

Comparing Properties of Audit Data and Routinely Collected Register Data in Case of Performance Assessment of Hip Fracture Treatment in Finland

R. Sund¹, I. Nurmi-Lüthje², P. Lüthje³, S. Tanninen³, A. Narinen², I. Keskimäki¹

¹National Research and Development Centre for Welfare and Health, Helsinki, Finland

²Health Centre of Kouvola Region, Kouvola, Finland

³Kuusankoski Regional Hospital, Kuusankoski, Finland

Summary

Objectives: To compare prospective audit data and secondary administrative register data in the production of performance assessment information in the case of hip fracture treatment, and to cross-validate the quality of information.

Methods: First, a conceptual model for the performance assessment of hip fracture treatment was defined. This model was then utilized in comparisons between the prospective audit data concerning 106 consecutive hip fracture patients from the Kuusankoski Regional Hospital and corresponding register data from the Finnish Health Care Register and the Causes of Death Register. We examined the completeness of registration of patients and also the accuracy and degree of completeness of the registered data. Observed differences were checked against the medical records.

Results: Register data lack clinical detail, but outperform prospective data in the recording of inpatient care history. Completeness of the register data is very good. The accuracy of easily measurable variables in the register is at least 95%. The agreement between register and audit data was 86.3% for detailed hip fracture diagnosis. Polyserial correlation between the functional dependency variables was 0.68.

Conclusions: Register and audit data have certain limitations and problems, but both seem to be suitable for the performance assessment of hip fracture treatment. To improve the feasibility of the administrative register data, the voluntary input of additional hip fracture event data to the register should be made possible. Standardized instructions that guide the use of available register classifications in a sensible way would improve the quality of data.

Keywords

Validity, completeness, accuracy, information systems, conceptual model

Methods Inf Med 2007; 46: 558–566

doi:10.1160/ME0382

Introduction

The theoretical principles of the performance assessment of health systems have been widely studied [1, 2]. Performance assessment requires a multidisciplinary approach and is closely connected to issues such as quality of health care [3, 4], systems analysis [5], knowledge management [6], and decision support systems [7]. Continuous improvement techniques are required in the design and management of information systems suitable for performance assessment [8, 9]. There exist several methodological challenges: complicated interactions between the physiological onset of disease, health system, treatment decisions, human behavior, society and expenditures need to be integrated under the same theoretical model [10, 11]. In practice, the most useful approach to performance assessment seems to be a disease-based comparison of episodes of care [12, 13]. Relevant performance criteria can be extracted from the evidence-based treatment guidelines [14]. Typically comparisons are made between the providers of care [15], though it requires the adjustment of known risk factors that may disturb the comparisons [16–18].

Practical and reliable information on evidence-based performance assessment in the case of hip fracture treatment is becoming increasingly necessary [19]. Good experiences of the Swedish Rikshöft-registry have led to the formalization of data production for the standardized audit of hip fractures in Europe (SAHFE) [20]. Data complying with the SAHFE standards have also been produced in some hospitals in Finland

[21–23]. This kind of prospective collection of clinical data obviously represents the practical “state-of-the-art” consensus for data requirements for assessing the treatment of hip fractures. Unfortunately, the separate data production requires extra work and resources, and therefore it is unlikely that the extensive voluntary data collection required by SAHFE type audit data would be feasible in all hospitals. It would be practical and cost-effective if routinely collected administrative data could be used for performance assessment purposes [24]. In fact, in Finland administrative registers have been utilized in performance assessment in the case of hip fractures [25–27].

In principle, the Finnish health registers offer a very attractive and flexible environment for research purposes, because the universal personal identification numbers are used in all registers [28]. Deterministic record linkage can be used and the potential difficulties with complex probabilistic linkages are thus avoided [29–31]. Also the accuracy of the most important variables in the Finnish registers is known to be good [32–35]. Unfortunately, the available evidence concerning the validity of register data can not be directly generalized to the case of hip fracture. One study of acute pelvic fractures in Finland during 1988 has investigated the correspondence between the register data and the medical records of 114 patients (10% of 1212 patients) [36]. The accuracy of the data of the register was found to be at least 95% for most important variables excluding secondary diagnosis (80%) and place of injury (75%). In another study investigating hip fracture incidence in

one health care district from 1982-3 and from 1992-3 in Finland, the data on hip fracture patients ($n = 668$) identified from hospital files was corrected using information from medical records [37]. The hospital discharge register missed 6.6% of patients in 1982-3 and 2.6% in 1992-3, and a high percentage of false or multiple data were found (41.8% in 1982-3 and 21.7% in 1992-3). Besides these, we do not know of any other Finnish studies that analyze the accuracy or completeness of the national hospital discharge register in regard to fracture patients.

From a general perspective, accuracy and completeness can not be defined uniquely without an assumption of a golden standard of measurement, and data validity must also be judged against the intended utilization purposes [38, 39]. In studying incidence, the definition of a golden standard is straightforward, but this is not the case in performance assessment. Fortunately, the data requirements for incidence calculation and for determining the starting points of care episodes in performance assessment are very similar [40]. This resemblance between two approaches implicitly shows that assumptions of a golden standard are very simplistic in incidence studies: certain observable facts are measured at a fixed time point. In performance assessment the main interest is on the dynamic processes of care, which essentially means dealing with different perspectives of measurement changing in time [41]. For instance, it is obvious that the detailed diagnosis of hip fracture is more important in the operating room than during the final stages of rehabilitation, and an assumption of constant accuracy is practically unrealistic. Moreover, the actual diagnosis of hip fracture is only an interpretation even after seeing x-rays, which suggests that the assumption of a golden standard may be erroneous [42]. This potential ambiguity is clearer in other diagnoses such as dementia or schizophrenia, and inevitable in the cases of more complex concepts such as post-operative complication or health status [43]. These problems become even more concrete and critical, if secondary data – in other words data collected originally for some other purpose – are to be used [24, 26, 44].

