Research Paper ■

Speech Recognition as a Transcription Aid: A Randomized Comparison With Standard Transcription

David N. Mohr, MD, David W. Turner, Gregory R. Pond, Joseph S. Kamath, Cathy B. De Vos, Paul C. Carpenter, MD

Abstract **Objective**. Speech recognition promises to reduce information entry costs for clinical information systems. It is most likely to be accepted across an organization if physicians can dictate without concerning themselves with real-time recognition and editing; assistants can then edit and process the computer-generated document. Our objective was to evaluate the use of speech-recognition technology in a randomized controlled trial using our institutional infrastructure.

Design. Clinical note dictations from physicians in two specialty divisions were randomized to either a standard transcription process or a speech-recognition process. Secretaries and transcriptionists also were assigned randomly to each of these processes.

Measurements. The duration of each dictation was measured. The amount of time spent processing a dictation to yield a finished document also was measured. Secretarial and transcriptionist productivity, defined as hours of secretary work per minute of dictation processed, was determined for speech recognition and standard transcription.

Results. Secretaries in the endocrinology division were 87.3% (confidence interval, 83.3%, 92.3%) as productive with the speech-recognition technology as implemented in this study as they were using standard transcription. Psychiatry transcriptionists and secretaries were similarly less productive. Author, secretary, and type of clinical note were significant ($p < 0.05$) predictors of productivity.

Conclusion. When implemented in an organization with an existing document-processing infrastructure (which included training and interfaces of the speech-recognition editor with the existing document entry application), speech recognition did not improve the productivity of secretaries or transcriptionists.

■ **J Am Med Inform Assoc.** 2003;10:85–93. DOI 10.1197/jamia.M1130.

Affiliations of the authors at the Mayo Clinic, Rochester, Minnesota: Division of Area General Internal Medicine (DNM); Section of Mayo Integrated Clinical Systems Coordinationa (DWT); Section of Biostatistics (GRP); Section of Information Servicesa (JSK, CBDV); Division of Endocrinology, Diabetes, Metabolism, Nutrition, and Internal Medicine (PCC).

The authors thank Ms Patricia K. Olevson and Ms Melissa L. Malley for their many contributions with language model building and pilot study support; Messrs Chris M. Grebin, Dean L. Sorum, and Richard W. Kingsley for their technical assistance with the preliminary study; Ms Michelle A. Rinn for her contributions with data collection; Jeffrey J. Huhn, MD, Ms Arleen M. Derynck, Ms Betsy A. Owen, and Ms Debra A. Ranfranz for their many contributions to the administration of this project; and Ms Jacklynn S. Conway and Ms Katherine T. Walters for their help with manuscript preparation.

Nothing in this publication implies that Mayo Foundation endorses any products mentioned.

Correspondence and reprints: David N. Mohr, MD, Division of Area General Internal Medicine, Mayo Clinic, 200 First Street SW, Rochester, MN 55905; e-mail: <dmohr@mayo.edu>.

Received for publication: 03/22/02; accepted for publication: 08/19/02.

Speech recognition is viewed as an important future option for electronic health record information entry. Although data input can be accomplished by template-based self-entry, the most common input process is dictation of the text and subsequent transcription. The magnitude of the resources needed to transcribe clinical notes is a substantial barrier to the move from a handwritten to an electronic medical record.

Speech recognition is a new technology that promises to reduce the costs of an electronic medical record system. Almost all of the evaluations of speech-recognition technology in medicine have focused on continuous recognition systems in which the health care provider acts as both the author and the sole editor.[1–4] These studies evaluated error rates but did not determine whether the editing process impaired provider productivity. When provider productivity was studied, dictation plus provider editing time with speech recognition was longer than with standard transcription.[5] In a preliminary study conducted 12 months before the current study, we also found that physician productivity was reduced when recognized speech was shown to the physician to edit during the dictation session (D. N. Mohr, unpublished data, 2001).

For our organization, the most popular use of speech recognition was standard dictation into a digital dictation system interfaced with a speech-recognition engine, which then presented the resulting document to a secretary or transcriptionist who edited the document and saved it (D. N. Mohr, unpublished data, 2001). Preliminary evidence indicated that this approach improved secretarial productivity by 24%. This does not change the physician dictation process or decrease physician productivity. Additionally, some transcription companies use the same model.

Secretaries in the preliminary study were volunteers and may not have been representative of all secretaries. Similarly, the five participating physicians had not been selected randomly. Additionally, one of the physicians dictated in a way that made speech recognition extremely inefficient; his documents were not included in the study. Finally, the preliminary study did not allow us to evaluate the interactions between author and transcriptionist.

