International Journal of Technology Assessment in Health Care, **21:1** (2005), 138–143.
Copyright © 2005 Cambridge University Press. Printed in the U.S.A.

Can electronic routine data act as a surrogate for patient-assessed outcome measures?

Hayley A. Hutchings, Wai-Yee Cheung, John G. Williams
University of Wales Swansea

David Cohen, Mirella F. Longo
University of Glamorgan

Ian Russell
University of Wales Bangor and *Velindre NHS Trust*

Objectives: There has been a rapid growth in the use of patient-assessed outcomes (PAOs) that are measured in the assessment of health technologies. The process of collection of such measures can be costly, and there may be problems associated with the ability of the patient to complete them. The use of electronically stored routine data may reduce costs and overcome the problems associated with patient completion. The feasibility of using routine data surrogates for the UK Inflammatory Bowel Disease Questionnaire (UKIBDQ) and the Short Form 36 (SF-36) was examined.
Methods: Clinical terms and codes for the UKIBDQ and SF-36 questions were identified, and data from electronic routine sources were sought on patients participating in a randomized controlled trial. The presence or absence of relevant symptoms was used to generate surrogate scores, which were compared with the original scores.
Results: Most questions in the UKIBDQ and SF-36 were codable but only one third of the terms were recorded routinely in electronic form. The surrogate total IBDQ score had reasonable reliability (Kuder–Richardson coefficient = 0.51), but this reliability could not be determined for the SF-36. Intraclass correlations between routine and designed data were poor to weak.
Conclusions: Although electronic routine data sources had the capacity to develop surrogate measures for patient assessed outcomes, there was evidence of wide underutilization of coding systems leading to an underreporting of symptoms. This finding is consistent with previous literature where only poor correlations were illustrated between patient assessed outcomes and surrogate scoring of symptoms.

Keywords: Patient assessed outcome, Inflammatory bowel disease, Routine data, Feasibility

There has been a rapid growth in the number of patient-assessed outcomes (PAOs) (15), and the research community is increasingly recognizing the importance of assessing PAOs in addition to clinical outcomes in health-care evaluation. However, as it forms part of the data collected for primary research, evidence of changes in PAOs is costly to obtain. Moreover, the benefit of these data tends to remain confined to the individual studies that generated them, that is, they are rarely used in routine clinical practice (16).

The amount of electronic stored data of patients' information has escalated over the past 20 years (8).

We are very grateful to the colleagues who allowed us to re-analyse their original study using routine data and to Jayne Morgan, Hayley Dickinson, Tracey Hughes, Teryl Stoneman, and Julia Demery for their coding, technical, and administrative support.

The study was funded by the National Health Service Health Technology Assessment National Research and Development Programme. The views and opinions expressed do not necessarily reflect those of the National Coordinating Centre for Health Technology Assessment.

https://doi.org/10.1017/S026646230505018X Published online by Cambridge University Press

Administrative databases are a relatively inexpensive source and patient outcomes can be tracked longitudinally for years, because these systems allow linking by means of a medical record, social security number, or other identification number for each episode of care (29). Thus, allied with increasing concerns about the quality and costs of health care, researchers have started investigating if patient outcomes resulting from health-care activities could also be systematically retrieved by querying these data (3;4;12;17;18;25). The information requested of routine data is shifting from data related to processes of care and patient outcomes such as mortality and morbidity to data related to more complex measures of health status (e.g., physical, mental, social, emotional, and role functioning) (3;4).

In England and Wales, much administrative, demographic, and clinical data are collected about patients in electronic form in the course of their health care. In the future, the quality and quantity of these data will increase (11;28) with a need to capture patient outcomes routinely along with it (10). Currently, most routine clinical systems in the United Kingdom have the provision to record diagnostic and procedure codes with some more elaborate systems, allowing the recording of patient symptoms and signs. There would thus appear to be considerable potential to examine the capacity of routine data systems to develop surrogate PAO measures and, hence, reduce the need for designed primary research and the costs associated with it.