To make any statements concerning the quality of (secondary) data, it is essential to first outline the properties of required measures carefully under a fixed conceptual model (determine how data result from the theory), and then evaluate the observed data against these requirements (examine how well the theory can be reconstructed using the properties of actual observed data). In other words, to analyze the quality of available data for performance assessment a pragmatically useful compromise between problem-oriented and data-driven theories is required. In this sense, the definition for a golden standard of measurement becomes hermeneutic: the key issue is to understand why the data are like they are without fixing the reality using any single data source. However, it can be assumed that primary data collected for hip fracture audit purposes correspond more closely to any reasonable golden standard than secondary register data that were originally collected for other purposes.

Objectives

The aims of this study were to define a conceptual model for producing data for performance assessment in the case of hip fracture treatment, to compare two different sources of data – prospective clinical audit data (designed for monitoring treatment quality) and secondary administrative register data (used mainly in compiling statistics) – in the production of performance assessment information, and also to cross-validate the quality of these data sources

Methods

For the definition of the conceptual model, the list of SAHFE-variables was chosen to represent an adequate base for data requirements (<http://www.sahfe.ort.lu.se/guide.html>). These are actual operationalizations of those patient level concepts considered important in the case of hip fracture. In this study, the dimensions of the conceptual model were abstracted from the actual oper-

ationalized measures, and then complemented using theoretical models of performance assessment [1, 18, 45, 46]. Each phenomenon represented by a concept was assumed to be attributable to an individual on a continuous time scale. It was also assumed that each concept can be described using a systems approach with a (limited or unlimited) number of states so that the system is always in some of these states at each time point. The concepts were then classified into more general groups in terms of their contextual interpretation, temporal stability and theoretical measurement properties. This qualitative classification procedure was repeated until the resulting conceptual model was considered satisfactory in the sense of Occam's razor principle.

Audit data were collected prospectively for 106 consecutive hip fracture patients in the Kuusankoski Regional Hospital between January 1, 1999 and January 31, 2000 [22, 23]. Patients treated in the surgical ward at the Kuusankoski Regional Hospital during the same period were identified from the Finnish Health Care Register, and all records of these patients from 1987-2002 were extracted from the Finnish Hospital Discharge Register, the Finnish Health Care Register, and the Causes of Death Register using the unique personal identification numbers of the patient population. Detected differences between data sources were further checked manually from the medical records of these patients (PL). The ethical committee of the Kymenlaakso Hospital District approved the study. Permission to use the data was obtained from the Kuusankoski Regional Hospital, the National Research and Development Centre for Welfare and Health, and also from Statistics Finland.

The selected measures of different data sources were first matched to the defined conceptual model. The completeness of registration of hip fracture patients as well as the accuracy and degree of completeness of the registered data were examined in both the audit and register data using standard methodology [38]. For some variables it was more reasonable to measure agreement than accuracy. This required approaches based on modeling the decision-making process, such as a polychoric correlation model (a form of latent trait model) [47]. The propor-

tion of agreements that were not due to chance was measured using a delta coefficient, which is based on the model of multiple-choice tests and avoids certain known deficiencies of the kappa coefficient [48]. Kappa may be considered as an approximation of delta, as both yield values that are very similar when at least one marginal distribution is balanced or when both marginal distributions are moderately unbalanced in the same direction [49].

Results

The dimensions of the conceptual model and selected measures attributable to these dimensions in various data sources are reported in Table 1. Of the dimensions, the biological constants do not change in time and therefore one measurement is generalizable to all times. Biological events, accident/

fall history, and hip fracture event represent dimensions for which actual values of measures are recorded in the proximity of some observable event. All other dimensions relate to phenomena that potentially change in time and should be continuously monitored. In practice, time must be fixed for actual measurement. In the audit data, the measurement is done in connection with a hip fracture event and for certain measures also two weeks, four months, and one year after the fracture. In the register data, the recording takes place at each discharge. As expected, the content of audit data is richer than in the register data, even after using indirect measures of certain concepts. However, the measurement frequency in register data is superior to audit data, allowing complete observations of event histories for inpatient care.

Completeness of Registration

There were 104 patients in the prospective data (two had a new fracture at the other side of the hip during the follow-up). In the register there were 111 patients who were admitted to the surgical inpatient ward at the Kuusankoski Regional Hospital during the

study period with a hip fracture diagnosis. It turned out that there were in total 105 patients with a confirmed hip fracture diagnosis (Table 2). The audit data missed one case and the register data two cases indicating very good completeness. The patient miss-

ing from the audit data had an impacted hip fracture (femoral neck) and was erroneously excluded from the prospective study. Both cases missing from the register data had a diagnosis of distal femur fracture recorded in the register. Three extra hip fracture patients with clearly false hip fracture diagnoses were found in the register. For the first patient, there was a coding error in the medical record diagnosis (S42.0 → S72.0) and the second patient had a distal femur fracture, but for the third patient any explanation for an obvious error was not found. One patient with hip fracture diagnosis in the register had a periprosthetic hip fracture and therefore this patient was not a hip fracture patient. Another patient was admitted because of suspected hip fracture, but no fracture was found. Three patients excluded from the audit data were having on-going treatment for a recent earlier hip fracture – one operated on at the Kuusankoski Regional Hospital and two at some other hospital. It also turned out that four patients included in the audit data had had – according to the register – an earlier hip fracture during the preceding ten years. The side of the fracture is recorded very seldom in the register, but the medical records revealed that only one of these patients had the new fracture on the same side as the earlier fracture. The positive agreement between register and audit data was very good (94.9%), and would be even higher if appropriate data abstraction rules for identification of false positives are defined. If data abstraction rules are not used, the register data will overestimate the number of new hip fractures.

