Automated Data Mining: An Innovative and Efficient Web-Based Approach to Maintaining Resident Case Logs

Pratik Bhattacharya, MD, MPH
Renee Van Stavern, MD
Ramesh Madhavan, MD, DM

Abstract

Background Use of resident case logs has been considered by the Residency Review Committee for Neurology of the Accreditation Council for Graduate Medical Education (ACGME).

Objective This study explores the effectiveness of a data-mining program for creating resident logs and compares the results to a manual data-entry system. Other potential applications of data mining to enhancing resident education are also explored.

Design/Methods Patient notes dictated by residents were extracted from the Hospital Information System and analyzed using an unstructured mining program. History, examination and ICD codes were obtained and compared to the existing manual log. The automated data History, examination, and ICD codes were gathered for a 30-day period and compared to manual case logs.

Results The automated method extracted all resident dictations with the dates of encounter and transcription. The automated data-miner processed information from all 19 residents, while only 4 residents logged manually. The manual method identified only broad categories of diseases; the major categories were stroke or vascular disorder 53 (27.6%), epilepsy 28 (14.7%), and pain syndromes 26 (13.5%). In the automated method, epilepsy 114 (21.1%), cerebral atherosclerosis 114 (21.1%), and headache 105 (19.4%) were the most frequent primary diagnoses, and headache 89 (16.5%), seizures 94 (17.4%), and low back pain 47 (9%) were the most common chief complaints. More detailed patient information such as tobacco use 227 (42%), alcohol use 205 (38%), and drug use 38 (7%) were extracted by the data-mining method.

Conclusions Manual case logs are time-consuming, provide limited information, and may be unpopular with residents. Data mining is a time-effective tool that may aid in the assessment of resident experience or the ACGME core competencies or in resident clinical research. More study of this method in larger numbers of residency programs is needed.

Background

Education and clinical training during undergraduate and graduate medical education is an ever-evolving venture. Over the years, the Residency Review Committee for Neurology of the Accreditation Council for Graduate Medical Education (ACGME) has considered implementing individual resident case logs to assess the quality and breadth of resident education. Case logs are mandatory for several specialties, particularly procedure-based specialties like surgery, obstetrics-gynecology, and neurological surgery. In 2007, a survey of neurology residents[1] regarding the utility of a web-based, manual patient log system found that most residents entered only 50% of their cases, and 54% of residents were not able to enter the correct International Statistical Classification of Diseases and Related Health Problems (ICD-9) code for the diagnoses in more than 50% of cases. In addition, 53% of responding residents agreed or strongly agreed that case logs interfered with their education, and 40% agreed or strongly agreed that case logs interfered with patient care. Due to these perceived inadequacies, patient logs are now voluntary for neurology residents. Similarly, they are also voluntary for residents in pathology subspecialties and pediatrics. Under the current duty hour limits, and with the proposed institution of added limits on resident hours,[2] residents from a range of specialties would welcome more efficient methods to comply with case log requirements.

There is a need for technologies that comply efficiently with this ACGME recommendation while providing meaningful feedback to residents. Documentation of a resident's personal patient experience may provide useful

Pratik Bhattacharya, MD, MPH, is Vascular Neurology Fellow in the Department of Neurology, Wayne State University School of Medicine; **Renee Van Stavern, MD,** is Associate Program Director, Neurology at Washington University School of Medicine, St Louis; and **Ramesh Madhavan, MD, DM,** is Assistant Professor and Director of the Neurology Residency Program and Associate Chief Medical Officer for Informatics in the Department of Neurology, Wayne State University School of Medicine.

Corresponding authors: Pratik Bhattacharya, MD, MPH, 248.794.5148, pdbhatta@med.wayne.edu and Ramesh Madhavan, MD, DM, 312.745.1540, rmadhava@med.wayne.edu. Wayne State University/Detroit Medical Center, 8C UHC 4201 St Antoine, Detroit, MI 48201.

information and guide individual learning, especially in areas where there are gaps in the resident experience. Data mining is a technology that extracts hidden information from documents. It facilitates the analysis of large amounts of data that are extracted by scouring documents for hidden patterns.[3] Our study describes a novel alternative, using automated data mining, to manual entry into web-based programs. We present results from a pilot project to compare this innovative data-mining method with the existing method of manual web-based input, and we consider wide-ranging potentials for enhancing the neurology education experience.

