Research Paper ■

Automated Tuberculosis Detection

GEORGE HRIPCSAK, MD, CHARLES A. KNIRSCH, MD, MPH, NILESH L. JAIN, DSc, ARIEL PABLOS-MENDEZ, MD, MPH

A b s t r a c t   **Objective:** To measure the accuracy of automated tuberculosis case detection.

Setting: An inner-city medical center.

Intervention: An electronic medical record and a clinical event monitor with a natural language processor were used to detect tuberculosis cases according to Centers for Disease Control criteria.

Measurement: Cases identified by the automated system were compared to the local health department's tuberculosis registry, and positive predictive value and sensitivity were calculated.

Results: The best automated rule was based on tuberculosis cultures; it had a sensitivity of .89 (95% CI .75–.96) and a positive predictive value of .96 (.89–.99). All other rules had a positive predictive value less than .20. A rule based on chest radiographs had a sensitivity of .41 (.26–.57) and a positive predictive value of .03 (.02–.05), and a rule that represented the overall Centers for Disease Control criteria had a sensitivity of .91 (.78–.97) and a positive predictive value of .15 (.12–.18). The culture-based rule was the most useful rule for automated case reporting to the health department, and the chest radiograph-based rule was the most useful rule for improving tuberculosis respiratory isolation compliance.

Conclusions: Automated tuberculosis case detection is feasible and useful, although the predictive value of most of the clinical rules was low. The usefulness of an individual rule depends on the context in which it is used. The major challenge facing automated detection is the availability and accuracy of electronic clinical data.

■ **J Am Med Inform Assoc.** 1997;4:376–381.

Tuberculosis detection is essential for initiating therapy, isolating patients, testing close contacts, and fol-

lowing the epidemiology of the disease. It has been argued that health information networks may improve the accuracy, timeliness, and confidentiality of case detection and reporting through automated surveillance, standardized definitions, real-time reporting, and security measures.[1–3] The need for new technologies for treatment, diagnosis, and prevention of tuberculosis has been noted.[4] We used an electronic medical record, natural language processing, a clinical event monitor, and a health information network to automate the detection of tuberculosis cases at Columbia–Presbyterian Medical Center, in order to improve case reporting to the New York City Department of Health and to improve tuberculosis isolation at the medical center.[5]

This paper describes a study of the accuracy of automated tuberculosis case detection. The positive predictive value and sensitivity of each of eight event monitor decision rules is reported, using the health department's existing tuberculosis registry as the reference standard. The appropriateness of individual

rules for case reporting and respiratory isolation is discussed.

Background

Case detection is one component of the "Applied Informatics" project.[6] The project is part of the U.S. Department of Commerce's Telecommunications and Information Infrastructure Assistance Program[7] to demonstrate the use of the national information infrastructure (NII) to coordinate health care. Three major health care providers that serve northern Manhattan —Columbia–Presbyterian Medical Center, New York City Department of Health, and Visiting Nurse Service of New York—teamed up to create the information infrastructure necessary to provide coordinated, effective care to patients in the home, clinic, doctor's office, and hospital. The primary focus area was tuberculosis—treatment is prolonged and crosses multiple organizations, providers, and locations,[8] making coordination critical—but the technology was applied to all patients in common among the organizations. The goals were to link the electronic patient records of the organizations to inform providers what other organizations are doing (for example, informing physicians of home care nurse observations); to improve adherence to accepted clinical protocols (respiratory isolation of tuberculosis patients); to improve and streamline reporting cases to the health department; to improve communications with mobile health care providers; to improve patient and provider access to educational materials; and to maintain appropriate patient privacy for all data transfer.

Independent of this project, the New York City Department of Health has mounted an intense effort to detect every case of tuberculosis in the city; a review revealed only 3 missed cases out of 3,000.[9] The effort includes clinicians who report cases based on clinical history, physical examinations, tuberculin skin testing, laboratory results, and chest radiographs; and public health workers who visit the city's many clinical laboratories looking for positive tuberculosis cultures. The cases are recorded in the health department's tuberculosis registry, which guides public health planning and treatment, and which is available over a wide-area network to the health department's chest clinics. Populating the registry is an enormous, resource-intensive process.