Completeness of Registration for Re-operations and New Fractures

Two of the 105 hip fracture patients needed an acute re-operation during the first admission. Neither of these re-operations was detectable from the register data. In fact, the

data structure of the Finnish Health Care Register allows only one operation day to be recorded during one period, and recording of two similar operation codes for the same period may seem to be an error. In both cases, the data in the register corresponded to the latter operation.

Three patients with a new hip fracture during the first year of follow-up were detected from the register. Two were also found from the audit data and the third was confirmed using the information in the medical record. The (confirmed) register data also revealed that one of these three patients had a third operation because of problems with the hip prosthesis during the first year of follow-up.

Accuracy and Completeness of Easily Measurable Variables

The completeness and accuracy of the register data were examined using 106 hip fracture events with complete audit data. Completeness indicates the percentage of cases with a recorded value for the variable in question. Accuracy tells the proportion of correctly recorded values for cases having at least some recorded values. The correctness – telling the overall utility (measured as the product of completeness and accuracy) of the variable – was very good for most of the variables (Table 3). The accuracy of admission source and discharge destination was improved by using record linkage instead of variables in the index admission period. Discharge day also required record linkage, because each transfer from one ward to another even in the same hospital results in a new discharge in the register. Missing operation codes are a nuisance in the case of hip fractures, because they may also indicate conservative treatment. The data structure of the register does not suit well to recording acute re-operations during the same hospital period, resulting in the recording of wrong codes and days for primary operation. The extra operation code indicating the side of fracture was used very rarely. Accuracy for the place of accident was poor. This was mainly because of incapability of the ICD-10 classification to

Table 1 Conceptual model dimensions and their realizations in different data sources

	SAHFE measure	Audit data measure	Register data measure
Biological facts			
Biological events	date of birth, death, year of menopause, menarche	date of birth and death	date of birth and death
Biological constants (dna)	sex	sex	sex
Biological measures			
Physical examinations	height, weight		
Laboratory tests	hemoglobin, creatine, albumin, bone density		
Demographic history			
Household composition	living alone	marital status, living alone	
Place of living	residential status	residential status, area of living	area of living, <i>level of care</i> ¹
Socioeconomic history			
Education, occupation, economic resources			
Health related behavior			
Risk behavior	smoking, alcohol intake	alcohol usage	<i>alcohol related inpatient care</i> ¹
Diet, activities			
Subjective quality of life			
Overall well-being	pain, psychological state, fear of fall	Pain	
Objective need for care			
Diseases/symptoms	comorbidity, complications	Complications	<i>recorded diagnoses, use of care</i> ¹
Physical functioning	walking	walking, activities of daily living	need for care
Cognitive functioning	Abbreviated mental test score		
Use of care			
Technical aid	walking aids	use of assistive technology, changes in living environment	
Medication		use of painkillers	
Health care utilization	provider, type of stay, length of stay	type of care, rehabilitation, dates for hospital visits	provider, type of care, admission and discharge dates, diagnoses, operations
Accident/fall history			
Properties of fall	date of fall, place of fall	date of fall, place of fall	<i>place of and reason for hospitalized injury</i> ¹
Hip fracture event			
Properties of fracture	time of fracture, occurrence place, fracture type, fracture side	day of fracture, reason for fracture, occurrence place, fracture type, fracture side	place of and reason for hospitalized injury
Initial treatment process	hospital id, admission time, admission source, start of operation, reason for operative delay, type of surgeon and anesthetist, type of anesthetic, operative method, length of surgery, type of prophylaxis, complications in operation, reoperations, time to mobilization, discharge time, discharge destination	admission day, admission source, day of operation, reason for operative delay, operative method, complications in operation, reoperations, discharge day, discharge destination	hospital id, admission day, admission source, diagnoses, day of operation, operation codes, discharge day, discharge destination
Clinical stability	ASA grade, comorbidities, coexisting fractures, reason for operative delay in medically unfit patients	ASA grade, reason for operative delay in medically unfit patients	<i>recorded diagnoses, use of care</i> ¹

1) Indirect measurement requiring record linkage

In both data sets: 98 new hip fracture patients 4 patients with earlier fracture	Missing from register: 2 false-negative register diagnoses
Missing from audit data: 1 verified new patient 2 operated in other hospital 1 recent earlier fracture 1 earlier total prosthesis 1 suspected hip fracture 3 false-positive register diagnoses	Positive agreement: 94.9% (95% CI: 90.3-96.8%) Overestimation in the register: 7.2% (95%-CI: 3.2-13.7%) Underestimation in the register: 1.9% (95%-CI: 0.2-6.7%)

Variable	Completeness ¹	Accuracy ²	Correctness ³
Personal identification number ⁴	100%	100%	100%
Home municipality	100%	100%	100%
Admission day	100%	100%	100%
Admitted from	100%	83.0%	83.0%
Admitted from ⁵	100%	95.3%	95.3%
Discharge day	100%	100% ⁵	100%
Discharge to	100%	96.2% ⁵	96.2%
Discharge to ⁵	100%	100%	100%
Operation day	99.1%	97.1%	96.2%
Primary operation	99.1%	98.1%	97.2%
Side of fracture	2.8%	100%	2.8%
Place of accident ⁶	93.4%	61.3%	57.3%
Main diagnosis (hip fracture)	100%	98.1%	98.1%