Methods

We analyzed the common inpatient and outpatient experiences of all 19 neurology residents at Wayne State University School of Medicine/Detroit Medical Center for a period of 1 month. To test the application of data-mining technology for resident case logs, a pilot project was designed among the neurology residents at our institution. Outpatient notes, histories and physicals, consultations, and discharge summaries are routinely dictated by residents and are transcribed onto an electronic medical-records system. The commercially available data-mining system (Healthcare SmartGrid, Process Proxy, Ellwood City, PA) accepts different kinds of files for real time using Health Level Seven International (HL7) listeners.

For our study, one of the authors (P.B.) copied and pasted the electronic notes dictated by neurology residents into the secure mining program. Keywords (determined a priori) were specified for the data miner, and the miner ran the keywords as search terms and gathered information including resident information; patient demographics; reason for visit; comorbidities; habits such as smoking, alcohol, and drug abuse; date of visit; date of dictation; referring physician; and name of the staffing physician. The data miner also picked up important phrases from the impression and plan section of dictations to determine primary and secondary neurological diagnoses, and automatically assigned ICD-9 codes to these. All patient identifiers were removed. The output created by the miner was in the form of a Microsoft Excel spreadsheet.

At our institution, patient information is currently logged manually by residents in a web-based system (New Innovations Inc., Uniontown, OH) that can also be accessed on a PDA; this system provides hospital systems and medical schools with data-management suites for managing medical education, including duty hour documentation, evaluations, and curricula. Neurology residents currently log the following information: broad diagnostic category, setting of the patient encounter (outpatient/intensive care unit/regular floor admission), hospital name, and patient demographic information (adult/child, sex). While not mandated by the ACGME, the Department of Neurology expects all residents to log their continuity clinic patients as well as in-hospital patient encounters on an ongoing basis. The logs are reviewed in quarterly face-to-face meetings with the program director.

Healthcare SmartGrid, the experimental tool in this study, uses a patent-pending data-mining technique for other uses in the health industry. It is useful for analyses involving unstructured text but can also be used to tabulate, cross-reference, and compare large amounts of data. The process uses algorithms to analyze electronic health information (transcriptions of history and physicals, consults, emergency department notes, radiology dictations, etc), coding data (principal/admitting diagnosis, length of stay), admission/discharge/transfer data, lab results, pharmacy orders, and computerized physician order entry in virtually any format. Data are processed in a secure computer using Health Insurance Portability and Accountability Act–compliant measures. The program extracts the relevant information from this data, a form of data mining, to populate Excel tables, which then enables further analysis.

For our study, all logs entered by residents for the month in question were collected after a lag period of 1 month, allowing residents some time to manually enter their case logs. Dictated notes and manually entered case logs for a period of 1 month (August 2008) were reviewed. Data collected in the 2 systems were compared and analyzed. Finally, other potential applications of this system were explored, particularly the evaluation of core competencies.

The study was approved by the Wayne State University School of Medicine Institutional Review Board.

Results

For the manually entered web-based logs, 4 of 19 residents made a total of 192 entries during the study period. The manual method captured the setting of patient encounter: emergency room consultation 77 (40%), clinic consultation 56 (29%), neurology floor admission 23 (12%), neurointensive care unit consultation 25 (13%), medical intensive care unit consultation 10 (5%), and surgical intensive care unit consultation 1 (1%). Manual case logs documented only broad categories of disease, not specific diagnoses with ICD-9 codes. The major categories seen by the 4 residents were stroke or vascular disorder 53 (27.6%), epilepsy 28 (14.7%), pain syndromes 26 (13.5%), syncope/other alteration of consciousness 19 (9.8%), disorder of peripheral nerves 12 (5.5%), and muscle disease 7 (3.7%). Details are shown in FIGURE 1.

The semiautomated data-mining method processed information from all 19 residents. A total of 540 dictations were mined. The data miner captured additional information compared with the manual method. The various reasons for neurological evaluation were analyzed, and headache 89 (16.5%), seizures 94 (17.4%), and low back pain 47 (9%) formed the majority of the chief complaints. The data miner was also able to pick up patient

