allocation of intervention. Two studies used a control group without randomizing the intervention. All three are among the earliest of the 12 studies. Two of the three controlled studies dealt with the improvement of paper records. Seven studies used a before–after design without a control group. Two studies appeared not to have aimed at an intervention from the outset. They merely seem to have analysed medical record data at several moments in time and named certain activities as interventions retrospectively. One paper did not reveal whether paper or computerized records were involved.

Interventions

In only two studies the interventions were described with enough detail to replicate, to compare across studies or to apply to a different setting. Nearly all studies (on paper as well as on computerized record improvement) seemed to have used some sort
### Table 1  
**Studies included in the review, methodology and results**

<table>
<thead>
<tr>
<th>Study</th>
<th>Nationality</th>
<th>(n) of GPs</th>
<th>Record</th>
<th>Design</th>
<th>Interval before/after</th>
<th>Intervention</th>
<th>Control group</th>
<th>Assessment</th>
<th>Outcome</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Gilliland et al., 1992¹⁷</td>
<td>UK</td>
<td>(5 GPs study) (4 GPs control)</td>
<td>CPR</td>
<td>Before–after, controlled</td>
<td>1, 3 months after each intervention</td>
<td>Three successive interventions: Informal individual meeting (10 min), no feedback. Informal individual meeting (10 min), with feedback (comparison with study and control group). Formal group meeting with feedback (comparison with study and control group).</td>
<td>No information</td>
<td>Completeness Recording (%) of: 1. Symptoms 2. Problems, diagnoses 3. Investigations</td>
<td>Effectiveness of A greater for 2.</td>
<td></td>
</tr>
<tr>
<td>3. Del Degan et al., 1996¹⁸</td>
<td>CAN</td>
<td>45</td>
<td>Paper</td>
<td>Before–after 6 weeks</td>
<td>Individual feedback (comparison with study and control group), instructions</td>
<td></td>
<td>Completeness % prescriptions with: 1. Concentration 2. Dosage 3. Quantity or length of time 4. Number of renewals</td>
<td>Significant effect on 1, 2, 3 and 4.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Nationality</td>
<td>(n) of GPs</td>
<td>Record Design</td>
<td>Interval before/after</td>
<td>Intervention</td>
<td>Control group</td>
<td>Assessment</td>
<td>Outcome</td>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>------------</td>
<td>---------------</td>
<td>-----------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>5. Thiru et al. 1999&lt;sup&gt;20&lt;/sup&gt;</td>
<td>UK</td>
<td>2</td>
<td>CPR</td>
<td>Before–after 5 years</td>
<td>No systematic intervention (. . . received training in Read nomenclature and feedback on morbidity recording)</td>
<td>Completeness</td>
<td>1. Percentage prevalence rates for 10 diagnoses 2. Percentage of alcohol/tobacco use 3. Percentage prevalence rates based on medications for four conditions.</td>
<td>Improvement for 1, 2 and 3.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Hiddema et al., 2001&lt;sup&gt;21&lt;/sup&gt;</td>
<td>NL</td>
<td>3 group practices</td>
<td>CPR</td>
<td>Before–after 6 months</td>
<td>Update of GP information system</td>
<td>Completeness and correctness</td>
<td>Percentage of missing indications of four medication groups Percentage of relevant indications of four medication groups</td>
<td>Decrease of missing indications for chronic medications. Slight increase of relevant indications. Significant improvement on all four items.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Code</td>
<td>Data</td>
<td>Duration</td>
<td>Setting</td>
<td>Methodology</td>
<td>Data Quality</td>
<td>Findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>----------</td>
<td>---------</td>
<td>-------------</td>
<td>--------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Brami and Doumenc, 2002</td>
<td>F</td>
<td>244</td>
<td>Paper/CPR</td>
<td>Before–after</td>
<td>6 months</td>
<td>Feedback consisting of comparison with regional/national scores</td>
<td>Completeness</td>
<td>Significant improvement on all 13 items.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. De Lusignan et al., 2002</td>
<td>UK</td>
<td>576</td>
<td>CPR</td>
<td>Before–after</td>
<td>1–9 years</td>
<td>Feedback on data quality markers. Specific feedback to below-average score (9 months)</td>
<td>Completeness</td>
<td>Significant improvement with time on 6, 7 and 8. Significant improvement of specific feedback to below-average scorers on 7.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Nationality (n) of GPs</td>
<td>Record Design</td>
<td>Interval before/after</td>
<td>Intervention</td>
<td>Control group</td>
<td>Assessment</td>
<td>Outcome</td>
<td>Results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------</td>
<td>---------------</td>
<td>-----------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of individual feedback to improve data recording, either with \textsuperscript{17,18,23} or without \textsuperscript{19,20,22,24–27} a comparison with group performance. Group meetings of recording GPs and training sessions were reported as well. \textsuperscript{17–19,22,25,27} Another study attempted to measure the effect of an update of a GP information system on data quality.\textsuperscript{21} One study used a self-audit system to improve paper recording.\textsuperscript{16}

