



ORIGINAL RESEARCH

Health reform: Is routinely collected electronic information fit for purpose?

Siaw-Teng Liaw,^{1,2,3} Huei-Yang Chen,² Della Maneze,² Jane Taggart,² Sarah Dennis,^{1,2} Sanjyot Vagholkar^{1,3} and Jeremy Bunker^{1,3}

¹School of Public Health and Community Medicine and ²Centre for Primary Health Care and Equity, University of New South Wales, and ³General Practice Unit, South West Sydney Local Health District, Sydney, New South Wales, Australia

Abstract

- Objective:** Little has been reported about the completeness and accuracy of data in existing Australian clinical information systems. We examined the accuracy of the diagnoses of some chronic diseases in an ED information system (EDIS), a module of the NSW Health electronic medical record (EMR), and the consistency of the reports generated by the EMR.
- Methods:** A list of ED attendees and those admitted was generated from the EDIS, using specific (e.g. angina) and possible clinical terms (e.g. chest pain) for the selected chronic diseases. This EDIS list was validated with an audit of discharge summaries, and compared with a list generated, using similar specific and possible Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT), from the underlying EMR database.
- Results:** Of the 33 115 ED attendees, 2559 had diabetes mellitus (DM), cardiovascular disease or asthma/chronic obstructive pulmonary disease; of these 2559, 876 were admitted. Discharge summaries were missing for 12–15% of patients. Only three-quarters or fewer of the diagnoses were confirmed by the discharge summary audit, best for DM and worst for cardiovascular disease. Proportion of agreement between the lists generated from the EDIS and EMR was best for DM and worst for asthma/chronic obstructive pulmonary disease. Possible reasons for this discrepancy are technical, such as use of different extraction terms or system inconsistency; or clinical, such as data entry, decision-making, professional behaviour and organizational performance.
- Conclusions:** Variations in information quality and consistency of the EDIS/EMR raise concerns about the ‘fitness for purpose’ of the information for care and planning, information sharing, research and quality assurance.
- Key words:** *electronic medical record, emergency department information system, fitness for purpose, information quality, routinely collected data, system consistency.*

Correspondence: Professor Siaw-Teng Liaw, General Practice Unit, Fairfield Hospital, PO Box 5, Fairfield, NSW 1860, Australia. Email: siaw@unsw.edu.au

Siaw-Teng Liaw, MB BS, PhD, FRACGP, FACHI, Professor of General Practice; Huei-Yang Chen, MSc, PhD, Research Fellow; Della Maneze, MB BS, Research Associate; Jane Taggart, MPH, Research Fellow; Sarah Dennis, MSc, PhD, Senior Research Fellow; Sanjyot Vagholkar, MB BS, MPH, FRACGP, Staff Specialist; Jeremy Bunker, MB BS, MMed, FRACGP Staff Specialist.

Introduction

The National Health and Hospitals Reform Commission,¹ National Primary Care Strategy Taskforce,² National Preventive Care Strategy Taskforce³ and various jurisdictional reviews, such as the Garling Special Inquiry into Acute Care in NSW,⁴ have emphasized the use of eHealth tools to collect, share and use information to optimize health reform, policy development and strategic work plans to implement safe, effective and coordinated care over the life cycle and across the 'patient journey' in the health system. However, many studies regularly report a range of deficiencies in routinely collected electronic information for clinical purposes.⁵⁻⁷ The 'fitness for purpose' of administrative data is better.^{8,9} This functional approach defines information quality as the degree to which information consistently meets the requirements and expectations of knowledge workers and end-users who use the information products and services required to accomplish the enterprise mission or personal objectives.¹⁰ Good-quality information is essential for health information exchange and interoperability because it can reduce costs and improve safety and quality of care.

The present paper reports on a study that examined: (i) whether the principal diagnoses generated by the ED information system (EDIS) were sufficiently accurate and complete to allow reliable identification of patients with type 2 diabetes, cardiovascular disease (CVD), asthma and chronic obstructive pulmonary disease (COPD); and (ii) whether the EMR system is sufficiently consistent in generating reports regardless of which window to the EMR database is used.

The present study was approved by the University of New South Wales and Sydney South West Area Health Service (SSWAHS) Ethics Committees.