PAO questionnaires aim to ascertain the well-being of the patient based on patient completed responses to either symptom-specific or generic health questions. A single generated score or several subscales can then give an indication of the "general health" (for generic questionnaires) or the "disease-specific" health (for disease-specific questionnaires) of a patient. This presupposes that the patient is able to complete the questionnaire, but may for various reasons be unable to do so (1). It has been suggested that, in some cases, a health-care professional or caregiver may be used as a proxy to provide useful information on PAOs, especially on observable aspects of PAOs (1). Furthermore, the patient's perception of health and its meaning will vary between individuals and within an individual over time (6).

Given these difficulties, as well as the increasing requirement to capture patient outcome and experience routinely (10), there is a need to explore whether routinely collected clinical data can be used as a surrogate for subjective patient-focused data collection. This potential, however, is dependent on the extent to which routine data are capable of generating valid condition-specific and/or generic PAOs. The assessment of PAOs using routine data should also give careful consideration to the validity of the data (3;4).

To assess the potential of the available electronic routine data in measuring PAOs, we have examined the possibility of developing routine data surrogates for one disease-specific [United Kingdom Inflammatory Bowel Disease Questionnaire, UKIBDQ (7)] and one generic [Short Form 36, SF-36 (5;27)] PAO measure.

METHODS

A previously completed pragmatic randomized controlled trial on patients with inflammatory bowel disease seen in two UK sites (Neath and Morriston) (32) was reanalyzed for this study using electronically held routinely collected data. This approach was in the context of a wider study on the use of routine data for health technology assessment (33). Design and analysis were undertaken by a statistician who was not involved with the original study but proceeded according to the principles set out in the protocol for the original trial.

Electronically held routine data were found in local and national systems (33). One of the study sites uses a generic clinical information system, GeneCIS® (31), on which coded clinical data are collected in detail to support individual patient management. Electronically held routine data were also sought from primary care. Each data source was queried to extract all the data for each of the study patients from the date of recruitment to the study end date.

Appropriate clinical terms and codes for the questions in the PAO instruments used in the original study [UKIBDQ (7) and SF-36 (27)] were identified by a clinical coding specialist using International Classification of Diseases version 10 (ICD-10) (30), Office of Population Censuses and Surveys version 4 (OPCS4) (23), and Clinical Terms version 3 (CTV3) (22) coding systems (see Tables 1 and 2 for examples). These codes representing relevant symptoms, signs, diagnoses, and procedures in the UKIBDQ and SF-36 were then sought in the data obtained from routine sources.

Surrogate subscale and total UKIBDQ scores and subscale SF-36 scores were generated by adding up the presence of relevant symptoms. The calculated baseline surrogate scores were reversed and transformed (20) to make them comparable with the original UKIBDQ and SF-36 scores obtained (32). They were then compared with the scores obtained from the original study using intraclass correlation (9). Kuder–Richardson coefficent (KR-20) or Cronbach's α values (26) were calculated for the surrogate UKIBDQ and SF-36 scores to determine the reliability of the new scales.

RESULTS

Patients studied at the site without the clinical information system GeneCIS® had limited information available, thus making any comparisons meaningless. The results are based on ninety-three patients from the site where data from GeneCIS® and departmental data systems were available.

Table 1 illustrates examples of UKIBDQ comparable coded data from electronic routine data sources. Of the original thirty inflammatory bowel disease questions, twenty-nine

https://doi.org/10.1017/S026646230505018X Published online by Cambridge University Press

Table 1. Examples of Some Original UKIBDQ Questions with Comparable Coded Symptoms, Signs, and Diagnoses from All Electronic Routine Data Sources