The numbers within the bars indicate actual numbers of diagnoses recorded by the manual method.
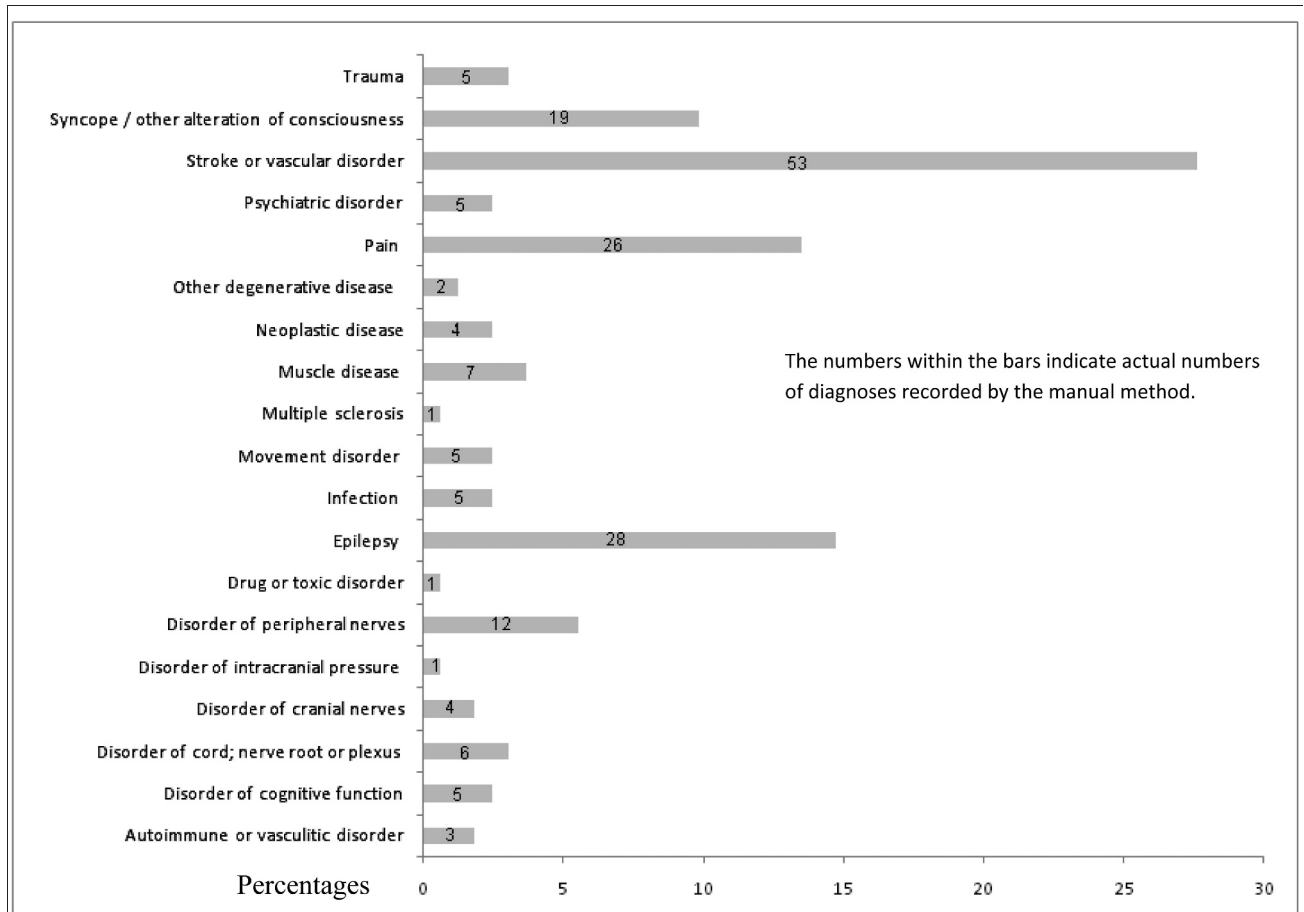
Percentages

FIGURE 1 | DISEASE CATEGORIES IDENTIFIED BY THE MANUAL METHOD (TOTAL 192 PATIENT ENCOUNTERS)

demographic data such as age, race, and sex, as well as whether the patients were smokers 227 (42%), used alcohol 205 (38%), or used drugs 38 (7%). Broad categories of disease could be obtained with the miner, and subcategories of specific disease states could also be determined (FIGURE 2). Epilepsy 114 (21.1%), cerebral atherosclerosis 114 (21.1%), and headache 105 (19.4%) were the most common primary diagnoses. The data miner was able to automatically assign ICD-9 codes based on primary and secondary diagnoses mined. Most importantly, the process of creating resident logs in this manner did not involve an additional time investment by residents. Two disease categories, trauma and neoplastic disease, were recorded by the manual method but not noted in the automated method. Retrospective review of the records showed that these encounters occurred in prior months and were logged in August 2008.