Methods

The present study focused on a community hospital ED in south-western Sydney that sees 30 000+ patients a year. It uses the EDIS, a module of the NSW Health electronic medical record (EMR). The EMR includes other modules for scheduling, clinical information, results reporting, orders, discharge referrals and operating theatres. Introduced in 2008, the EDIS has some unique tables and share some tables (e.g. personal details) with the EMR. The EDIS enables ED clinicians to access previous ED attendance and inpatient admis-

sion information, such as investigations and discharge summaries, from any area health service unit. The terminology (coding system) used is a part of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT). Only one 'principal diagnosis' is recorded, which limits the ability to identify cases of interest should it not be labelled as the 'principal diagnosis'. All ED clinicians (doctors, nurses and allied health professionals) can input data and receive a 3 h training session on the EDIS. The only mandatory field in the EDIS is related to poisoning; otherwise there are no explicit rules/directions governing data entry. This ED is not accredited for ACEM training.

A list of all patients who attended the ED during the 2009 calendar year and all patients admitted with a principal diagnosis of diabetes mellitus (DM), CVD, asthma or COPD was generated from the EDIS, which uses the SNOMED-CT codes for recording the diagnosis. The EDIS list included patient identifiers: sex, date of birth, age, discharge disposition, arrival date and time, triage category, time of triage conducted, presenting problem, diagnosis, mode of arrival, admitted ward, adult medicine officer, length of stay, time seen by doctor, discharge date and time, specialty, referral on discharge and referral to facility. The admission data report included patient identifiers, admission date and time, discharge date and time, SNOMED-CT code, corresponding SNOMED-CT description and the name and contact details of the primary general practitioner (GP). The list of principal diagnoses was checked by two clinicians in the research team (DM and S-TL) to ensure that all synonyms and acronyms for DM, CVD and asthma/COPD were included. The relevant key words from ICD-10-AM and the principal diagnoses in the EDIS were then combined to form a comprehensive list of search terms to generate a list of patients with a definite diagnosis.

Because some patients with these selected chronic diseases might have a less obvious but related principal diagnosis, we defined possible diagnoses to include, for example, chest pain for CVD or shortness of breath for asthma/COPD. Table 1 describes the key words that describe the definite and possible diagnoses.

Five members of the research team (HYC, DM, SV, JB and STL) independently confirmed the accuracy of these possible diagnoses of DM, CVD or asthma/COPD by auditing the discharge summaries of these patients, available as digital documents in the EMR. Interrater consensus was achieved by discussion. The discharge summary information was designated as the benchmark for comparison. Where information was not available for the specific admission, the discharge summaries

Table 1. Key words for *definite* and *possible* diagnoses of patient with specific chronic diseases

	Definite diagnoses (key words)	Possible diagnoses (key words)
CVD	Acute coronary syndrome, angina, ischaemic heart disease, Myocardial infarct, ST segment elevation myoc, and NSTEMI	Chest wall pain, chest pain (discomfort/tightness), and atypical chest pain
DM	Diabetes, hyperglycaemia, hyperosmolar non-ketotic, metformin, insulin injection, and NIDDM	Hypoglycaemia, metabolic acidosis, and persistent hyper-insulinaemic
COPD/asthma	COPD (chronic obstructive pulmonary disease), asthma, chronic obstructive airway, chronic airflow limitation, chronic airway disease/obstruction, bronchiectasis, bronchitis, bronchial spasm, and emphysema	Shortness of breath, respiratory distress, and dyspnoea

CVD, cardiovascular disease; DM, diabetes mellitus; NIDDM, non-insulin-dependent diabetes mellitus; NSTEMI, non-ST-elevation myocardial infarction.

of the previous and subsequent admissions were audited to ensure that patients who have had a diagnosis made at other admissions were not excluded. There were therefore two levels of accuracy – for the specific episode of care and for a period around the specific episode. This meant that the report generated was as inclusive as possible.

To examine for potential sources of errors and inconsistencies within the EMR system, the principal diagnoses obtained via the EDIS module was compared with principal diagnoses identified directly from the tables of the EMR, using SNOMED-CT terms equivalent to those used for the list generated via the EDIS.