This setting provided an excellent standard with which to test the hypothesis that automated case detection could be made accurate and timely by exploiting the Applied Informatics infrastructure. Similar projects[10–13] have focused on electronic registries and sometimes electronic reporting, but not completely

Table 1 ■

Centers for Disease Control Definition of a Countable Case of Tuberculosis

Any of the Following Laboratory Criteria	Or All of the Following Criteria for a Clinical Case Definition
Isolation of *M. tuberculosis* from a clinical specimen	A positive tuberculin skin test
Demonstration of *M. tuberculosis* from a clinical specimen by DNA probe or mycolic acid pattern on high-pressure liquid chromatography	Other signs and symptoms compatible with tuberculosis, such as an abnormal, unstable (worsening or improving) chest x-ray or clinical evidence of current disease
Demonstration of acid-fast bacilli in clinical specimen when a culture has not been or cannot be obtained	Treatment with two or more antituberculosis medications
	Completed diagnostic evaluation

automated detection. In this study, case detection occurred without human intervention; it depended solely on the medical center's existing electronic medical record.[14]

Methods

The objective was to detect "countable" cases of tuberculosis, as defined in Table 1 from the Centers for Disease Control.[15] The automatically detected cases were compared to the health department's tuberculosis registry, which is based on these same criteria.

We used an automated decision support system called a clinical event monitor,[16] which tracks all the clinical data that are available electronically in the medical center. The monitor contains rules that are triggered by clinical events such as the storage of data or the admission of a patient. If, based on the electronic medical record, a clinically relevant situation is found, a message is generated and sent to the appropriate provider or department. The monitor's rules are based on a national standard called the Arden Syntax for Medical Logic Modules.[17]

Much of the data necessary to test the criteria in Table 1 were not available electronically. For example, the tuberculin skin test and details of the clinical course were not generally available in the electronic record. Therefore, it was not possible to automate the tuberculosis criteria exactly.

Table 2 ■

Automated Rules

R1. Positive *Mycobacterium* tuberculosis culture
R2. Positive acid-fast bacillus smear or any positive *Mycobacterium* culture
R3. Chest radiograph report whose description or impression section is either:
  (a) highly suspicious for tuberculosis (radiologist mentions tuberculosis or cavitary disease)
                    OR
  (b) moderately suspicious (various forms of infiltrate, granuloma, effusion, or pleurisy) with immunocompromise (implied by use of medications that treat complications of immunocompromise or by tests that demonstrate immunocompromise)
R4. Inpatient use of anti-tuberculosis medications
R5. R1 or R2 (approximates laboratory criteria)
R6. R3 and R4 (approximates clinical criteria)
R7. R1 or R2 or {R3 and R4} (approximates overall criteria)
R8. R1 or R2 or R3 or R4 (most sensitive rule)

Instead, we generated a series of rules (Table 2) that used whatever data was available, such as tuberculosis cultures, acid fast bacillus smears, other laboratory tests, chest radiographs, and medications. Rules R1 and R2 approximate the laboratory criteria; rules R3 and R4 cover what was available for the clinical criteria.

Chest radiograph data were available as electronic narrative reports (dictated by radiologists and typed by transcribers). Narrative data are not useful to most automated systems, because the systems cannot decipher the meaning of English phrases. For example, "possible infiltrates" (moderate probability that infiltrates are present) must be distinguished from "vague infiltrates" (definite presence of infiltrates that are vague). We therefore exploited a computer program called a natural language processor,[18] which takes the narrative data and turns it into unambiguous coded form. The processor had previously been shown to detect clinical conditions in chest radiographs in a manner that was not distinguishable from physicians but superior to lay persons and alternative computer programs,[19] and it showed reasonable accuracy (92% agreement with a clinician's opinion) in detecting chest radiographs suspicious for tuberculosis.[20]

The study period was July 1, 1995 to May 31, 1996. All patients seen at the medical center (inpatient and outpatient) were assessed. During the study period, the automatically detected cases were kept separate from the health department's tuberculosis registry, which contained only manually reported cases. The two sets of cases were matched by name and date of birth for analysis.

Accuracy was measured in terms of positive predictive value and sensitivity. The positive predictive

value of a rule was defined as the proportion of cases identified by the rule that actually had countable tuberculosis. Only those cases identified by the rule during the study period were included. A case was considered countable if it was recorded as such in the tuberculosis registry some time from 12 months before the study period to 2 months after the period; this allowed for delayed identification by either the automated system or workers recording cases in the tuberculosis registry. (For example, the automated system might detect a case before the proper paperwork was generated to report it to the health department's tuberculosis registry. Conversely, a culture-negative case might be reported to the health department before sufficient electronic evidence of tuberculosis was amassed to satisfy the automated rules.)