1) Percentage of cases with recorded value for the variable
2) Proportion of correctly recorded values for cases with some recorded value
3) The overall utility of the variable (product of completeness and accuracy)
4) Contains date of birth and sex
5) Identification by utilizing record linkage
6) Incompatible classification in the register

Table 4 Accuracy of hip fracture diagnosis in the register

	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Femoral neck fracture	96.7% (91.5-100%)	80.5% (68.2-91.7%)	88.1% (79.7-95.3%)	94.3% (85.7-100%)
Trochanteric fracture	68.6% (52.9-83.7%)	98.5% (95.1-100%)	96.0% (87.0-100%)	85.7% (77.5-93.1%)
Subtrochanteric fracture	83.3% (42.9-100%)	96.9% (92.8-100%)	62.5% (25.0-100%)	98.9% (96.7-100%)

separate home accidents from the accidents occurring in residential care and falls occurring outdoors from the ones occurring indoors.

Accuracy and Reliability of Hip Fracture Diagnoses

Register diagnosis was recorded as hip fracture (ICD-10: S72.0, S72.1, S72.2) in

Table 2
Completeness of data sets

98.1% (95% CI: 93.4-99.8%) of cases (Table 3). However, the evaluation of the exact accuracy of the main diagnosis was not straightforward, as different classifications of hip fractures were used in the register and in the audit data. Following the methodological suggestions, the accuracy of diagnoses was addressed by considering the bone anatomy as a biological entity identifying true fracture status [50].

Hip fractures are commonly classified into intra- and extra-capsular according to their relationship to the capsular attachment of the hip. In terms of current data, fractures of the neck of the femur are intra-capsular, and other hip fractures are extra-capsular.

The bone anatomy may be used to subdivide extra-capsular fractures further into basocervical, trochanteric and subtrochanteric fractures. Unfortunately, there is no separate diagnosis code in ICD-10 for basocervical fracture. Using a detailed fracture classification of femoral neck fracture, trochanteric fracture or subtrochanteric fracture, the agreement between audit data and register was 86.3% (95% CI: 79.4-92.2%). Pairwise agreement analyses revealed that femoral neck and subtrochanteric fractures were correctly identified in the register (high sensitivity), but identification of trochanteric fractures was not as accurate (Table 4). This is most probably because the extremities are easier to identify correctly. This interpretation is supported by the fact that the register diagnoses of trochanteric fracture were practically always true trochanteric fractures (high positive predictive value). This was not the case for femoral neck fractures or subtrochanteric register-diagnoses which were occasionally used for true trochanteric fractures (low positive predictive value). Misclassification at the borders of the trochanteric region of the femur may be one reason. Another cause is the ICD-10 classification utilized in Finland, which suggests the use of femoral neck diagnosis for unspecified hip fractures. The proportion of agreements that were not due to chance was very good as measured with the delta coefficient (81.3%; 95% CI: 71.3-91.3%) or the kappa coefficient (73.5%; 95% CI: 61.2-85.7%). In addition, since it was reasonable to assume that the underlying trait is continuous, the poly-

Table 3
Completeness and accuracy of easily measurable variables in the register

choric correlation coefficient was calculated indicating a very high consistency between the audit data and register (0.92; 95% CI: 0.86-0.98).

Agreement between Functional Dependency Measures

In the audit data, functional dependency was measured using an activities of daily living (ADL) type of scale [23]. In the Finnish Health Care Register, overall dependency (and need for care) is recorded using a scale of five categories [51]. The completeness of dependency variable in the register was 100%. The boxplots describing dependency measures (at discharge) are shown in Figure 1. The upper categories of the register-based measure are combined due to the small number of observations. The polyserial correlation between the variables is 0.68 (95% CI: 0.55-0.81), showing rather good consistency. Means and medians behave reasonably, but Figure 1 indicates that the most commonly used category in the register data has a very wide range. It seems that high register categorizations truly reflect bad functionality, but absolute interpretation of the register scale at an individual level may be erroneous.

Follow-up Information

State diagrams for the register and audit data summarizing the follow-up information are shown in Figure 2. The overall shapes are pretty similar, but the register data have more accuracy. The dates of deaths, lengths of initial hospitalizations, and residential statuses at four months and at one year after the fracture in the audit data are accurate, but the follow-up information concerning acute care subsequent to the initial hospitalization in the audit data is partly incomplete and therefore technically difficult to handle. This can be most clearly seen from the proportion of patients in nursing homes, which is first overestimated and later on underestimated (Fig. 2).