Simple descriptive and analytic statistics were used. The distribution of personal characteristics and clinical factors, such as triage category, were examined by the χ^2 trend test. SAS Power and Sample Size 3.1¹¹ was used to calculate the power of the study to test the null hypothesis that ‘90% the selected principal diagnoses reported from EDIS will be present in the discharge summary’.

Finally, the findings were discussed as part of a data-quality management process with local clinicians (three junior medical officers and one staff specialist) and two information managers associated with the EMR to obtain an insight into any technical and professional and organizational reasons for any data quality or system consistency issues found.

Results

The ED treated 33 115 patients during the study period (2009). A total of 2559 patients with a principal diagnosis of DM ($n = 118$), CVD ($n = 1392$) or asthma/COPD ($n = 1049$) attended ED (Table 2). Of these patients, 876 were admitted; the admission pattern was similar across

the diagnostic groups, with the lowest rate for CVD (31%) and highest for DM (40%).

A consistent and significant rate of missing discharge summaries was found for 12–15% for specific diagnoses across the four chronic diseases. When combined with terms for possible diagnoses, the rate of missing information was 11–20%.

The discharge summary audit (Table 2) excluded missing discharge summaries from the denominator ($n = 730$). It confirmed up to 61% of the definite diagnoses (definite $n = 436$) for the specific episode and up to 96% for the period, that is, when previous and subsequent episodes were included with the specific episode: DM (75% for specific episode and 97% for period), asthma/COPD (62% and 90%) and CVD (64% and 77%). When the possible (possible $n = 440$) cases are included, the figures (total $n = 730$) are: DM (71% for specific episode and 97% for period), CVD (59% and 70%) and asthma/COPD (48% and 74%).

Generally, fewer cases of ED attendees were obtained from the EMR tables directly (definite $n = 764$; total $n = 1921$) compared to obtaining them via an EDIS report (definite $n = 1059$; total $n = 2559$). Agreement on the principal diagnoses (Table 3) was comparable among the chronic diseases: DM (60% for specific episode and 74% for period), asthma/COPD (51% and 69%) and CVD (58% and 81%). When the possible ($n = 1157$) diagnoses are included, the figures are: CVD (87% for episode and 92% for period), DM (69% and 84%) and asthma/COPD (38% and 51%).

The indicators of ‘information quality’ (% confirmed for episode) were cross-tabulated with some patient attributes, clinical discipline and indicators of workload and workflow, such as Triage Category, waiting time and length of stay. The cross-tabulations did not achieve statistical significance ($P > 0.05$).

Table 2. Concordance of principal diagnosis from the EDIS with diagnosis from audit of discharge summary

	Audit of discharge summaries of patients admitted from the ED					
	EDIS	<i>n</i> = <i>N</i> – missing	Missing (%)	% confirmed (95% CI) (episode)	% confirmed (95% CI) (period)	Not a case (95% CI)
CVD definite Dx	198	170	28 (14.1)	64.1 (56.7–70.9)	12.9 (8.7–18.8)	22.9 (17.3–29.8)
CVD possible Dx	306	233	73 (23.9)	54.5 (48.1–60.8)	10.3 (7.0–14.9)	35.2 (29.3–41.5)
CVD subtotal	504	403	101 (20.0)	58.6 (53.7–63.3)	11.4 (8.7–14.9)	30.0 (25.8–34.7)
DM definite Dx	33	28	5 (15.2)	75.0 (56.6–87.3)	21.4 (10.2–39.5)	3.6 (0.6–17.7)
DM possible Dx	14	10	4 (28.6)	60.0 (31.3–83.2)	40.0 (16.8–68.7)	0.0 (0–27.8)
DM subtotal	47	38	9 (19.1)	71.1 (55.2–83.0)	26.3 (15.0–42.0)	2.6 (0.5–13.5)
COPD/asthma definite Dx	205	180	25 (12.2)	61.7 (54.4–68.5)	28.3 (22.3–35.3)	10.0 (6.4–15.3)
COPD/asthma possible Dx	120	109	11 (9.2)	26.6 (19.2–35.6)	21.1 (14.5–29.7)	52.3 (43.0–61.4)
COPD subtotal	325	289	36 (11.1)	48.4 (42.7–54.2)	25.6 (20.9–30.9)	26.0 (21.2–31.3)

CI, confidence interval; COPD, chronic obstructive pulmonary disease; CVD, cardiovascular disease; DM, diabetes mellitus; Dx, diagnosis; EDIS, ED information system.