The sensitivity of a rule was defined as the proportion of countable cases that were identified by the rule. Countable cases included only those that were recorded in the tuberculosis registry during the study period and attributed to Columbia–Presbyterian Medical Center (the registry collects cases from the entire city). To allow for delayed identification, this set of countable cases was matched against cases identified by the automated system from when the system was first turned on (June 20, 1995) to 2 months after the period. Because the system was running only 11 days before the study period, the electronic medical records of countable cases that were missed by the rule were reviewed manually for data that might have triggered an identification up to 12 months before the study period. This latter information is noted in the results section but is not included in the reported sensitivity.

For those cases in common between the automated system and the tuberculosis registry, the average difference in reporting date was also calculated. Some countable cases in the tuberculosis registry are first

Table 3 ■

Positive Predictive Value

Rule	Total Cases Identified by Rule	Countable in TB Registry	Positive Predictive Value (95% CI)
R1. (MTB culture)	74	71	.96 (.89–.99)
R2. (AFB smear)	419	71	.17 (.13–.21)
R3. (CXR)	834	29	.03 (.02–.05)
R4. (anti-TB medications)	595	46	.08 (.06–.10)
R5. (laboratory criteria)	419	71	.17 (.13–.27)
R6. (clinical criteria)	125	24	.19 (.13–.27)
R7. (overall criteria)	481	72	.15 (.12–.18)
R8. (most sensitive rule)	1527	73	.05 (.04–.06)

reported as suspicious and then converted to countable. Therefore, the automated system was compared to both the earliest reporting date (whether suspicious or countable) and the actual date the case was reported as countable.

Exact binomial confidence intervals[21] were calculated for positive predictive value and sensitivity. Confidence intervals for difference in reporting date were approximated with the Student's t distribution.[22]

Results

Approximately 450,000 unique patients were assessed by the automated system during the study period. Table 3 shows the number of cases identified by each rule during this period, how many of those cases were marked as countable in the tuberculosis registry, and the corresponding positive predictive values. Rule R1, which was based on tuberculosis cultures, had the highest predictive value. Its three false positive cases were due to laboratory errors. (The false-positive laboratory-error rate was .0003, based on 3 errors in about 10,000 tuberculosis cultures during the 11-month period.) The other rules, which were based on less definitive data, had positive predictive values less than 0.20. No known countable cases were identified by the automated system but missed in the tuberculosis registry.

During the study period, 45 countable cases were attributed to Columbia–Presbyterian Medical Center in the tuberculosis registry. One of these cases was attributed incorrectly by the health department (the patient had been diagnosed and treated at another hospital, so there was no paper or electronic evidence of tuberculosis at the medical center); therefore, only 44 tuberculosis registry cases were used. (This is less than the 73 true positive patients detected by rule R8 because some medical center cases were attributed to other hospitals in the registry and because the automated system detected several prevalent cases early in the study period that had already been recorded in the tuberculosis registry before the study period.) Table 4 shows how many of the 44 cases were identified by each rule and the corresponding sensitivity.

Rule R1 detected 39 of the 44 cases. The other 5 cases were culture-negative tuberculosis, indicating a culture-negative tuberculosis rate of 11% in this sample. Two of the culture-negative cases were detected by the rules that approximate the clinical criteria (R3, R4), and two more would have been detected if those rules had been running at least four months before the study period. The last culture-negative case would have been detected if the rules exploited any of the following: referring information section of chest ra-

Table 4 ■

Sensitivity

Rule	Number of TB Registry Cases (out of 44) Identified by Rule	Sensitivity (95% CI)
R1. (MTB culture)	39	.89 (.75–.96)
R2. (AFB smear)	39	.89 (.75–.96)
R3. (CXR)	18	.41 (.26–.57)
R4. (anti-TB medications)	30	.68 (.52–.81)
R5. (laboratory criteria)	39	.89 (.75–.96)
R6. (clinical criteria)	16	.36 (.22–.52)
R7. (overall criteria)	40	.91 (.78–.97)
R9. (most sensitive rule)	41	.93 (.81–.99)

diograph reports (the ordering physician's description of the reason for the radiograph), financial codes, or pathology reports; the effect of including these data on positive predictive value was not measured.

