Evaluation of a computer tracking program for resident-patient encounters

BRIAN H. ROWE, MD, CCFP(EM)
DALE T. RYAN, BSC, BED
JOHN V. MULLOY, MD, MCFP(EM)

OBJECTIVE To examine the effectiveness of a formal tracking system for residents' clinical experiences.

DESIGN We examined three shifts, selected at random, for each resident (without residents' knowledge) during emergency rotations. Information from patient charts was compared with residents' computerized records for rotation (location and preceptor) and patient (age, sex, diagnosis, and procedure) information.

SETTING The Northeastern Ontario Family Medicine Program, a program designed to provide remote, rural, and northern resident experience.

PARTICIPANTS First-, second-, and third-year residents on emergency rotations in the academic years 1992 to 1994.

MAIN OUTCOME MEASURES Compliance, reliability, and validity of tracking records.

RESULTS Residents recorded patient encounters 88% of the time. Compliance with rotation information was high (100% rotation, 94% preceptor). Agreement on patient age and sex was high. Procedure compliance was somewhat lower (83%). Intrarater reliability (91%) and inter-rater reliability (78%) are acceptably high, as is validity when compared with a gold standard entry (88%).

CONCLUSIONS Regular entry of reliable and valid data is facilitated by the computerized resident-patient encounter tracking program. This computer tool should prove useful for multilevel program evaluation in the future.

OBJECTIF Examiner l'efficacité réelle d'un système formel de suivi pour l'expérience clinique des résidents.

CONCEPTION Pour chacun des résidents, nous avons examiné trois périodes de travail choisies de façon aléatoire (sans en informer le résident) pendant les stages de médecine d'urgence. On a comparé les renseignements notés au dossier des patients à ceux des dossiers informatisés des résidents lors de ce stage (lieu et précepteur) et les renseignements concernant le patient (âge, sexe, diagnostic et intervention).

CONTEXTE Le programme de médecine familiale du Nord-Est de l'Ontario, conçu pour offrir au résident une expérience nordique en régions rurale et éloignée.

PARTICIPANTS Les résidents de première, deuxième et troisième années pendant leur stage de médecine d'urgence au cours des années universitaires 1992 à 1994.

PRINCIPALES MESURES DES RÉSULTATS Observance, fiabilité et validité des dossiers de suivi.

RÉSULTATS Les résidents ont inscrit 88% de leurs rencontres avec les patients. La fidélité de l'information entourant le stage fut élevée (stage 100%, précepteur 94%). La concordance sur l'âge et le sexe du patient fut forte. Quant aux interventions, la fidélité fut un peu plus faible (83%). Comparées à une entrée «étalon-or», la concordance entre l'observateur lui-même (91%) et la concordance inter-évaluateurs (78%) furent élevées tout comme ce fut le cas pour la validité (88%).

CONCLUSIONS Le programme informatique du suivi des rencontres résident-patient facilite l'entrée régulière et fiable de données valides. Cet outil informatique devrait s'avérer utile dans le futur pour l'évaluation d'un programme à multiples niveaux.

Can Fam Physician 1995;41:2113-2120.

Dr Rowe *is Research Director of the Northeastern Ontario Family Medicine (NOFM) Program and of the Sudbury General Hospital Lead Trauma Program.*
Mr Ryan *was a Research Assistant with NOFM and is now a medical student at McMaster University.*
Dr Mulloy *is Program Director at NOFM and Director of the Emergency Department at Sudbury Memorial Hospital.*

EXAMINING MEDICAL STUDENTS', interns', and residents' records of clinical experiences is one way of evaluating both an individual learner's clinical training exposure and the overall experience offered by the training program.[1,2] The time-honoured tradition of the informal diary has been replaced recently by specific log or tracking systems.

There are several reasons for this trend toward formal data collection. Most notably, in Canada and other

nations with remote and rural training needs, programs have become decentralized. To ensure standardized training, decentralized programs need thorough evaluation.[3-6]

A variety of logs are currently available, including generic computer tracking programs, program-specific tracking programs, billing package summaries, and written records.[1-16] The most popular are written records and individualized computer tracking programs. Written records have the advantage of simplicity and low cost. However, with the advent of inexpensive computer programs, the benefits of computer tracking become obvious: easy data management and analysis, potentially fewer transcription errors, and participant entry of data.