Using SAS,¹¹ we determined that the study had sufficient power (>0.9) to detect the difference in accuracy between the EDIS report and EMR discharge summary with alpha 0.05 for all the selected chronic conditions except diabetes. In the case of diabetes, the null hypothesis of 90% concordance overall was rejected for specific episodes but accepted for the period.

Discussion

The study examined the accuracy of the diagnoses in the EDIS (against the discharge summary diagnoses) and the consistency of the EMR system by comparing reports generated by the EDIS and standard queries of the EMR database. The variations in information accuracy and system consistency, along with the 12–15% missing discharge summaries, raise important questions for clinicians, services and information managers and researchers.

Allowing for these variations, generating reports using comprehensive clinical terms for specific diseases can reliably identify up to 61% of definite diagnoses for a specific attendance. Using less specific key words for possible cases can increase the yield. If we include an audit of the previous and subsequent discharge summaries, the yield can be up to 96%. This is a significant improvement, although some might question the cost-efficiencies with the added human resources required.

The EDIS/EMR technical support and information managers consulted suggested that the use of specifically tailored clinical terms to generate a report from the EDIS would give a better yield than the use of equivalent SNOMED-CT terms; that is, using codes to retrieve records means some patients with the diagnosis will be missed. They believed that the current implementation of SNOMED-CT as the reference terminology in the EDIS, replacing the current ICD-10-AM, was not likely to be a reason for the EDIS–EMR discrepancy. Australian health information managers have consistently used ICD-9-CM and ICD-10-AM well,⁹ with coding discrepancy being more frequently caused by universal difficulties associated with the coding of hospital records, especially more complex cases, than any unique local problems.¹² Coding validity tends to improve as coders gain experience with new coding systems, for example, transition from ICD-9 to ICD-10-AM.¹³

From the ED perspective, these variations suggest potential ‘errors’ at different points in the information generation cycle: from data collection and entry to storage and use. The differences are consistent with a

Table 3. Comparing diagnostic data from the EDIS with diagnostic data from tables of the EMR

	Dx reported via EDIS	Dx reported via tables of the EMR					
		% confirmed (episode)		% confirmed (period)		% confirmed (episode + period)	
		<i>N</i>	<i>n</i> % (95% CI)	<i>n</i> % (95% CI)	<i>n</i> % (95% CI)	<i>n</i> % (95% CI)	<i>n</i> % (95% CI)
CVD definite diagnoses	228	132	57.9 (51.4–64.1)	52	22.8 (17.8–28.7)	184	80.7 (75.1–85.3)
CVD possible diagnoses	1164	1077	92.5 (90.9–93.9)	23	2.0 (1.3–3.0)	1100	94.5 (93.0–95.7)
Subtotal	1392	1209	86.9 (85.0–88.5)	75	5.4 (4.3–6.7)	1284	92.2 (90.7–93.5)
DM definite diagnoses	70	42	60 (48.3–70.7)	10	14.3 (8.0–24.3)	52	74.3 (63.0–83.1)
DM possible diagnoses	48	39	81.3 (68.1–89.8)	9	18.8 (10.2–31.9)	48	100.0 (92.6–100.0)
Subtotal	118	81	68.6 (59.8–76.3)	19	16.1 (10.6–23.8)	100	84.7 (77.2–90.1)
COPD/asthma definite diagnoses	761	393	51.4 (48.1–55.2)	135	17.7 (15.2–20.6)	528	69.4 (66.0–72.6)
COPD/asthma possible diagnoses	288	9	3.1 (1.7–5.8)	0	0.0 (–)	9	3.1 (1.7–5.8)
Subtotal	1049	402	38.3 (35.4–41.3)	135	12.9 (11.0–15.0)	537	51.2 (48.2–54.2)

CI, confidence interval; COPD, chronic obstructive pulmonary disease; CVD, cardiovascular disease; DM, diabetes mellitus; Dx, diagnosis; EDIS, ED information system; EMR, electronic medical record.