Therefore, we undertook a randomized, controlled study that included all staff physicians, secretaries, and transcriptionists in two specialty areas. Both secretaries and physicians were randomly assigned to a document-processing approach to determine which was more productive. Various subsets of authors, transcriptionists, and document types were predefined for analysis to determine whether certain subgroups or combinations would better lend themselves to speech-recognition technology.

Methods

Standard Transcription Process

At our institution, most clinical notes are dictated by the provider. A campus-wide, telephone-based, digital dictation system supports this dictation. Each author follows a generic template that includes sections for chief complaint, history of present illness, review of systems, examination, and impression. As they dictate, providers categorize sections of the note. These sections improve document clarity and subsequent navigational ease. There are also specific note types for consultations, general examinations (or histories and physicals), limited examinations, return visits, and miscellaneous notes. Using the telephone keypad, the dictating physician indexes his or her dictation job to indicate physician, patient number, and the dictation job work type, which correlates with a note type. Voice files are stored digitally on central dictation file servers. A work-waiting application follows the status of dictation jobs on all dictation servers. Dictation jobs are routed to appropriate medical secretaries (who transcribe on a part-time basis) or transcriptionists (who transcribe full-time). Medical secretaries and transcriptionists call the desired voice files to their workstation and transcribe the dictation into a clinical note using the Mayo-developed Clinical Notes application. This application supports the generic template described above. Notes are stored on a central server and are viewed on a Mayo-developed document browser. This system can be used at more than 12,000 workstations.

Server-based Speech-recognition Process: Using Speech Recognition as a Secretarial Aid

The provider dictates, using the system described above. However, the voice file is sent to a server-based speech-recognition system instead of a secretary or transcriptionist. This system converts voice text using a generic model for all authors within a medical specialty or a voice and language model assigned to a specific author. The document produced and its associated voice file are then routed to a secretary or transcriptionist to edit in the transcription editor. (All secretaries and transcriptionists received 6 hours of training over 3 consecutive days; after training, they had at least 4 weeks of practice before the start of the study.) After it is edit-

ed, the file is converted to a note in the Clinical Notes application. To do this, an interface application uses tags in the dictation to create sections in the editor and convert these dictated sections to Clinical Note sections (e.g., chief complaint, history of present illness). This interface application process takes less than 10 seconds if there are no errors in the tags (e.g., duplicate or incorrect tags, invalid numeric values). The note is proofread in the Clinical Notes application to ensure that the appropriate text has been put into the appropriate sections of the note. The note is then saved and the job is marked transcribed.

Model Building and Testing

Domain (medical specialty) acoustic and language models were built on site using LTI model-building software provided by the vendor (Linguistics Technology, Inc, Edina, MN). To create acoustic models, 40–80 megabytes of voice files for each author were exported from the digital dictation system to another server and stored in folders specific to each author and specialty. To create language models, corresponding text files were exported from the clinical notes repository and stored in a folder specific to the specialty. Five medium-sized voice files (5–8 minutes in length) and the associated text were randomly extracted and placed into a separate folder specific to each author. These latter files were used to determine which combination of acoustic and language models was best for each author.

A speaker-independent language model containing a list of the most commonly used words was created from the text files for each medical specialty. The language model was then reviewed to ensure that it contained normalized word forms, accurate spelling, proper conversion of numerical values (e.g., "one twenty over eighty" changed to "120/80"), proper conversion of punctuation, accurate drug names, and proper handling of duplicate word formats (e.g., "Mr. Cook" and "a cook"). Because some authors often use words not contained in their medical specialty language model, a speaker-adapted language model was created for each author.

For each medical specialty, a speaker-independent acoustic model containing the phonetic sounds for commonly used words was created from the voice files. Speaker-adapted acoustic models also were created.

The final task in the model-building process was to select for each author the specific language model and acoustic model that provided the best speech recognition. This involved testing four models (speaker-independent acoustic model, speaker-independent language model, speaker-adapted acoustic model, speaker-adapted language model) for each author and selecting the combination of language and acoustic models that would be best for that author.