Original question (Subscale)	Comparable ICD10 code	Comparable CTV3 Code
Q1. On how many days over the last two weeks have you had loose or runny bowel movements? a) None b) On one or two days c) On three to seven days d) On eight to fourteen days (BOWEL I)	b)–d) R19.4- "Change in bowel habit"	b)–d) 19F2 Y30c8- "Diarrhea", 19EA.Y7EfC- "Change in bowel habit"
Q2. On how many days over the last two weeks have you felt tired? a) None b) On one or two days c) On three to seven days d) On eight to fourteen days (SYSTEMIC)	b)–d) R53- "Malaise and fatigue (includes tiredness)"	b)–d) Xa96S Y7DO0- "Tiredness", R0075- "Tiredness", 168Z- "Tiredness symptom NOS, 1682- "Fatigue", XE0qk Y7DO7- "Tired all the time"
Q3. In the last two weeks have you felt frustrated? a) No, not at all b) Yes, some of the time c) Yes, most of the time d) Yes, all of the time (EMOTION)	None	b)–d) Ua164 Ya5sa- "Feeling frustrated"
Q4. In the last two weeks, has your bowel condition prevented you from carrying out your work or other normal activities? a) No, not at all b) Yes, for one or two days c) Yes, for three to seven days d) Yes, for eight to fourteen days (SOCIAL)	b)–d) Z73.8- "Other problems related to life management difficulty"	b)–d) ZV4KB Ym9n2(0){V} – "Other problems related to life management difficulty"

To generate surrogate scores, for the above questions, 0 = symptom not recorded, 1 = symptom recorded. A lower UKIBDQ score indicates worse health. The surrogate measures were reversed and transformed to adopt the same convention, i.e., a lower score indicating worse health.

had a comparable CTV3 code and twenty had an ICD-10 code. However, the surrogate symptoms were only recorded routinely for ten of them. A standardized item KR-20 of 0.51 was calculated for the surrogate total IBDQ score. KR-20 scores could not be calculated for the UKIBDQ subscales due to the limited number of recorded symptoms in the subscales. The surrogate scores were consistently higher than the original UKIBDQ subscales and total scores. Intraclass correlations between the original UKIBDQ and surrogate scores were poor to weak (see Table 3).

Table 2 illustrates examples of the SF-36 comparable coded data from electronic routine data sources. Of the original thirty-six SF-36 questions, thirty-four had a comparable CTV3 code and thirty had an ICD-10 code. The surrogate symptoms, however, only were recorded routinely for ten of them. Cronbach's α or Kuder–Richardson scores could not be calculated for any of the SF-36 surrogate subscales due to the nonrecording of many of the SF-36 surrogate symptoms. The surrogate SF-36 subscale scores were generally higher than the original SF-36 subscales scores, with the exception of vitality and mental health. Intraclass correlation between the original SF-36 and surrogate scores were poor to weak (see Table 3).

DISCUSSION

This study was a retrospective exploratory study. It explored a unique approach to the measurement of PAOs. This strategy has been possible because of the availability of detailed clinical data that were routinely collected and stored in GeneCIS®. This information system uses a clinically rich terminology for coding (Clinical Terms version 3 (22)), which includes many terms that correspond to the questions in the UKIBDQ and SF-36. More "standard" hospital administration systems record diagnoses only, using ICD-10 (30). Our experience also suggests that data collected on departmental systems under clinical ownership are more reliable than those collected through processes under managerial control (33). This evidence is consistent with the literature (13;14;24). Although many of the questions in the UKIBDQ and SF-36 questionnaire have a corresponding term and code in ICD-10, these are clinically unwieldy and rarely recorded in coded

https://doi.org/10.1017/S026646230505018X Published online by Cambridge University Press

Table 2. Examples of Some Original SF-36 Questions with Comparable Coded Symptoms, Signs and Diagnoses from All Electronic Routine Data Sources

Original question (Subscale)	Comparable ICD10 code	Comparable CTV3 code
Q4. During the past 4 weeks have you had any of the following problems with your work or other regular daily activities as a result of your physical health a) cut down on the amount of time you spent on work or other activities b) accomplished less than you would like c) were limited in the kind of work or other activities d) had difficulties performing the work or other activities (for example, it took extra effort) (ROLE – PHYSICAL)	a)–d) Z73.6- "Limitation of activities due to disability" (Score 1) Symptom not recorded (Score 0)	a)–d) Xa844 YaVcu- "Difficulty performing laboring activities" (Score 1) Symptom not recorded (Score 0)
Q5. During the last 4 weeks have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious) a) cut down on the amount of time you spent on work or other activities b) accomplished less than you would like c) didn't do work or other activities as carefully as usual (ROLE – EMOTIONAL)	a)–c) R45.8- "Other symptoms and signs involving emotional state" (Score 1) Symptom not recorded (Score 0)	a)–c) Xa844 YaVcu- "Difficulty performing laboring activities", X00SO YM3yH- "Depressed", XE0rb Ya5sj- "Feeling anxious" (Score 1) Symptom not recorded (Score 0)