Computer programs have been developed specifically to track patient encounter experiences. Despite the many different tracking formats, their psychometric properties are rarely, if ever, investigated before use.[1] If we are to incorporate such methods of data collection into multifaceted program evaluation, we must make a formal scientific study of learners' compliance with data entry and of data validity and reliability.

This project was designed to examine the psychometric properties of residents' entries into a computer tracking program currently used by all residents in the Northeastern Ontario Family Medicine (NOFM) Program. Specifically, we examined resident compliance with data entry during emergency medicine rotations to determine the validity of the data entered and to suggest factors that improve compliance and data validity.

METHODS

Setting

The University of Ottawa supports a family medicine residency program in Northeastern Ontario. This program is designed to provide residents with experience in remote, rural, and northern medical practices. There are 12 residents in each of the 2 years of the program, and two third-year residents specializing in emergency medicine.

Residents rotate through 16 weeks of family medicine and 8 weeks of rural family medicine in first and second year. Each resident completes 4 weeks of emergency medicine rotations per year. During the 2 years, residents also get training in obstetrics and gynecology (12 weeks), internal medicine (12 weeks), pediatrics (8 weeks), surgery (4 weeks), and behavioural medicine (4 weeks). Eight weeks are reserved for elective experience.

Tracking program

Each resident is given a diskette containing a computer program designed to track resident-patient exposure in a variety of settings.[2] Residents are instructed on how to use the program during orientation. Residents are required to record all patient encounters in all rotations. They can either enter data directly as patients are seen or keep records on paper and transfer them.

Once opened, the program is visible on one computer screen and prompts residents for the required information. At every entry field, a key (F9) allows residents to scroll through available options. After data entry, residents save the data to the disk. This process should be completed for each encounter. Computers are made available wherever residents complete core rotations: laptops in remote locations and personal computers elsewhere.

Residents return disks monthly or every 2 months depending on location. This type of tracking provides information on:

- the resident (ID number and name);
- the patient (age, sex, diagnosis, procedures);
- the rotation (setting, preceptor name, specialty); and
- the degree of responsibility for diagnosis and procedures (eg, whether residents performed, assisted, or observed).

The classification of diagnoses and procedures is unique to this computer program[2]; it represents a compromise between complexity and user-friendliness. Diagnoses are grouped by body system into 12 categories. Procedures are grouped into five categories, such as orthopedic and gynecologic; there are 80 possible procedures.

Selection of rotations

We examined resident rotations in emergency medicine during the 1992 to 1994 academic years. These rotations were chosen because they are an important component of primary care training, are busy, and include frequent procedures. Daily encounter records are regularly kept and are readily available. (In contrast, data from family physicians in community practice were sketchy and incomplete because the practices' billing systems were not linked and did not regularly identify which patients were seen by residents and which by staff.)

Each resident's daily encounters were assessed on 3 randomly selected days in the Sudbury General Hospital Emergency Department. Residents were unaware of the research study; however, results were made available following evaluation. Only aggregate data were reported.

Patient information was examined and recorded before examining residents' computerized data records in order to reduce bias. Since assisting or "show-and-tell" encounters are not easily tracked, we thought compliance data might be misleading. Information on such encounters, therefore, was not examined. Data were collected only from encounters in which residents were primarily responsible (ie, the resident had interviewed and examined the patient as the health care provider). In the emergency department, this was indicated by the resident having completed and signed the chart.

Each record was examined by two researchers, and disagreement was resolved through consensus or third-party ruling. Diagnoses and procedures were recorded from patient charts. We noted the number of days from patient encounter to data entry.

Compliance

We assessed resident compliance with data entry in the following ways.

Entry of any data for the rotation. This measure has been used by others and is difficult to interpret, since NOFM requires residents to complete this task as part of the program.

Exact compliance with data entry. Correct patient age, sex, diagnosis, date of encounter, and rotation should be entered.

Partial compliance. If diagnosis, patient age and sex, rotation, and procedure could be identified on another day, we assumed the resident had entered data correctly, except for the date.

Shift compliance. Overall shift compliance was determined for each day examined and represents the number of patients recorded for the shift divided by the number of patients actually seen. For example, if a resident had seen 10 patients on a certain shift, 10 records should have been entered in the tracking program. If the resident did not turn in a diskette, compliance would be 0%. If the diskette was returned and only eight patients were entered on that day, "exact" and "shift" compliance would be 80%. If those eight encounters were found to be entered under the previous day's records, exact compliance would be 0%, while the partial and shift compliance would be 80%.