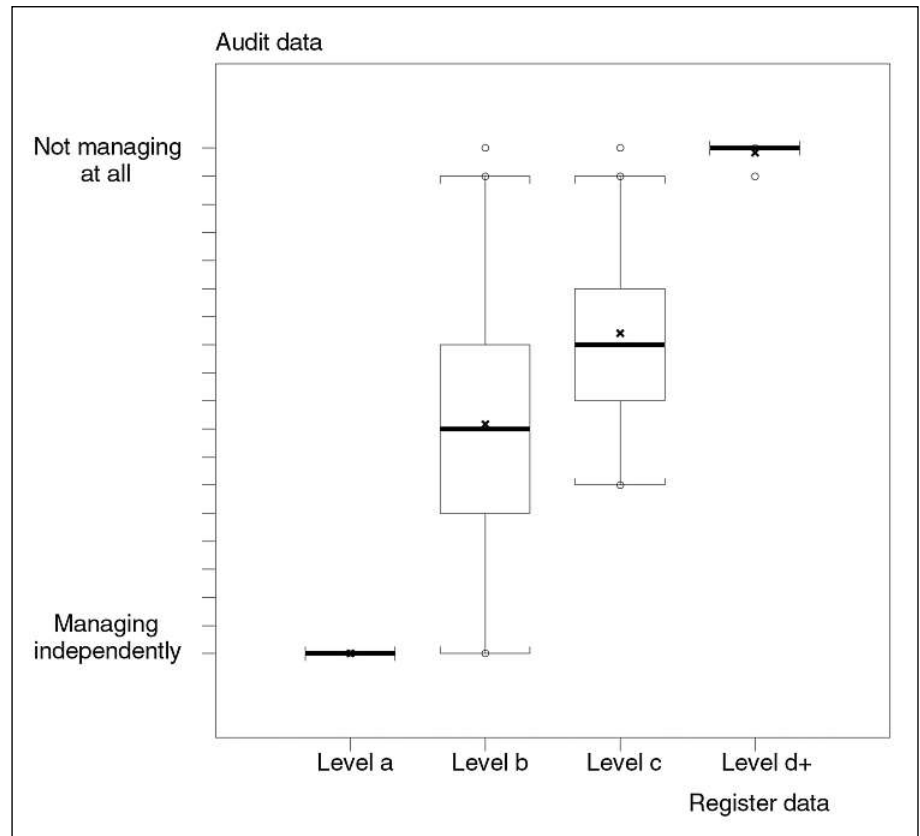


Fig. 1 Functional dependency measures in register and audit data. Audit data has an activities of daily living (ADL) type of sum scale [23]. Register data levels are: a) completely or almost independent, b) occasional need for care, c) recurrent need for care, d+) almost (or totally) continuous need for care [51]. Cross refers to mean in boxplot.

Discussion

In this study two forms of data production for performance assessment purposes were compared in the case of hip fracture treatment. A justifiable comparison required a definition of the conceptual model, which allowed for a systematic structuring of links between observable measures and theoretical concepts. Such models are only rarely reported [46], even though the explicit introduction of a conceptual model obviously significantly improves the understanding of the problem and makes the most important assumptions affecting the actual results of analyses visible.

Systematic approaches can be used for the development of a conceptual model [52]. Quite often conceptual models had to be represented so that it becomes possible to consider several different perspectives

simultaneously [53-56]. This is particularly challenging if data are used for other purposes than originally intended [57]. In this study, a pragmatic mapping between concepts from a performance assessment theory, SAHFE metadata, and structured data entry in prospective audit data and secondary register data was constructed. In spite of the generality of the current model, it is limited in the sense that it deals with individual level concepts only. For benchmarking purposes, provider level phenomena that give explanations of observed differences between providers of care should also be considered. One very important individual level measure – namely costs – is also missing from the model. It would have special measurement properties, because it is cumulative in time.

As expected, the register data lacked clinical detail. However, most of the more detailed measures in the audit data were ac-

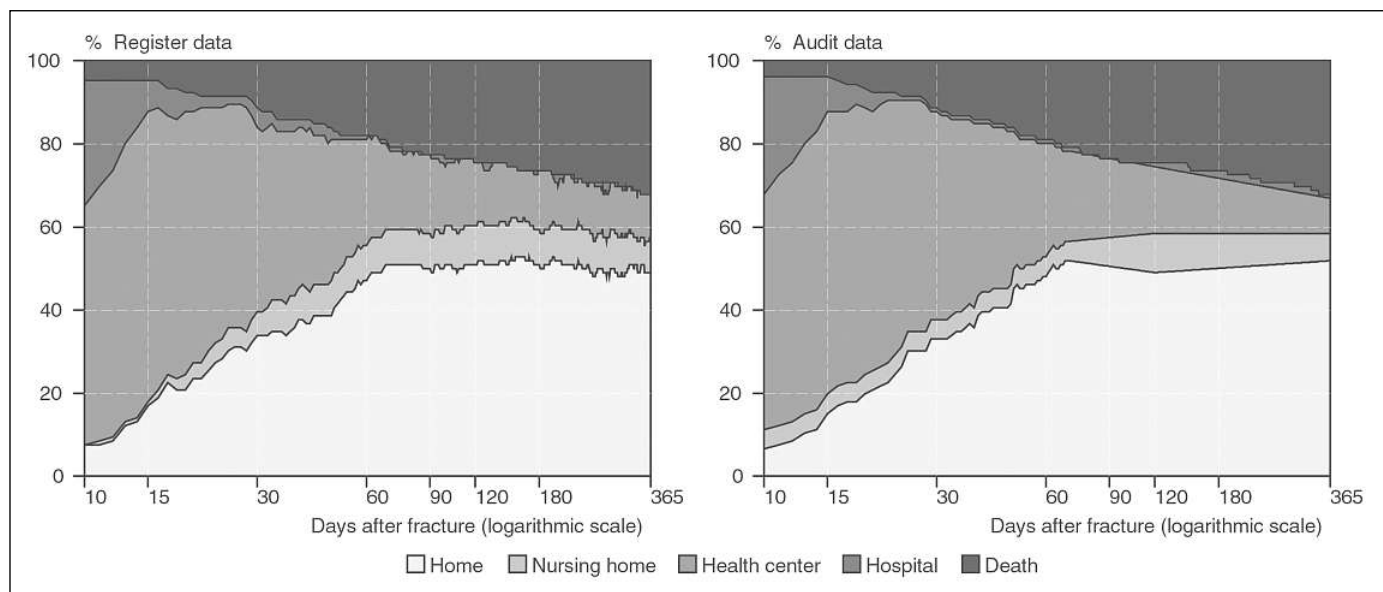


Fig. 2 State diagrams for register and audit data

tually related to phenomena that continuously change in time. Such data require tacit knowledge to be useful in a more general context [58]. On the other hand, register data have a data structure that allows the complete observation of inpatient care history, and therefore outperforms prospective data in this sense if some delay in data production is tolerable. Making use of demographic, socioeconomic and medication histories data – which are, in principle, available in other Finnish registers – could also improve certain details of register data, but it requires extensive work to obtain and preprocess such – possibly expensive – data.