A lower SF-36 score indicates worse health. The surrogate measures were reversed and transformed to adopt the same convention, i.e., a lower score indicating worse health.

form. Routine systems that do not use a clinically rich terminology, therefore, may miss information that could be useful in measuring PAOs.

Table 3. Intraclass Correlation between the Original and Surrogate Scores for the UKIBDQ and SF-36

Item		Intraclass correlation
UKIBDQ	Bowel I	0.04
	Bowel II	0.22
	Emotion	0.16
	Systemic	0.04
	Social	No surrogate available
	Total	0.08
SF-36	General health	0.00
	Social function	0.04
	Vitality	0.25
	Mental health	0.27
	Bodily pain	0.26
	Physical function	No surrogate available
	Role physical	No surrogate available
	Role emotional	0.07

UKIBDQ – United Kingdom inflammatory bowel disease questionnaire; SF-36 – Short Form 36.

Questions that had a corresponding code were generally scored as present if the patient exhibited the code or absent if the code was not present. One flaw with this approach is that it cannot be assumed that a patient did not have a particular symptom just because it was not recorded in the system. Also the presence of a recorded symptom gives no indication of the level of severity of that symptom (29), and the symptoms recorded in records cannot generally substitute for the response to a refined and validated questionnaire (12).

Only ten of the thirty UKIBDQ and ten of the thirty-six SF-36 available codes were ever used in the electronic routine data sources in this study. This finding reflects routine clinical practice and the ad hoc collection of data in the conventional outpatient setting. Routine data, therefore, will consistently underrecord symptoms that might contribute to the assessment of PAOs, even when captured in a clinically rich information system. This finding is reflected in the weak to poor intraclass correlations between the surrogate and original UKIBDQ scores. Our findings are in agreement with published literature that highlights the discrepancy between patient and providers perception of general health status (3;4). Although correlations between the surrogate and

https://doi.org/10.1017/S026646230505018X Published online by Cambridge University Press

original scores were only poor to weak, the surrogate total UKIBDQ scale was shown to be fairly reliable with a Kuder–Richardson value of 0.51. This compares reasonably well with the UK validation study (7). Although symptoms were only recorded as present or absent, it appears that the clinical information system GeneCIS® could be useful in measuring changes in PAOs in disease-specific conditions such as inflammatory bowel disease.

In contrast, although three of the surrogate SF-36 subscales showed a weak correlation with the original SF-36 items, the Cronbach's α or Kuder–Richardson values could not be calculated, as symptoms were absent for the large majority of questions. Because the SF-36 questionnaire is generic, this finding may account for the limited number of questions with symptoms recorded as present. Retrospectively routine data systems appear to have less value in assessing changes in generic PAOs after interventions than condition-specific changes.

Due to the retrospective nature of this study, some codes may have been missed due to extraction difficulties or problems with the systems at the time of the study. There may also be differences in coding procedures and codes used at different study sites that may affect the result of PAO testing with routine data (29). In a prospective trial, it would be possible to identify and record routinely the presence or absence of all symptoms and signs that are normally recorded in disease-specific or generic questionnaires. Context-specific, structured proformas for data collection have been shown to improve performance (19). This approach could eliminate the need for patient-administered questionnaires and the associated difficulties such as patient comprehension and poor response rates (1).