Scottish ED study where a discharge diagnosis accuracy of 87% have been reported⁶ and comparable with a sensitivity of 85.1% and specificity of 60.4% demonstrated by a Swiss study of electronic queries about drug-related problems, conducted by clinical pharmacists, from a hospital EMR.¹⁴ Data quality issues are not unique to Australia.

What are some reasons for this imperfect information quality?

The major causes of unavailable discharge summaries reported by clinicians include a lack of time or doctors having finished their clinical rotations before completing their paperwork. Hospital policies and work environments that are not supportive will not promote a culture that values good documentation. An unsupportive culture also means little training and support for clinicians to collect good-quality data for clinical information systems as well as system barriers, such as workstations being unavailable, or technical barriers, such as unfriendly or unusable user interfaces or an inadequate terminology to cover the range of clinical options. In addition to change management, embedding efficient data collection into the clinical workflow is fundamental to the implementation of clinical information systems.¹⁵

The ED clinician is the primary point of contact for many patients with acute health problems. Uncertainty is inherent at point of first contact because the clinical picture is usually unclear and evolves and changes with time in the ED and during the transition to admission. For instance, chest pain managed appropriately as 'pos-

sible acute coronary syndrome' or 'suspected myocardial ischaemia' in the ED might declare itself as a peptic ulcer following the patient's admission into CCU. This dynamic nature of ED care requires an accurate and complete history and its documentation to build a knowledge base to enable better understanding of the clinical course of acute symptoms. An incomplete history and documentation might miss comorbidity, for example, diabetes or CVD, which can lead to inadvertent changes in medications or unsafe treatments. Just as important, an improved recording culture will enable the use of electronic decision support systems¹⁶ to improve diagnosis and management in the ED.¹⁷

What needs to be done?

Training and assisting clinicians to enter data into structured data fields is a key strategy to address the 'missing data' issue as much of the required information is often found in the narrative text of discharge summaries. Customized templates can improve the quality of documentation and can also prompt best clinical practice.¹⁸ Other strategies include workflow re-engineering, change management, training and support and regular data audits and feedback to enable and maintain a culture of good documentation and structured data entry. There is an urgent need to understand the personal, professional, organizational and socio-technical factors that might influence the completeness and accuracy of data in clinical information systems.

Software engineering is required to ensure reliability and validity at the user interface and in the database architecture with well-designed software. Regular audit

of data quality and feedback is essential to avoid the adverse impacts of poor-quality information in eHealth tools and systems. Workplace training and support in the consistent use of the EMR are essential. Education providers must incorporate informatics competencies in their curricula, aiming to graduate specialist health informatics practitioners as well as eHealth-savvy clinicians and managers.

Emergency department information should also be linked to similar information from other health professionals involved with the patient's care before and after attending ED. A lack of standard protocols for referral, preadmission, discharge and other 'clinical handover' arrangements between services and sectors is a key issue as is the lack of routine feedback and accountability protocols to ensure the quality of data exchanged.

A systematic study of information quality, including 'consistency, correctness, currency, time-stamped, meet industry standards, and relevance'¹⁰ of data repositories created and used by different primary and secondary care services will provide a comprehensive and accurate picture of the health and care of individuals and communities. It will improve policy development, planning and implementation, and quality monitoring. A data-quality metric to specify and measure the quality requirements for a database as well as techniques, such as quality policies and data-quality systems, is essential. We need to understand the link between poor data quality and procedures to detect and eliminate problems during data product creation, the costs of external data failure and complementary costs of data-quality assurance.¹⁹

Finally, the study is constrained by the lack of a cost-efficient way to obtain similar information for all patients with chronic diseases who attended ED and were not admitted. This exclusion of ED attendees who were not admitted meant that only the sensitivity (% agreement) could be measured. Studies with sufficient resources to enable the calculation of Kappa and specificity as well as sensitivity are needed.

Conclusions

The present study highlights what we already know – we are only as good as the information we have! The various purposes designated for the EMR as part of health reform are unrealistic if we cannot rely on data quality and consistency of our information systems. The current information quality and consistency of the

EDIS/EMR system in a typical community hospital with an unaccredited ED needs improvement to easily support research. If it is not good enough for research, it is probably not good enough to promote and support safe and effective clinical practice.