Reliability and validity

Using a random numbers table, samples of the first 320 encounters were selected to collect five cases from each of the 12 body-system categories; these 60 cases were used for reliability testing.

We identified three levels of experience: experts (tracking program authors and clinicians), experienced

(second-year residents), and inexperienced (first-year residents). Two residents were chosen at random from among the first- and second-year residents and asked to participate in the reliability study.

Residents and clinicians were asked to code primary diagnosis, age, and sex from the list of cases. All confidential information was deleted from the chart. Residents were blinded to the original code for the chart and the reasons for the study. At least 3 weeks after completing the first rating, residents and staff were asked to repeat the task on a similar list. All coders were masked to any previous coding.

Table 1. **Compliance with tracking**

COMPLIANCE	% (95% CI)
Preceptor (name correct for day of encounter)	94 (87-89)
Exact (correct diagnosis, date, and rotation)	77 (68-85)
"Ever" (correct diagnosis, age, sex, rotation, and procedure identified on another day)	88 (80-94)
More than 75% correct patient records as a proportion of total encounters on a single shift	89 (74-90)
Overall	83 (74-90)

Validity was determined by comparing the resident's entry with the appropriate "gold standard" diagnosis determined a priori by consensus of the authors. To be considered gold standard, the diagnosis appearing on the chart had to be classified into a single body system and a specific diagnosis. In several ambiguous cases, two or more diagnoses were accepted as valid. The simple percentage of valid entries was then determined by comparing the number of valid diagnoses for each reviewer.

Data analysis

Data were analyzed using Statview SE statistical software.[17] Categorical data were reported as counts and proportions, continuous data as means and standard deviations. Compliance data were reported as counts and percentages.

Agreement on dichotomous variables is reported as a κ statistic,[18] agreement on continuous variables as a correlation coefficient. Agreement statistics were evaluated using previously accepted standards for κ values: excellent ($\kappa \geq 0.8$), good ($0.6 \leq \kappa < 0.8$), moderate ($0.4 \leq \kappa < 0.6$), fair ($0.2 \leq \kappa < 0.4$), and poor ($\kappa < 0.2$).[18] Student's *t* test and analysis of variance (ANOVA) were used to identify differences in compliance based on sex, shift time, and level of training.

RESULTS

We examined 25 resident rotations from October 1992 to May 1994. On average, we examined three shifts for each resident (one day, one afternoon, and one night shift) for a total of 77 shifts and 705 patients. The average number of patients seen per shift was eight for first-year residents, 10 for second-year residents, and 13 for third-year residents.

Patients' most frequent presenting complaints were musculoskeletal (32%), head and neck (12%), and abdominal or gastrointestinal (11%). Fewer patients presented with chest (8%), infective (7%), cardiovascular (6%), gynecologic (5%), neurologic (5%), dermatologic (5%), or psychiatric (4%) emergencies. Very few patients (< 2%) had endocrine, metabolic, or oncologic emergencies. These percentages closely reflect the pattern of emergency cases in our centre.

Compliance

Compliance with data entry was high (*Table 1*). Exact compliance was identified in 77% of all cases. When we searched for valid data entered under incorrect dates, compliance increased to 88%. In all, 89% of cases were correctly recorded and entered more than 75% of the time on a single shift.

Compliance with data entry on procedures was less impressive, although acceptable. Only 18% patients seen

required procedures; residents recorded 83% of these. No invalid entries were identified.

Compliance with data entry was inversely correlated to delay between the time a patient was seen and the time of data entry for exact ($P=0.02$) and "ever" ($P=0.04$) compliance. For ever compliant entries, no differences were identified based on residents' sex ($P=0.81$), shift time ($P=0.83$), or level of experience ($P=0.88$).

Reliability

The 60 randomly selected cases were given to each of the four residents and two "expert" coders. They were asked to recode primary diagnoses from the chart summary provided. Code-recode reliability (intra-observer) was determined by giving the residents 60 encounters from a previous coding. Interobserver reliability was determined by examining 60 standard cases. Overall, reliability was higher for intrarater coding (91%) than interrater coding (78%). No differences were apparent between expert and inexperienced raters.

Validity

Validity of coding against a gold standard was determined for the same 60 encounters used for reliability testing (*Table 2*). Overall, 88% of the diagnostic selections entered by the two staff and four residents were considered valid.