This study used data from a generic clinical information (GeneCIS®), which can be customized for use in different clinical contexts. Symptoms and signs can be routinely recorded in coded form, as well as diagnoses and procedures. It also has the facility for local set up of structured questionnaires, which facilitate the capture of multi-item scores (PAOs, patient preferences, disease severity, disability, activities of daily living, and so on). Such systems are not presently widely available but will be needed to meet the specification of requirements published by the Academy of Medical Royal Colleges (2). If the specification is met, patient-focused outcomes could be measured by proxy across a broad range of clinical conditions and in the course of routine care.

If routinely collected data are used for prospective randomized trials, the data will be recorded by those not blind to the treatment given, who may also have a vested interest in the success of the experimental regimen. Thus, a randomized trial using routine data might fail to comply with the CONSORT guidelines (21). The validity of such adaptations in measuring PAOs needs to be explored further. In addition, for greatest accuracy, outcome data should be captured as close to the source as possible and are best reported by the patients' themselves. Systems, therefore, should allow the patient to report these outcomes directly (35). Validated measures (such as the UKIBDQ or SF-36) could be routinely completed by patients when attending for consultation (18) or admission and held on a clinical system such as GeneCIS®. This strategy may address some of the issues concerning the standardization of use of PAO measures (15) and the process of linking clinical variables with health-related quality of care (34) but may not overcome the difficulties associated with patient completion (1). The feasibility and cost of this process should be explored.

CONCLUSIONS

This exploratory study has shown that, where clinical data are collected in detail in structured form, it may be possible to develop surrogate measures for PAOs. Where only diagnoses and procedures are collected, this development does not appear to be possible. This inability is due, not to the lack of available codes, but due to the underutilization of the terms and codes within the routine data systems. Both the surrogate UKIBDQ and SF-36 scores were consistently higher than the original scores, indicating an underreporting of symptoms. To develop good surrogate measures for PAOs, the consistent, accurate, and complete collection of routine data is required.

CONTACT INFORMATION

Hayley A. Hutchings, PhD, Lecturer (h.a.hutchings@swan.ac.uk), **Wai-Yee Cheung**, PhD, Senior Lecturer (w.y.cheung@swan.ac.uk), Centre for Health Improvement Research and Evaluation, Swansea Clinical School, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, Wales
John G. Williams, Professor of Health Services Research (j.g.williams@swan.ac.uk), Clinical School, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, Wales
David Cohen, BCom, MPhil, Professor of Health Economics (dcohen@glam.ac.uk), **Mirella F. Longo**, PhD, Research Fellow (mflongo@glam.ac.uk), School of Care Sciences, University of Glamorgan, Treforest, Pontypridd CF37 1DL, Wales
Ian Russell, PhD, Director (ian.russell@bangor.ac.uk), Institute of Medical and Social Care Research, University of Wales Bangor; Professor of Public Health, National Public Health Services for Wales, Velindre NHS Trust, Wheldon Building, Bangor Gwynedd LL57 2UW, Wales

REFERENCES

1. Addington-Hall J, Kalra L. Who should measure quality of life? *BMJ*. 2001;322:1417-1420.
2. Academy of Colleges Information Group. *Specification of core requirements for clinical information systems in support of secondary care*. Available at: www.aomrc.org.uk; Accessed: 2001.

https://doi.org/10.1017/S026646230505018X Published online by Cambridge University Press