Document Processing and Randomization

All staff physicians, secretaries, and transcriptionists in the Division of Endocrinology, Diabetes, Metabolism, Nutrition, and Internal Medicine (hereafter called endocrinology) and the Department of Psychiatry and Psychology (hereafter called psychiatry) were included unless they were scheduled to be away from work for most of the 4-week study. Physicians must have been on the staff for 18 months so that an adequate number of previous notes were available for the building of language models. From a list of random numbers, staff physicians were assigned to one of two groups. The physicians were experienced in using the telephone dictation system and received no new training in dictation. They were unaware of their group assignment and therefore did not know how their dictation would be processed. Initially, dictation jobs from the first group were assigned to standard transcription and jobs from the second group to speech recognition. After 2 weeks, the document-processing assignments were reversed. Secretaries and transcriptionists also were assigned randomly to two groups. The first group used speech recognition, and the second used digital dictation. These assignments were changed every other working day except for the seven secretaries in psychiatry, who, for purposes of confidentiality, remained linked to individual physicians. Secretaries in psychiatry changed their documentation-processing mode at 2 weeks instead of every other day. Secretaries who processed part-time were studied separately from transcriptionists who did this activity full-time.

Participant Exclusions

Physicians, secretaries, and transcriptionists were excluded if they did not complete at least one dictation or transcription in each of the two document-processing approaches. No physicians were excluded because of dictation behavior, but a subgroup analysis was performed for those whom the secretaries believed dictated inefficiently.

Dictation Job Types

To reduce productivity variations resulting from the use of a wide variety of dictation job types, the only

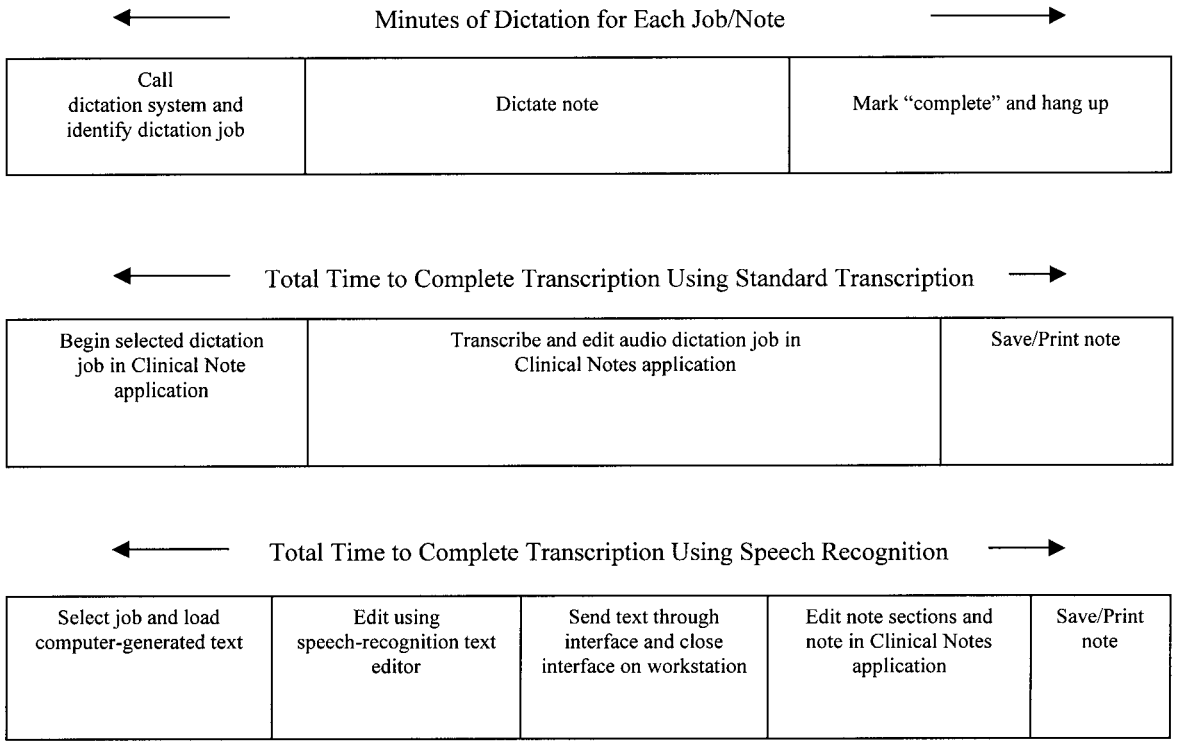

Figure 1. Document-processing steps in standard transcription and transcription using speech recognition.

job types used were the following: multisystem history and physical examination (work type 1 in both endo-crinology and psychiatry), limited examination (work type 2 in endocrinology and work type 1 in psychiatry), consultation (work type 3 in endocrinology and work type 1 in psychiatry), return office visit (work type 4 in both endocrinology and psychiatry), and supervisory notes (work type 14 in both endocrinology and psychiatry).