In this study, it was assumed that all hip fracture patients were identifiable from the register or audit data. This assumption is valid so far as patients were treated at the orthopedic inpatient ward at the Kuusankoski Regional Hospital, because it is very unlikely that two complete, but mutually different registrations would miss a patient. In fact, even the patients treated conservatively or operated on at another hospital are routinely referred to an orthopedic inpatient ward and would be registered. In theory, some patients admitted to the outpatient emergency department who were not resident in the operation area of the Kuusankoski Regional Hospital and thus directly

transferred to their own local hospital, could have been missed from the data. However, in our study capture-recapture analyses estimating the total number of cases were not applicable, because there was structural dependency between data sources (the occurrence of a patient in audit data also indicated expected occurrence in register data) [59, 60]. Therefore the completeness was evaluated by identifying all hip fracture cases from audit data and register data, and then all incompatibilities were checked against the information in medical records. In this sense, the differences in completeness between data sources indicate the different definitions of hip fracture or errors in register data.

The completeness of audit data was excellent, and also the completeness of register data seemed to be very good. In fact, there is a tendency to overestimation if no appropriate data abstraction rules are used. For example, it is more reasonable to count new hip fracture cases than every patient with an ongoing hip-fracture-related care episode [61, 62]. However, unless the clinical judgment of a hip fracture case is accurately and completely recorded, register-based estimates for the numbers of hip fractures are prone to bias. The appropriateness of such data abstraction rules can be partially verified by comparing the hospital-

specific numbers of consecutive hip fractures during particular time periods – available from prospective studies – to register-data-based estimates for the same hospitals and the same time periods. It must be noted that even this kind of a definition has some drawbacks in incidence calculations. The number of patients is hospital-specific, but the available risk-population data refers to geographical areas and one can therefore argue that patients living in the same area but operated in some other hospital should also be counted as cases, and patients from other areas should be excluded.

Our definition for a golden standard of measurement in this study was hermeneutic. In practice, three types of comparisons between data sources were utilized (even though the boundaries between types are not clear cut). The first type was applied in accuracy comparisons: the same rather easily measurable variable was available in both data sources, and detected differences could be further checked from the medical records (e.g. Table 3). The second type was for the comparison of theoretically similar information that has substantially different operationalizations in both data sources (e.g. Figs. 1 and 2). The third type of comparison was a combination of the first two types: both approaches can be applied (e.g. analyses of the detailed fracture classification).

The key idea was to make versatile comparisons that are methodologically justified and practically relevant.

The accuracy of the most easily measurable variables in the register is very good. However, the data structure for the recording of performed operations is not optimal in the register. For example, only one operation day is allowed for one hospitalization even though the patient may be re-operated on during the same hospitalization. The content of these fields could also be improved by giving standardized instructions that guide practitioners to also record the relevant additional operation codes available in the classification, such as which side the fracture is on. Such additional metadata should also be feasible to use to improve the quality of register data in general in cases where the use of structured data entries is known to vary between coders [39, 63, 64].

For the purposes of performance assessment, follow-up information is particularly essential. Mortality and length of initial hospitalization period were recorded equally in both data sets, but otherwise it was not easy to derive information that was exactly compatible. Information on the use of inpatient health services on a daily basis was available directly from the register while deriving similar information from the audit data required certain assumptions and interpolation. The collection of follow-up information from the audit data requires extra work and is prone to different biases as is the routine collection of register data. Even though the use of health services is recorded better in the register, other measurements of useful concepts such as use of technical aid, functional disability and pain offer information that is not directly available from the register. However, because the register includes a disability measure, it actually offers complementary information (at admission, discharge and census points). This information would be more valuable if some commonly used functional disability or quality of life measure had been available in the register instead of a non-validated ad hoc measure [65].

On the other hand, the register includes information concerning the provider-specific use of health services with diagnostic-

related group information for the total patient population. This allows the calculation of costs by summing up the suitably (price)-weighted information on the use of health services. This same idea for calculating costs can be further utilized, though the weights this time correspond not to prices but to the mean health or disability status of persons receiving specific forms of care, and this results in health status measure which is unbiased (in the same sense as costs can be unbiased) at the population level. Examination of such ideas is out of the scope of this study, but may offer fruitful directions for further research.

As a conclusion, register and audit data both have certain limitations and problems, but seem to be suitable for the performance assessment of hip fracture treatment. Voluntary input of additional hip fracture event data to the register should be made possible, because the best compromise for practical performance assessment purposes would be reached by a prospective recording of hip fracture event data combined with extraction of care histories of these patients from the registers.

Acknowledgments

This work was financially supported by a research grant from the Academy of Finland. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report.