3. Bakken Henry S, Lush M, Costantino M, et al. Health status measurement in computer-based patient record systems. *Nurs Admin Q*. 1997;21:50-60.
4. Bakken Henry S, Morris J, Holzemer W. Using structured text and templates to capture health status outcomes in the electronic health record. *J Qual Improv*. 1997;23:667-677.
5. Brazier J, Harper R, Jones N, et al. Validating the SF-36 health survey questionnaire: A new outcome measure for primary care. *BMJ*. 1992;305:160-164.
6. Carr A, Gibson B, Robinson P. Is quality of life determined by expectations or experience? *BMJ*. 2001;322:1240-1243.
7. Cheung W-Y, Garratt A, Russell I, Williams J. The UK IBDQ- A British version of the inflammatory bowel disease questionnaire: Development and validation. *J Clin Epidemiol*. 2000;53:297-306.
8. Cole L, Houston S. Integrating information technology with an outcomes management program. *Crit Care Nurs Q*. 1997; 19:71-79.
9. Deyo R, Diehr P, Patrick D. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials*. 1991;12:142S-158S.
10. Department Of Health. *The NHS Plan— a plan for investment, a plan for reform*. London: Stationary Office; July 2000.
11. Department Of Health. *Building the information core: Implementing the NHS plan*. London: Department of Health; 2001.
12. Dolin R. Outcome analysis: Considerations for an electronic health record. *M.D. Computing*. 1997;14:50-56.
13. Galland R, Magee T, Berridge D, et al. Accuracy of centrally recorded OPCS codes for vascular surgery in the United Kingdom. *Eur J Vasc Endovasc Surg*. 1998;16:415-418.
14. Galland R, Whatling P, Crook T, Magee T. Regional variation in varicose vein operations in England 1989-1996. *Ann R Coll Surg Engl*. 2000;82:275-279.
15. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick. Quality of life measurements: Bibliographic study of patient assessed health outcome measures. *BMJ*. 2002;234:1417-1419.
16. Higginson I, Carr A. Using quality of life measures in the clinical setting. *BMJ*. 2001;322:1297-1300.
17. Johnson R, Hornbrook M, Nichols G. Replicating the chronic disease score (CDS) from automated pharmacy data. *J Clin Epidemiol*. 1994;47:1191-1199.
18. Johnston R. Implementation of a computer-based patient record and an outcomes data-collection system at the Department of Orthopaedic Surgery, University of Iowa. *J Bone Joint Surg Am*. 2000;82:1502-1506.
19. Mann R, Williams J. Standards in medical record keeping. *Clin Med*. 2003;3:329-332.
20. Medical Outcomes Trust. *SF-36 Health Survey Scoring Manual for English Language Adaptations*. Australia/New Zealand, Canada, United Kingdom. Boston, MA: Medical Outcomes Trust; 1994.
21. Moher D, Schulz K, Altman D. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191-1194.
22. National Health Service Information Authority. *Clinical terminology and classification services*. Available at: http://www.coding.nhsia.nhs.uk.
23. Office of Population Censuses and Surveys. *Tabular list of the classification of surgical operations and procedures*. 4th revision. London: HMSO; 1990.
24. Pollock AM, Vickers N. Reducing DCO registration through electronic matching of cancer registry data and routine hospital data. Br J Cancer. 2000;82:712-717.
25. Rietman L, Hendricks H. Speaking the same language: Improving evaluation of patient-focused outcomes using standardized terminology. *Top Health Inform Manage*. 2000;20: 35-43.
26. Streiner G, Norman R. *Health measurement scales. A practical guide to their development and use*. Oxford: Oxford University Press; 1995.
27. Ware J, Sherbourne C. The MOS 36-item short form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483.
28. Welsh Assembly Government. *Informing healthcare- towards a knowledge-based NHS*. Cardiff: Welsh Assembly Government; 2002.
29. Whitman G. Outcomes research. Getting started, defining outcomes, a framework, and data sources. *Crit Care Nurs Clin N Am*. 2002;14:261-268.
30. World Health Organization. *International statistical classification of diseases and related health problems*. 10th revision. Geneva: World Health Organization; 1992.
31. Williams J, Morgan J, Howlett P, Severs M. Let there be light. *Br J Healthcare Comput*. 1993;10:30-32.
32. Williams J, Cheung W, Russell I, et al. Open access follow up for inflammatory bowel disease: Pragmatic randomised trial and cost effectiveness study. *BMJ*. 2000;320:544-548.
33. Williams J, Cheung W, Cohen D, et al. Can randomized trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment. *Health Technol Assess*. 2003;7(26).
34. Wilson I, Cleary P. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995;273:59-65.
35. Zielstorff R. Capturing and using clinical outcome data: Implications for information systems design. *J Am Med Informatics Assoc*. 1995;2:191-196.

https://doi.org/10.1017/S026646230505018X Published online by Cambridge University Press