DISCUSSION

Using clinical encounter logs or tracking programs for evaluating both learner experience and program objectives is not yet widespread. The concept is still unfamiliar and still has not been proven effective. Logs can be difficult to keep, sometimes use complicated coding systems, and might lack proven psychometric stability. But use of these tracking methods makes eminent sense, especially for programs in which teaching experiences are distributed over different specialties and wide geographic areas.

Formal examination of resident data entry compliance and data validity and reliability is both worthwhile and necessary because such information allows learners to identify areas of weakness. The information might also be used to validate a learner's experience following completion of training so that he or she can apply for hospital privileges or further residency training. Tracking data can benefit the academic program as a whole by identifying the strengths and weaknesses of learning sites, rotations, and preceptors.[2] Comparisons of rotations and experiences between and within a program can be made. Other logs or tracking programs have been used successfully by medical students[1,3-5,7-10,12,13,15] and residents.[1,12,16] Such tracking has allowed comparisons of clinical experiences between learners[8,9,12] and across specialties.[1,3,16]

Table 2. **Percentage of entries of 120 encounters agreeing with "gold standard" diagnosis**

EXPERT 1	EXPERT 2	FIRST-YEAR RESIDENT	FIRST-YEAR RESIDENT	SECOND-YEAR RESIDENT	SECOND-YEAR RESIDENT
92% (95% CI 85-96)	88% (95% CI 80-94)	92% (95% CI 85-96)	85% (95% CI 76-91)	86% (95% CI 78-92)	88% (95% CI 80-94)
90% (95% CI 82-95)		88% (95% CI 80-94)		87% (95% CI 79-93)	
88% (95% CI 80-94)					

Our training program distributes residents over a large area of rural and northeastern Ontario. The inexpensive, readily available, user-friendly software has allowed us to provide evaluation and feedback to residents, preceptors, and accreditation bodies. However, such data are of little value, and potentially misleading, if they are not valid and reliable.

We believe this to be the first comprehensive attempt to examine the issues of validity, reliability, and compliance with encounter recording. Our results show that residents' data are both reliable and valid. Compliance with data entry is good, and strategies to improve data entry have reduced tracking errors.

Compliance

Compliance can be measured in different ways. For example, Vanek and colleagues[1] calculated compliance by the percentage of students submitting their log forms. Not surprisingly, compliance was high (97% and 100%, respectively, for medical students and clerks). This measure is too crude for our system: only one resident in 3 years did not track any encounter data, making compliance by this measure 97%. Data recording is now a compulsory component of the NOFM program.

The measures we use allow us to be confident that residents are tracking most patient encounters in a typical rotation. Errors in data entry are minor and appear to be mainly omissions (eg, residents frequently entered data for 2 days at the same time, without changing the date of encounter). On only two of the 75 shifts reviewed had residents not entered any data.

Residents do not appear to be fabricating data. If anything, residents tend to underrepresent rather than overrepresent their clinical encounters.

Reliability

Residents and staff alike demonstrated high intrarater reliability when 60 typical encounters were coded twice at different times. Not surprisingly, interrater reliability was lower, but still acceptable. These findings suggest that the data are reliable and can be used to examine the issues for which tracking was designed.[2]

Validity

All tracking programs have difficulties with diagnostic codes. A free text option gives too much leeway for accurate coding and aggregation of data; limiting the diagnostic choices can result in loss of information, frustrated coders, and erroneous conclusions from data analyses. Large diagnostic data sets, such as billing codes from centralized systems, appear too expansive; the options provided by others,[1] at least for primary care, appear too restrictive.

Our diagnostic coding, with approximately 245 options, is a compromise. Iterative resident feedback allows changes to facilitate data entry. Residents need not memorize codes: they are available on screen, grouped by category, through a single-key scroll function. Both clusters of related diagnoses and single diagnoses can be analyzed.

Our study examined validity of entry using the codes for body system and for exact diagnosis. Validity was excellent (88%). We have noticed that exact diagnostic validity improves when redundant diagnoses are removed (eg, upper respiratory tract infection was listed under two categories: head and infection).

Limitations and concerns

First, this study focused on emergency department records and might not accurately reflect compliance in other rotations. Second, the psychometric properties might be specific to computer-based data logs or to this computer program. Third, although residents were unaware of the testing, over time omission and fabrication could occur. Since resident compliance has not been reexamined, this issue awaits further study. Finally, logistical concerns, such as access to compatible computers

and resource requirements to maintain the database, might limit the use of programs such as this elsewhere. Despite these concerns, we believe our findings are generalizable to other programs where resident encounter information is being used for program evaluation.