References

- Aday LA, Begley CE, Lairson DR, Slater CH. Evaluating the healthcare system: Effectiveness, efficiency, and equity. 2nd edition. Chicago: Health Administration Press; 1998.
- OECD. Measuring up – Improving health system performance in OECD countries. Paris: Organisation for Economic Co-operation and Development; 2002.
- Hasman A, Safran C, Takeda H. Quality of health care: informatics foundations. *Methods Inf Med* 2003; 42 (5): 509-18.
- Ammenwerth E, Haux R, Kulikowski C, Bohne A, Brandner R, Brigl B, et al. Medical informatics and the quality of health: new approaches to support patient care – findings from the IMIA Yearbook of Medical Informatics 2003. *Methods Inf Med* 2003; 42 (2): 185-9.
- Ammenwerth E, Ehlers F, Eichstadter R, Haux R, Pohl U, Resch F. Systems analysis in health care: framework and example. *Methods Inf Med* 2002; 41 (2): 134-40.
- Stefanelli M. Knowledge management to support performance-based medicine. *Methods Inf Med* 2002; 41 (1): 36-43.
- Fieschi M, Dufour JC, Staccini P, Gouvernet J, Bouhaddou O. Medical decision support systems: old dilemmas and new paradigms? *Methods Inf Med* 2003; 42 (3): 190-8.
- Leonard KJ. Total quality in information systems management: issues for the health care industry. *Methods Inf Med* 1998; 37 (2): 156-60.
- Nykänen P, Karimaa E. Success and failure factors in the regional health information system design process – results from a constructive evaluation study. *Methods Inf Med* 2006; 45 (1): 85-9.
- Stefanelli M. Knowledge and process management in health care organizations. *Methods Inf Med* 2004; 43 (5): 525-35.
- Garde S, Wolff AC, Kutscha U, Wetter T, Knaup P. CSI-ISC – Concepts for smooth integration of health care information system components into established processes of patient care. *Methods Inf Med* 2006; 45 (1): 10-8.
- OECD. A disease-based comparison of health systems. What is best and at what cost? Paris: Organisation for Economic Co-operation and Development; 2003.
- Grady ML, Schwartz HA, editors. Medical effectiveness research data methods. Rockville: Agency for Health Care Policy and Research; 1992.
- Kumar A, Smith B, Pisanelli DM, Gangemi A, Stefanelli M. Clinical guidelines as plans – an ontological theory. *Methods Inf Med* 2006; 45 (2): 204-10.
- Ghali WA, Hall RE, Ash AS, Rosen AK, Moskowitz MA. Evaluation of complication rates after coronary artery bypass surgery using administrative data. *Methods Inf Med* 1998; 37 (2): 192-200.
- D'Hoore W, Sciotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson comorbidity index. *Methods Inf Med* 1993; 32 (5): 382-7.
- O'Connell RL, Lim LL. Utility of the Charlson comorbidity index computed from routinely collected hospital discharge diagnosis codes. *Methods Inf Med* 2000; 39 (1): 7-11.
- Iezzoni LI. Risk Adjustment for measuring health care outcomes. Third edition. Chigago: Health Administration Press; 2003.
- Treml J, Kroker PB. Hip fracture audit: time for standards to be evidence-based. *Int Orthop* 2000; 24 (4): 181-3.
- Parker MJ, Currie CT, Mountain JA, Thorngren KG. Standardised audit of hip fracture in Europe (SAHFE). *Hip Int* 1998; 8 (1): 10-5.
- Heikkinen T, Willig R, Hänninen A, Koskinen K, Mannismäki P, Alavaikko A, et al. Hip fractures in Finland – a comparison of patient characteristics and outcomes in six hospitals. *Scand J Surg* 2004; 93 (3): 234-40.
- Nurmi I, Narinen A, Luthje P, Tanninen S. Cost analysis of hip fracture treatment among the elderly for the public health services: a 1-year prospective study in 106 consecutive patients. *Arch Orthop Trauma Surg* 2003; 123 (10): 551-4.