Document Processing

To compare server-based speech recognition with standard transcription for a dictated clinical note, rules were developed to create a consistent work flow that would allow unbiased comparisons. These best practices included being certain that the Clinical Notes application was launched before the speech-recognition editor was used. This avoided including the launch time in the secretarial or transcriptionist work time. Other best practices included avoiding breaks after a job had been selected; avoiding listening to the job before starting work in Clinical Notes or the editor; saving printing jobs that required unusual work, such as calling the author; and avoiding starting documents that could not be completed in 1 session.

Data Generation

From various time stamps, the durations and ratios for the different document-processing steps were determined (Figure 1).

1. **Total time to complete transcription using standard transcription**. This duration started when the secretary or transcriptionist began to transcribe the dictated job and stopped when the job was fully transcribed, saved, and printed. The total time included all editing done within the Clinical Notes application. This measurement was performed automatically through time stamps in the Clinical Notes application (i.e., time from the selection of "New Note" to the selection of "Save&Print").

2. **Total time to complete transcription using speech recognition**. This duration started when the voice file and computer-generated text were loaded into the transcription editor where the note with section tags was edited. The first step of this duration ended when the note was passed to the interface application. This second step ended when the properly tagged document was saved. The third step of the duration started when the file was imported to the Clinical Notes application and

ended when a final edit was completed and the note was printed.

3. **Minutes of dictation for each note**. This duration started at the beginning of the telephone call into the dictation system and ended when the author hung up. It did not include any time that the dictation system was put into pause mode by the author.

4. **Hours of secretarial work**. The number of hours of secretarial work (hsw) was determined for each day for each secretary by calculating the total time to complete transcription per secretary. This measured all time spent typing, editing, and proofreading dictated documents and speech-recognized documents. (Category includes transcriptionists.)

5. **Secretarial productivity**. Secretarial productivity was the ratio of minutes of dictation time (mdt) to hsw. (Category includes transcriptionists.)

Dictation Job Exclusion Rules

A job was defined as a unit of dictation related to a note. Dictation jobs were excluded if (1) a note was worked on several times by the secretary or transcriptionist, (2) an author used more than one dictation session for a given note, (3) there was a pause in keyboard activity greater than the length of the dictation, (4) the dictation time for the job was zero minutes, and (5) record storage identification criteria between the dictation database and the Clinical Notes database did not correlate.

Statistical Analysis

Data on each group (endocrinology secretaries, psychiatry secretaries, and psychiatry transcriptionists) were analyzed separately. Descriptive statistics, such as the mean, sum, and interquartile range, were calculated for each measurement within each group. The total mdt of all jobs was divided by the total hsw for each group and used to estimate productivity. The estimate of productivity for standard transcription was used as the current productivity level for each group. Regression analyses with a logarithmic transformation were performed to identify significant predictors of productivity (e.g., week of the study, transcriptionist, author, job type) per dictation job. Interactions between potential predictors also were investigated. Statistical significance was defined as a p value less than 0.05; all tests were two-sided.

The total productivity estimate using speech recognition was subtracted from the total productivity estimate using standard transcription to obtain an estimate of productivity difference. Five hundred sample productivity difference estimates were calculated by bootstrapping on the dictation job, from which 95% confidence intervals (CI) were obtained using percentile methods. Productivity gains or losses using speech recognition were expressed as percentages of current standard transcription productivity. This percentage was computed overall, for significant predictors identified using regression analyses, and for potential predictors identified retrospectively through discussions with secretaries and transcriptionists involved in the study. Interactions between potential predictors also were investigated.

Results

Overall Descriptive Statistics

Endocrinology

Thirty-nine endocrinology authors and 18 secretaries completed 2,878 dictation jobs. Nine authors with their associated 328 jobs and no secretaries were excluded because they did not complete at least one job in both standard transcription and speech recognition. On the basis of dictation job exclusion rules, 196 additional jobs were excluded. Hence, for purposes of analysis, 30 authors and 18 secretaries completed 2,354 jobs.

Psychiatry Full-time Transcriptionists

Of the 47 psychiatry authors, 45 sent 386 dictation jobs to the 6 full-time transcriptionists (who transcribed only work type 1 jobs). Four authors and their associated 13 jobs were excluded because they did not complete at least one job in both standard transcription and speech recognition. On the basis of dictation job exclusion rules, 30 additional jobs were excluded. Hence, analysis was done for 41 authors, 6 transcriptionists, and 343 jobs.