### Assessment of data quality

None of the studies complied with the criterion Hogan and Wagner\textsuperscript{26} formulated for the assessment of completeness and correctness of (computerized) medical records: an external reference standard for comparing the medical records with. Instead, in 10 studies internal indicators or markers of data quality were formulated as items of the medical record considered most essential (i.e. % of diagnoses coded, % of allergies recorded), and then a change in that indicator was used as an outcome measure.\textsuperscript{16–19,21–24,26,27} Two studies used even less direct measures of data quality: a before–after change in prevalence rates and a comparison of prevalence rates with criterion prevalence rates.\textsuperscript{20,25}

All studies were directed at improving the completeness of medical records. In two studies the correctness of medical records was targeted as well.\textsuperscript{21,25}

When completeness was measured, the items concerned various aspects of patient-oriented information, but often included the presence of medical diagnosis or assessment\textsuperscript{16,17,19,22–24} and/or prescriptions.\textsuperscript{21,24,27}

### Effect of interventions

All studies did report improvement on at least some outcome measures. In the non-controlled studies the effect of interventions could, of course, not be disentangled from a possible Hawthorn effect. In the three controlled studies, interventions did improve the (completeness of) reporting of, amongst other items, diagnoses or assessments, while the control group did not show an improvement in time.\textsuperscript{16,17,19}

Two of these used audit\textsuperscript{16,19} to improve the quality of paper records; one combined several different types of interventions\textsuperscript{17} to improve the quality of computerized records. The latter study showed that an informal meeting without feedback improved the completeness of diagnosis recording better than meetings with feedback. However, as the interventions were cumulative, it could simply have been a case of the first intervention cutting the deepest.

There is some evidence that the extent to which diagnoses are coded can be sustained and improved over rather long periods of time: 2–9 years.\textsuperscript{24,26} In these studies the interventions, although insufficiently described, seemed to consist of a continuous cycle of assessment, feedback and training.

While the studies did report effects, no conclusion went beyond the reporting of statistical significance; none of them commented on the absolute level achieved, and whether the improvement was large enough.

### Conclusions

Our review showed that data quality improvement studies in general practice are few and very often not up to the standard of intervention study methodology. Most attempted interventions are described insufficiently and aimed at improving adherence to ‘norms of good record keeping’. As far as could be deciphered it seems that some sort of confrontation with one’s own recording effort has been the most prevailing intervention. This feedback is sometimes followed by instruction/training sessions. However, this type of intervention has not been put to a sound methodological test.

Striving for completeness seems the first broad step in improving the quality of general practice records. Only when the quality of recording for specific diseases or diagnoses is studied is correctness targeted as well. Assessment of both completeness and correctness has only been indirect, and not patient-based.

Keeping these methodological shortcomings in mind, it seems certain that record keeping is a skill that can be influenced, though it is not clear whether it is time, being part of a registration effort or an intervention that brings it about.

The background of the practices studied were diverse: explicit registration and research networks,\textsuperscript{20,21,24,26} separate group practices, health centres and teaching practices,\textsuperscript{16,18,22,23,25,27} and participants in a national morbidity study.\textsuperscript{19} The 12 studies had their origin in 7 different countries, the 4 most recent ones were from the United Kingdom. All this shows how the importance of the subject is considered irrespective of the various uses of the medical record in general practice. The implications of this issue seem nevertheless most urgent for networks of practices where continuously data are recorded for research purposes. Here, the consequences of over- and under-recording of diagnoses, for instance, can be calculated respectively in terms of extra efforts to include cases for studies and the bias due to selection of more serious or outstanding cases. The benefit of quality improvement can then be clearly weighted.

Interestingly enough, the interventions used to improve computerized medical records have not been drastically different than those used with paper records. In just one instance was a change in software studied as drastically different than those used with paper records. In just one instance was a change in software studied as

In our final search strategy we did not use the terms ‘data’ and ‘quality’, nor the MeSH headings ‘quality assurance’, ‘quality control’ or ‘quality indicators’. In our iterative approach we found that inclusion of these terms and headings sharply increased the number
of articles to be screened, but still missed some key studies. The combination of terms ‘data’ and ‘quality’ draws large numbers of studies where medical record data are used to study aspects of quality of care.

It has been found that our method of using primarily one reviewer to decide on inclusion of studies may lead to 8% of relevant literature being missed. In our search, this would mean that ultimately one other relevant study has been missed.

Given all the attention for data quality in general practice and for the space and need for improvement, it seems surprising that so few studies have been dedicated specifically to the topic of quality improvement. More rigorous testing is clearly needed, with adherence to the following standards:

- a sufficiently detailed description of the interventions
- a before-after control design,
- outcomes measured in terms of both completeness and correctness
- assessment of completeness and correctness using independent patient-based reference measures.

Funding to pay the Open Access publication charges for this article was provided by Academic Medical Center, University of Amsterdam.

References