23. Nurmi I, Narinen A, Luthje P, Tanninen S. Functional outcome and survival after hip fracture in elderly: a prospective study of 106 consecutive patients. *J Orthopaed Traumatol* 2004; 5 (1): 7-14.
24. Powell AE, Davies HT, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003; 12 (2): 122-8.
25. Rissanen P, Sund R, Nordback I, Rousi T, Idänpään-Heikkilä U. Register based effectiveness measurement in the case of hip fracture treatment (in Finnish). Helsinki: National Research and Development Centre for Welfare and Health; 2002.
26. Sund R. Utilisation of administrative registers using scientific knowledge discovery. *Intelligent Data Analysis* 2003; 7 (6): 501-19.
27. Sund R, Liski A. Quality effects of operative delay on mortality in hip fracture treatment. *Qual Saf Health Care* 2005; 14 (5): 371-7.
28. Gissler M, Haukka J. Finnish health and social welfare registers in epidemiological research. *Norsk Epidemiologi* 2004; 14 (1): 113-20.
29. Roos LL, Wajda A. Record linkage strategies. Part I: Estimating information and evaluating approaches. *Methods Inf Med* 1991; 30 (2): 117-23.
30. Wajda A, Roos LL, Layefsky M, Singleton JA. Record linkage strategies: Part II. Portable software and deterministic matching. *Methods Inf Med* 1991; 30 (3): 210-4.
31. Oberaigner W, Stuhlinger W. Record linkage in the Cancer Registry of Tyrol, Austria. *Methods Inf Med* 2005; 44 (5): 626-30.
32. Keskimäki I, Aro S. Accuracy of data on diagnoses, procedures and accidents in the Finnish Hospital Discharge Register. *Int J Health Sci* 1991; 2 (1): 15-21.
33. Lahti RA, Penttilä A. The validity of death certificates: routine validation of death certification and its effects on mortality statistics. *Forensic Sci Int* 2001; 115 (1-2): 15-32.
34. Leppälä JM, Virtamo J, Heinonen OP. Validation of stroke diagnosis in the National Hospital Discharge Register and the Register of Causes of Death in Finland. *Eur J Epidemiol* 1999; 15 (2): 155-60.
35. Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Räihä P, Kärjä-Koskenkari P, et al. The validity of the Finnish Hospital Discharge Register and Causes of Death Register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil* 2005; 12 (2): 132-7.
36. Luthje P, Nurmi I, Kataja M, Heliövaara M, Santavirta S. Incidence of pelvic fractures in Finland in 1988. *Acta Orthop Scand* 1995; 66 (3): 245-8.
37. Huusko TM, Karppi P, Avikainen V, Kautiainen H, Sulkava R. The changing picture of hip fractures: dramatic change in age distribution and no change in age-adjusted incidence within 10 years in Central Finland. *Bone* 1999; 24 (3): 257-9.
38. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol* 1996; 25 (2): 435-42.
39. Winthereik BR. "We fill in our working understanding": on codes, classifications and the production of accurate data. *Methods Inf Med* 2003; 42 (4): 489-96.
40. Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement and use. *Med Care Rev* 1985; 42 (2): 163-218.
41. Chin MH, Muramatsu N. What is the quality of quality of medical care measures? Rashomon-like relativism and real-world applications. *Perspect Biol Med* 2003; 46 (1): 5-20; discussion 21-3.
42. Audige L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand* 2004; 75 (2): 184-94.
43. Zhang J. Representations of health concepts: a cognitive perspective. *J Biomed Inform* 2002; 35 (1): 17-24.
44. Wray NP, Ashton CM, Kuykendall DH, Hollingsworth JC. Using administrative databases to evaluate the quality of medical care: a conceptual framework. *Soc Sci Med* 1995; 40 (12): 1707-15.
45. Kane RL, editor. Understanding health care outcomes research. Gaithersburg: Aspen Publishers; 1997.
46. Shroyer AL, London MJ, Sethi GK, Marshall G, Grover FL, Hammermeister KE. Relationships between patient-related risk factors, processes, structures, and outcomes of cardiac surgical care. Conceptual models. *Med Care* 1995; 33 (10 Suppl): OS26-34.
47. Hripcsak G, Heitjan DF. Measuring agreement in medical informatics reliability studies. *J Biomed Inform* 2002; 35 (2): 99-110.
48. Andres AM, Marzo PF. Delta: A new measure of agreement between two raters. *Br J Math Stat Psychol* 2004; 57 (1): 1-19.
49. Andres AM, Marzo PF. Chance-corrected measures of reliability and validity in $K \times K$ tables. *Stat Methods Med Res* 2005; 14 (5): 473-92.
50. Audige L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma* 2005; 19 (6): 401-6.
51. Björkgren MA, Fries BE, Häkkinen U, Brommels M. Case-mix adjustment and efficiency measurement. *Scand J Public Health* 2004; 32 (6): 464-71.
52. Schnabel M. Representing and processing medical knowledge using formal concept analysis. *Methods Inf Med* 2002; 41 (2): 160-7.
53. Degoulet P, Fieschi M, Chatellier G. Decision support systems from the standpoint of knowledge representation. *Methods Inf Med* 1995; 34 (1-2): 202-8.
54. Brandt CA, Sun K, Charpentier P, Nadkarni PM. Integration of Web-based and PC-based clinical research databases. *Methods Inf Med* 2004; 43 (3): 287-95.
55. Brandt CA, Cohen DB, Shifman MA, Miller PL, Nadkarni PM, Frawley SJ. Approaches and informatics tools to assist in the integration of similar clinical research questionnaires. *Methods Inf Med* 2004; 43 (2): 156-62.
56. van Ginneken AM. Considerations for the representation of meta-data for the support of structured data entry. *Methods Inf Med* 2003; 42 (3): 226-35.
57. Haux R. Knowledge-based decision support for diagnosis and therapy: on the multiple usability of patient data. *Methods Inf Med* 1989; 28 (2): 69-77.
58. Berg M, Goorman E. The contextual nature of medical information. *Int J Med Inform* 1999; 56 (1-3): 51-60.
59. IWGDMF. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: History and theoretical development. *Am J Epidemiol* 1995; 142 (10): 1047-58.
60. Brenner H. Application of capture-recapture methods for disease monitoring: potential effects of imperfect record linkage. *Methods Inf Med* 1994; 33 (5): 502-6.
61. Langley J, Stephenson S, Cryer C, Borman B. Traps for the unwary in estimating person based injury incidence using hospital discharge data. *Inj Prev* 2002; 8 (4): 332-7.
62. Alsop JC, Langley JD. Determining first admissions in a hospital discharge file via record linkage. *Methods Inf Med* 1998; 37 (1): 32-7.
63. Los RK, Roukema J, van Ginneken AM, de Wilde M, van der Lei J. Are structured data structured identically? Investigating the uniformity of pediatric patient data recorded using OpenSDE. *Methods Inf Med* 2005; 44 (5): 631-8.
64. Williams JG. Measuring the completeness and currency of codified clinical information. *Methods Inf Med* 2003; 42 (4): 482-8.
65. Maly M, Vondra V. Generic versus disease-specific instruments in quality-of-life assessment of chronic obstructive pulmonary disease. *Methods Inf Med* 2006; 45 (2): 211-5.

Correspondence to:

Reijo Sund
National Research and Development Centre
for Welfare and Health (STAKES)
Outcomes and equity research
P.O. Box 220
00531 Helsinki
Finland
E-mail: reijo.sund@stakes.fi