Discussion

Health care delivery systems have evolved rapidly in the 21st century. The incorporation of novel technologies in health care may improve quality of care, reduce errors, streamline processes, and enhance physician performance. Data-mining technology is used in various fields in the health care industry, primarily for performance assessment and quality improvement.[3]

Traditionally, the ACMGE has required case log data only in procedure-oriented specialties. However, properly maintained case logs have many other utilities across a range of specialties. Program directors could use the data to monitor the quality of education provided and to identify diagnoses to which residents are not being exposed. Residents could use the logs themselves to assess their strengths and weaknesses.

In this pilot project, we were able to demonstrate a unique application of data-mining technology to enhance resident education. The described method of creating patient logs has several advantages. First, the automated method did not require the resident's participation, saving time for residents to complete other tasks. In fact, if the program can be set up to directly transfer electronic medical-record information to the miner, the process would not need any time investment on the part of residents or residency program staff. Second, adopting an automated
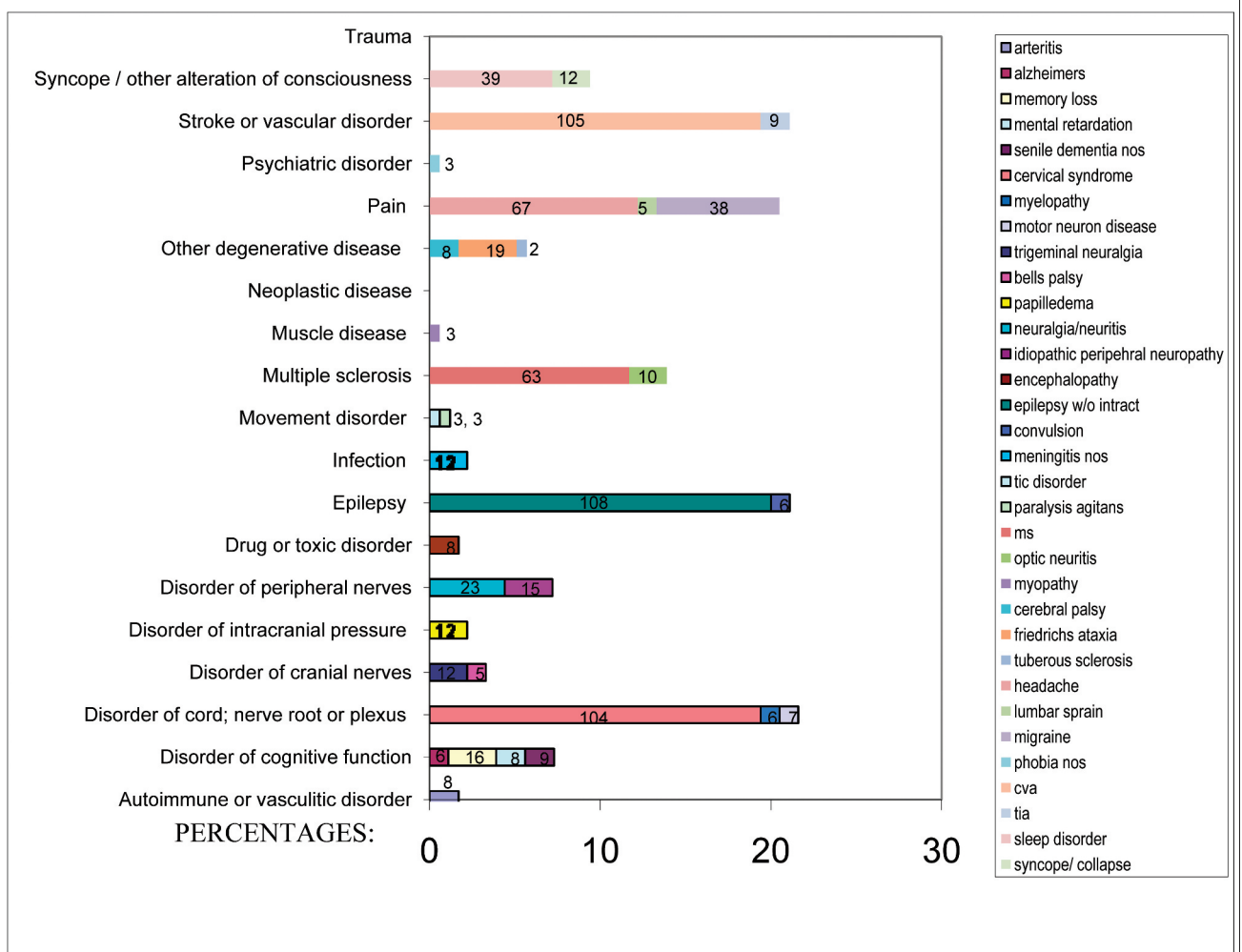
system generated a complete and accurate record of patients, while the completeness of a manual, web-based record depends on residents' recall and attention to timely entry. Third, the automated system was able to retrieve more precise diagnostic information, rather than broad disease categories in the manual web-based system. This would help the ACGME assess the quality of education in different residency programs, could provide additional data to program directors to make necessary program changes, and could also be a tool for individual resident reflection and assessment of educational needs. Finally, the automated method captured additional demographic information about patients that may prove useful in resident research projects. Electronic chart reviews would become less labor intensive. Residents can use this tool effectively for