Strategies to improve data entry

We have made several changes to the program to improve the accuracy of and compliance with data entry.

- Recommendations from residents have been solicited and applied to change the tracking program, including filling "holes" in the classification systems.
- We advise immediate data entry for maximum accuracy; immediate entry in a handwritten log for later transfer also reduces error.
- We encourage preceptors to allot time at the end of the day for residents to do data entry.
- Tracking is compulsory. Others have recommended this.[19]
- We have designated a research assistant specifically to work with residents on the tracking program.

Conclusion

Accurate, valid, and reliable data on residents' experiences with patients, sites, and preceptors can be collected. This computer tool should prove useful for multilevel program evaluation in the future. ■

Acknowledgments

We thank the residents of the NOFM Program for cooperating during development and implementation of this program. We especially thank the NOFM residents who helped with reliability testing. We also thank Mrs Cindy Cavanagh for securing resident emergency department encounter sheets and Dr Peter C. Rowe for his helpful comments during manuscript preparation.

***Correspondence to**: Dr Brian H. Rowe, Northeastern Ontario Family Medicine Program, Health Sciences Education Resource Building, Laurentian University, Sudbury, ON P3E 2C6*

References

1. Vanek EP, Barriga-Unal RM, Hekelman FP, Hull Al, Lindley BD, Barley GE, et al. Use of patient encounter documentation (log) systems at three medical schools. *Teach Learn Med* 1993;5:164-8.
2. Mulloy JV, Leuschen M, Rowe BH. Computer-based patient encounter tracking. Development of a system for family medicine residents. *Can Fam Physician* 1995;41:1742-51.
3. Ferrell BG. Demonstrating the efficacy of the patient logbook as a program evaluation tool. *Acad Med* 1991;66:549-51.
4. Hobbs J, Mongan P, Miller MD. A system for assessing clerkship experience using a logbook and microcomputers. *Fam Med* 1987;19:287-90.
5. Ruane TJ, Smith BWH. Using a computerized database to manage a decentralized ambulatory care clerkship. *Acad Med* 1990; 65:438-9.
6. Hobbs J, Mongan PF, Tollison JW, Miller MD, Wilson OR. A decentralized clerkship: strategies for standardizing content and instruction. *Fam Med* 1987;19:133-7.
7. Witzke DB, Koff NA, McGeagh AM, Skinner PD. Developing a computer-based system to document and report students' clinical patient encounters. *Acad Med* 1990; 65:440-1.
8. Parkerson GR, Michener JL, Muhlbaier LH, Falcone JC. Clinical experience of medical students in model family practices and private family practices. *J Fam Pract* 1986;23:361-6.
9. Rosen RL, Eshler B, Bone RC. Comparison of students' encounters with patients in university and community hospitals. *J Med Educ* 1986;61:601-3.
10. Rabinowitz HK. Student clinical experiences in family medicine: a comparison of clerkships and preceptorships. *Fam Pract Res J* 1989;8:92-9.
11. Duckwall JM, Alul IH, Arnold LM, Willoughby TL, Sirridge MS, Hamburger SC. Educating students in a general medicine ambulatory clinic: a pilot study. *Teach Learn Med* 1989;1:92-6.
12. Phillips WR, Rosenblatt RA, Gordon MJ, Fletcher RM. Clinical content of the WAMI community clerkship in family medicine. *J Med Educ* 1982;57:615-20.
13. Soler NG, Mast TA, Anderson MB, Kienzler LM. A logbook system for monitoring student skills and experiences. *J Med Educ* 1981;56:775-7.

Continued on page 2120

14. Links PS, Foley F, Feltham R. The educational value of student encounter logs in a psychiatry clerkship. *Med Teach* 1988; 10:33-40.
15. Munro JGC. Computer analysis of the student log diary: an aid in the teaching of general practice and family medicine. *Med Educ* 1984;18:75-9.
16. Bordley DR, Dolan JG, Hall WJ. Monitoring and improving the content of medical residents' ambulatory care experience. *J Gen Intern Med* 1987;2:174-7.
17. Abacus Concepts, Inc. *Statview SE.* Berkeley, Calif: Abacus Concepts, Inc, 1988.
18. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
19. Curtis P, Resnick J, Greganti MA. Documentation of inpatient experiences of resident physicians. *J Med Educ* 1985; 60:408-10.

• • •