Further systematic examination of the large repositories of routinely collected data is required to determine the fitness for purpose, which might be to share good-quality information among clinicians or to support electronic decision support systems. A logical next step is to repeat the present study with a representative sample of EDs and include a qualitative study of clinicians and information managers to understand the underlying reasons for any variations in data quality.

Acknowledgements

We'd like to extend our thanks to the SSWAHS Divisions of GP Executive Liaison Committee, SSWAHS Information Management and Technology Directorate, Fairfield Hospital ED (Prairiewood, NSW, Australia), Enrico Coiera, Mark Harris, Gawaine Powell-Davies, Elizabeth Comino and staff at the SSWAHS/University of New South Wales GP Unit. We acknowledge our co-author JB who passed away on 4 May 2011.

Author contributions

STL conceptualized and directed the implementation of the project and led the writing of the paper. HYC coordinated the project and conducted the statistical analyses. All authors contributed to the medical records audit and abstraction as well as to the interpretation of the results and writing of the paper.

Competing interests

None declared.

Accepted 2 August 2011

References

1. National Health & Hospital Reform Commission. *A Healthier Future for All Australians – Final Report of the National Health and Hospitals Reform Commission – June 2009*. Canberra: Department of Health & Ageing, Commonwealth of Australia, 2009.

2. Australian Government Department of Health & Ageing. *Building a 21st Century Primary Health Care System. A Draft of Australia's First National Primary Health Care Strategy*. Canberra: Commonwealth of Australia, 2009.
3. National Preventive Care Taskforce. *Australia: The Healthiest Country by 2020 – National Preventative Health Strategy – Overview*. Canberra: Commonwealth of Australia, 2009.
4. Garling P. *Final Report of the Special Commission of Inquiry: Acute Care in NSW Public Hospitals, 2008 – Overview*. Sydney: NSW Government, 2008.
5. Azaouagh A, Stausberg J. [Frequency of hospital-acquired pneumonia – comparison between electronic and paper-based patient records]. *Pneumologie* 2008; **62**: 273–8.
6. Mitchell J, Westerduin F. Emergency department information system diagnosis: how accurate is it? *Emerg. Med. J.* 2008; **25**: 784.
7. Moro ML, Morsillo F. Can hospital discharge diagnoses be used for surveillance of surgical-site infections? *J. Hosp. Infect.* 2004; **56**: 239–41.
8. Lain SJ, Roberts CL, Hadfield RM, Bell JC, Morris JM. How accurate is the reporting of obstetric haemorrhage in hospital discharge data? A validation study. *Aust. N. Z. J. Obstet. Gynaecol.* 2008; **48**: 481–4.
9. Quan H, Li B, Saunders LD *et al.* Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv. Res.* 2008; **43**: 1424–41.
10. Wang RY, Pierce EM, Madnick SE, Fisher CW, eds. *Information Quality*. Armonk: ME Sharpe, 2005.
11. SAS Institute. *SAS: Statistical Software. 9.2 Ed.* Cary: SAS Institute, 2008.
12. MacIntyre CR, Ackland MJ, Chandraraj EJ, Pilla JE. Accuracy of ICD-9-CM codes in hospital morbidity data, Victoria: implications for public health research. *Aust. NZ J. Public Health* 1997; **21**: 477–82.
13. Henderson T, Shephard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. *Med. Care* 2006; **44**: 1011–19.
14. Roten I, Marty S, Beney J. Electronic screening of medical records to detect inpatients at risk of drug-related problems. *Pharm. World Sci.* 2010; **32**: 103–7.
15. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. *BMC Health Serv. Res.* 2010; **10**: 231.
16. Michalowski W, Farion KJ, Pike J, Rubin S. Design and development of a mobile system for supporting emergency triage. *Methods Inf. Med.* 2005; **44**: 14–24.
17. Horsky J, Gutnik L, Patel VL. Technology for emergency care: cognitive and workflow considerations. *AMIA Annu. Symp. Proc.* 2006; **2006**: 344–8.
18. Ragoo MA, McNaughton G. Improving documentation of head injured patients admitted to the emergency department ward. *Scott. Med. J.* 2005; **50**: 99–100.
19. Wang RY, Storey V, Firth C. A framework for analysis of data quality research. *IEEE Trans. Knowl. Data Eng.* 1995; **7**: 623–40.