Psychiatry Secretaries

Thirty-six psychiatry authors and 7 secretaries completed 405 dictation jobs. Two authors with their associated 11 jobs and no secretaries were excluded because they did not complete at least one job in both standard transcription and speech recognition. Thirty-nine jobs were excluded on the basis of dictation job exclusion rules. Analysis therefore included 34 authors, 7 secretaries (who transcribed work type 4 and 14 jobs for their particular physician only), and 355 jobs.

Standard Transcription

Endocrinology

A total of 1,053 jobs were completed. There were 134 work type 1 jobs, 138 work type 2 jobs, 328 work type 3 jobs, 210 work type 4 jobs, and 243 work type 14 jobs. The average job required 4.6 minutes (median: 4.0 minutes) to dictate and 20.8 minutes (median: 16.0 minutes) to transcribe. Secretaries required 21,914.5 minutes to complete 4,874.0 minutes of author dictation. Thus, productivity was 13.34 mdt/hsw.

Psychiatry Full-time Transcriptionists

A total of 123 jobs were completed. The average job required 7.2 minutes (median: 7.0 minutes) to dictate and 30.9 minutes (median: 24.6 minutes) to transcribe. Transcriptionists required 3,796.6 minutes to complete 889.1 minutes of author dictation. Productivity was 14.05 mdt/hsw.

Psychiatry Secretaries

A total of 111 jobs were completed. The average job required 5.3 minutes (median, 4.6 minutes) to dictate and 25.2 minutes (median, 17.0 minutes) to transcribe. Secretaries required 2,794.4 minutes to complete 590.3 minutes of author dictation. Productivity was 12.67 mdt/hsw.

Speech Recognition

Endocrinology

A total of 1,301 jobs were completed. There were 169 work type 1 jobs, 142 work type 2 jobs, 362 work type 3 jobs, 326 work type 4 jobs, and 302 work type 14 jobs. The average job required 4.1 minutes (median: 3.4 minutes) to dictate and 19.6 minutes (median: 15.5 minutes) to transcribe using computer-based speech recognition. A total of 25,447.7 minutes of secretarial time was needed to finish 5,333.9 minutes of author dictation, a productivity of 12.58 mdt/hsw.

Psychiatry Full-time Transcriptionists

A total of 220 jobs were completed. The average job required 9.3 minutes (median: 8.5 minutes) to dictate and 55.9 minutes (median: 48.1 minutes) to transcribe. A total of 12,296.2 minutes of transcriptionist time was needed to finish 2,042.7 minutes of author dictation, a productivity of 9.97 mdt/hsw.

Psychiatry Secretaries

A total of 244 jobs were completed. The average job required 4.9 minutes (median: 4.3 minutes) to dictate and 31.6 minutes (median: 24.7 minutes) to transcribe. A total of 7,716.9 minutes of secretarial time was needed to finish 1,200.5 minutes of author dictation, a productivity of 9.33 mdt/hsw.

Regression Analyses

Endocrinology

Each of the following was a significant predictor of productivity: author, secretary, document-processing approach, and work type ($p < 0.001$ for each). Week of study ($p = 0.89$) was not significant. Second-order interactions also were investigated, and only work type by author interaction ($p = 0.22$) was not a significant predictor of productivity.

Psychiatry Full-time Transcriptionists

Work type could not be investigated in this portion of the study because all transcriptionists transcribed only work type 1 jobs. Transcriptionist, author, document-processing approach, and the second-order interaction between transcriptionist and document-processing approach were significant predictors of productivity ($p < 0.001$ for each). Week of study ($p = 0.41$), author by document-processing approach interaction ($p = 0.26$), and author by transcriptionist interaction ($p = 0.59$) were not significant predictors of productivity.

Psychiatry Secretaries

Work type was not investigated. Secretary, author, and document-processing approach were significant predictors of productivity ($p < 0.001$ for each). Author by document-processing approach interaction ($p = 0.076$) and secretary by document-processing approach interaction ($p = 0.086$) neared significance; author by secretary interaction ($p = 0.96$) was not a significant predictor of productivity.

Comparison of Speech Recognition with Standard Transcription

Table 1 shows productivity for speech recognition in terms of that for standard transcription.