continuous quality improvement projects by determining current practices and the effects of different practice interventions to improve them.

Like other specialties, neurology has looked for tools for assessing the 6 ACGME competencies,[4] and the data miner could assist in this. The tool could screen dictations for keywords indicative of appropriate documentation and management issues (patient care); diagnosis and differential generation, investigative work-up and analysis (medical knowledge); patient education and counseling (interpersonal skills and communication); use of ancillary services like physical and occupational therapists, speech therapists, nutritionists, case managers, and social workers (systems-based practice); and addressing referring physicians and timeliness of dictations (professionalism).

Residents may be assessed at several time points over the course of their residency to evaluate practice-based learning and improvement. The system could be set up to send periodic e-mails to residents giving them feedback about their performance on the various core competencies and could assist in self-assessment and improvement. The data-mining technique also could enhance the quality of resident case logs by giving electronic alerts to residents who do not comply with treatment guidelines. Additionally, the system could look for keywords to suggest certain risk classes (eg, disabled patients at risk for falls) and could provide alerts for appropriate risk management strategies. It can identify current and previous smokers, and treating physicians may be alerted to target them for counseling and pharmacologic treatment.

Finally, the tool may be used to provide incentive or credit to residents (medical knowledge during work) and continuing medical education credits to practicing physicians on a daily basis. The primary and secondary diagnoses mined out of dictations may be used as search terms in web-based academic resources. For each patient, an analysis of relevant evidence-based medicine guidelines obtained from Web resources can be provided to the physician in easily digested formats. The program can then provide continuing medical education credit to physicians who subsequently demonstrate adherence to the guideline. This method can help physicians to gain these "microcredits" and keep up with the requirements of the state licensing authorities and medical boards.

As with any novel technology, one limitation is that data mining has yet to be validated for use, though it is being used for quality improvement in several hospitals.[3] Adopting a new technology may be difficult for some programs due to additional cost and the need for technical personnel to establish and support the system. The adaptability of this unique data-mining technology to populate already existing web-based systems may

contribute to improved efficiency for residents. While specialties like neurology do not mandate the use of case logs, the method used in this study can be implemented in procedure-based specialties in which maintaining logs is mandatory. More studies are needed with the tool tested on a larger scale and with data taken over a longer period of time prior to implementation. Studies would also be helpful for enhancing the accuracy of the machine learning process. Finally, resident surveys to gauge their perceptions of the benefit of using this system and insights that the information provides would be helpful.

Conclusion

Manual case logs are time-consuming, provide limited information, and are unpopular with residents. The new data-mining program is a time-effective tool for data collection and analysis, as demonstrated in this pilot study. In addition to meeting case log requirements, data-mining technology could be used by programs for program improvement, to document the ACGME core competency requirements, to aid in self-evaluation and evaluation of the program by residents, for research, and to provide rapid feedback to residents without affecting duty hours. Larger ER studies involving multiple residency programs are needed to develop this method further.

References

1 Gill DJ, Freeman WD, Thoresen P, Corboy JR. Residency training the neurology resident case log: a national survey of neurology residents. *Neurology.* 2007;68(21):E32–E33.

2 Nasca TJ, Day SH, Amis ES Jr. The new recommendations on duty hours from the ACGME Task Force. *N Engl J Med.* 2010;363(2):e3. http://www.nejm.org/doi/full/10.1056/NEJMsb1005800

3 Veluswamy R. Golden nuggets: clinical quality data mining in acute care. *Physician Exec.* 2008;34(3):48–53.

4 Peltier WL. Core competencies in neurology resident education: a review and tips for implementation. *Neurologist.* 2004;10(2):97–101.