On average, the automated system identified cases 21 (95%; CI 15–27) days earlier than they were reported in the tuberculosis registry as countable, and 10 (7–13) days earlier than they were reported in the tuberculosis registry as either suspicious or countable. Rule R1 identified cases 16 (10–22) days earlier than they were reported in the tuberculosis registry as countable, and 4 (1–7) days earlier than they were reported in the tuberculosis registry as either suspicious or countable.

Discussion

A wide range of performance was exhibited by the different rules. Rule R1 had the best predictive value and good sensitivity; this is not unexpected because tuberculosis cultures are pathognomonic for tuberculosis disease and culture-negative tuberculosis is infrequent. The other rules added little sensitivity with a large drop in predictive value. The usefulness of individual rules, however, varied greatly depending on whether they were applied to case reporting or tuberculosis respiratory isolation.

An effective tuberculosis registry must be accurate,[9] so only rule R1 is now used to report cases directly to the health department's tuberculosis registry. Cases detected by rule R1 are encoded in the Health Level Seven[23] clinical data messaging standard, encrypted, and transferred via a modem to the health department. There they are automatically inserted into the tuberculosis registry as suspicious cases. Even rule R1's accuracy is not sufficient, so official designation in the registry as "countable" requires human inter-

vention (for example, to rule out a laboratory error). Culture-negative cases are still identified by medical center epidemiologists and physicians, using rules R2, R3, and R4 as adjunct screening tools. The main impact of the system has been the availability of more timely microbiology data, which has facilitated the health department clinicians' work. The potential monetary value of automated reporting is about $2,000 per year per medical center, based on the costs to send a health department worker in person to the medical center's clinical laboratory one half day twice per month (the health department's current standard practice).

Rules R1, R2, and R3 were also used to detect patients with active tuberculosis who were not in respiratory isolation rooms.[5] Whenever an inpatient who was not assigned to an isolation room fulfilled these criteria, an alert was sent to the hospital epidemiologist, who assessed the case. In a 12-month period, the system identified 7 tuberculosis patients for isolation who would otherwise have been missed,[5] avoiding the infection of other patients and hospital personnel. Furthermore, infected patients received more rapid, accurate therapy because physicians who failed to isolate their patients frequently missed tuberculosis entirely or used inappropriate treatment. This is especially important in light of recent evidence that a delay in treatment is significantly correlated with mortality.[24]

Rule R3 detected four of the patients, R1 detected one, and R2 detected two. The positive predictive value and sensitivity of rule R3 was the least impressive of the three rules, yet it was the most effective. There are several reasons. Perhaps most important, a physician who ordered a tuberculosis test (which later triggered rule R1 or R2) probably suspected tuberculosis and was likely to isolate the patient appropriately. Rule R3, triggered only by chest radiographs, was able to fire even in cases when tuberculosis was never suspected, and this is where it had the greatest effect. In addition, culture-driven alerts (R1) were usually too late to make a difference in isolation, whereas chest radiograph-driven alerts (R3) occurred soon after admission in many patients. (On average, rule R3 did not fire significantly earlier than rule R1, but for those cases in which it did fire earlier, it made a difference in isolation.) The isolation study's positive predictive value (.07) and sensitivity (.33) for rule R3 agreed fairly well with this study; some difference can be attributed to the difference in populations (inpatients versus all patients).

Therefore, context is critical in judging the usefulness of rules. Positive predictive value and sensitivity were helpful as predictors of behavior, but taking them

alone would have underestimated the impact of rule R3 on tuberculosis isolation. This illustrates the divide between studying a system's function (positive predictive value and sensitivity) and studying its impact on health care[25]; unexpected issues that affect impact frequently arise during real use.

The automated system requires an electronic medical record. The rules in Table 2 exploit data collected from common ancillary departmental systems: laboratory results, radiology reports, pharmacy orders. Therefore, it should not be difficult to collect similar data in other centers; one would have to install a clinical event monitor and, if the chest radiograph rules are used, a natural language processor. The rules could be used as-is or modified to fit a particular patient population with locally determined positive predictive values and sensitivities.

The system could be improved with more accurate, more complete clinical data, which would support more accurate clinical rules and perhaps better ability to differentiate laboratory errors from positive cultures. Clinical history, tuberculin skin test results, and clinical course would be most valuable. Collecting these data would require providers to record their clinical notes on a computer, either in coded form or through natural language processing of typed or dictated notes.

Conclusion

Automated tuberculosis case detection is feasible. The rules attained reasonable sensitivity, but most had low predictive value. Nevertheless, even in its current form, the automated system facilitated both case reporting to the local health department and respiratory isolation in the hospital. More complete, accurate clinical data would be most useful to improve the system.

References ■

1. Friede A, Blum HL, McDonald M. Public health informatics: how information-age technology can strengthen public health. Ann Rev Pub Health. 1995;16:239–52.
2. Thacker SB, Stroup DF. Future directions for comprehensive public health surveillance and health information systems in the United States. Am J Epidemiol. 1994;140:383–97.
3. Kilbourne EM. Informatics in public health surveillance: current issues and future perspectives. MMWR Morbidity and Mortality Weekly Report. 1992;41Suppl:91–9.
4. Miller B, Castro KG. Sharpen available tools for tuberculosis control, but new tools needed for elimination. JAMA. 1996; 276:1916–7.

5. Knirsch CA, Jain NL, Pablos-Mendez A, Friedman C, Hripcsak G. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system. Infection Control and Hospital Epidemiology, in press.

6. Hripcsak G, Jain NL, Knirsch C, Frieden T, Stazesky RC, Fulmer T, Pablos-Mendez A. Applied Informatics: using the NII to coordinate health care. In: Humphreys BL (ed). Proc AMIA Spring Congress. 1996; June 5–8; Kansas City. Washington, DC: American Medical Informatics Association, 1996;103.

7. US Department of Commerce National Telecommunications and Information Administration. Telecommunications and Information Infrastructure Assistance Program: Notice of Solicitation of Grant Applications and Guidelines for Preparing Applications. Washington, DC: US Department of Commerce, 1996.

8. Brudney K, Dobkin J. Resurgent tuberculosis in New York City: human immunodeficiency virus, homelessness, and the decline of tuberculosis control programs. Am Rev Respir Dis. 1991;144:745–9.

9. Frieden TR, Fujiwara PI, Washko RM, Hamburg MA. Tuberculosis in New York City: turning the tide. N Engl J Med. 1995;333:229–33.

10. Serra T, Salema A, Lopes H, Antunes ML. Tuberculosis surveillance and evaluation system in Portugal. Tuber Lung Dis. 1992;73:345–8.

11. Chapman KA. Georgia information network for public health officials. In: Humphreys BL (ed). Proc AMIA Spring Congress. 1996; June 5–8; Kansas City. Washington, DC: American Medical Informatics Association, 1996;79.

12. Valleron A-J, Garnerin P, Menares J. The French National Communicable Diseases Network: an overview. In: Duisterhout JS, Hasman A, Salamon R (eds). Telematics in Medicine. Amsterdam: North-Holland, 1991;193–203.

13. Snodgrass I, Chew SK. A national computer-based surveillance system for tuberculosis notification in Singapore. Tuber Lung Dis. 1995;76:264–70.

14. Johnson SB, Forman B, Cimino JJ, Hripcsak G, Sengupta S, Sideli R, Clayton PD. A Technological Perspective on the Computer-Based Patient Record. In: Steen EB (ed). Proceedings of the First Annual Nicolas E. Davies CPR Recognition Symposium. 1995; April 4–6; Washington (DC). Washington, DC: Computer-based Patient Record Institute, 1995; 35–51.

15. Wharton M, Chorba TL, Vogt RL, Morse DL, Buehler JW. Case definitions for public health surveillance. MMWR Morbidity and Mortality Weekly Report. 1990;39(RR-13):1–43.

16. Hripcsak G, Clayton PD, Jenders RA, Cimino JJ, Johnson SB. Design of a clinical event monitor. Computers and Biomedical Research. 1996;29:194–221.

17. Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden Syntax. Computers and Biomedical Research. 1994;27:291–324.

18. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Natural Language Engineering. 1995;1:83–108.

19. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med. 1995;122:681–8.

20. Jain NL, Knirsch C, Friedman C, Hripcsak G. Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. J Am Med Inform Assoc. 1996;3:542–6 (supplement).

21. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934;26:404–13.

22. Snedecor GW, Cochran WG. Statistical Methods. Ames: Iowa State University Press, 1989.

23. HL7 Working Group. Health Level Seven, Version 2.1. HL7 Working Group, 1990.

24. Pablos-Mendez A, Sterling TR, Frieden TR. The relationship between delayed or incomplete treatment and all-cause mortality in patients with tuberculosis. JAMA. 1996;276:1223–8.

25. Friedman CP, Wyatt JC. Evaluation Methods in Medical Informatics. New York: Springer–Verlag, 1996.