Endocrinology

Productivity with standard description was 13.34 mdt/hsw. By means of bootstrapping, it was estimated that productivity with standard transcription would exceed that with speech recognition by 1.70 (CI, 1.03, 2.23) mdt/hsw. Thus, endocrinology secretaries were 87.3% (CI, 83.3%, 92.3%) as productive

Table 1 ■

Productivity for Speech Recognition Versus Standard Transcription in Three Groups and Selected Subsets

	Endocrinology secretaries			Psychiatry transcriptionists			Psychiatry secretaries		
Variable	n_1	n_2	Productivity (95% CI), %*	n_1	n_2	Productivity (95% CI), %*	n_1	n_2	Productivity (95% CI), %*
All dictation jobs	1,301	1,053	87.3 (83.3, 92.3)	220	123	63.3 (54.0, 74.0)	244	111	55.8 (44.6, 68.0)
Work type									
1	169	134	84.2 (75.0, 94.8)	NA	NA	NA	NA	NA	NA
2	142	138	89.2 (79.8, 99.1)	NA	NA	NA	NA	NA	NA
3	362	328	102.5 (93.5, 115.6)	NA	NA	NA	NA	NA	NA
4	326	210	82.5 (73.3, 92.4)	NA	NA	NA	NA	NA	NA
14	302	243	73.2 (65.4, 80.5)	NA	NA	NA	NA	NA	NA
Slow secretaries or transcriptionists	645	522	104.6 (100.0, 109.4)	138	63	93.6 (85.4, 101.0)	114	27	88.6 (73.6, 102.3)
Selected authors only	981	796	86.6 (81.7, 93.0)	172	93	61.6 (50.9, 73.6)	200	88	52.6 (39.3, 65.0)
Jobs > 3 min in dictation length	706	625	97.1 (90.1, 105.1)	190	106	52.8 (41.7, 64.1)	178	86	60.5 (46.6, 74.7)

CI, confidence interval; n_1, number of dictated jobs using speech recognition; n_2, number of dictated jobs using standard transcription; NA, not applicable.
*Productivity for speech recognition as a percentage of productivity for standard transcription.

using speech recognition as they were using standard transcription ([(13.34–1.7) ÷ 13.34] × 100).

Psychiatry Full-time Transcriptionists

Productivity with standard transcription was 14.05 mdt/hsw. By means of bootstrapping, it was estimated that productivity with standard transcription would exceed that of speech recognition by 5.16 (CI, 3.66, 6.46) mdt/hsw. Thus, psychiatry transcriptionists were 63.3% (CI, 54.0%, 74.0%) as productive using speech recognition as they were using standard transcription.

Psychiatry Secretaries

Productivity with standard transcription was 12.67 mdt/hsw. By means of bootstrapping, it was estimated that productivity with standard transcription would exceed that of speech recognition by 5.60 (CI, 4.05, 7.02) mdt/hsw. Thus, psychiatry secretaries were 55.8% (CI, 44.6%, 68.0%) as productive using speech recognition as they were using standard transcription.

Subset Analyses

Removal of Selected Authors

Seven endocrinology authors and 7 psychiatry authors had heavily accented speech or a conversational style of dictation. In this subset analysis, these authors were excluded. However, the resulting productivity (81.6%) was not significantly different ($p = 0.84$) from that of the overall group (see Table 1).

Removal of Selected Secretaries and Transcriptionists

Secretaries and transcriptionists who appeared slow (< 15 mdt/hsw [arbitrarily chosen cutoff point] in standard transcription) were identified, and an analysis based on data from these individuals only was conducted. Ten endocrinology secretaries, 4 psychiatry transcriptionists, and 3 psychiatry secretaries were included in this category. For this subset, productivity was better with use of speech recognition than with use of standard transcription (see Table 1). Further analysis showed that this subset's productivity with speech recognition was no different from that for the larger group and that the improved productivity with speech recognition simply reflected their decreased productivity with standard transcription.

Work Type

Work type could be analyzed only in endocrinology because the psychiatry secretaries and transcriptionists were restricted to particular work types. Use of speech recognition was more productive than use of standard transcription only for work type 3 jobs, although the difference was not statistically significant ($p = 0.91$). In comparison with standard transcription, use of speech recognition was least produc-

tive for work type 14 jobs, followed by work type 4 jobs. It was also noted that the psychiatry secretaries, who transcribe only work type 4 and 14 jobs, had much poorer productivity using speech recognition than did the endocrinology secretaries. With standard transcription, there was little variation in productivity between work types, with slightly better productivity for work type 1 jobs and slightly worse productivity for work type 14 jobs. With speech recognition, however, productivity for work type 14 jobs was worse than that for other work types, and productivity for work type 3 jobs was better.

Dictation Length

Upon completion of the study, the endocrinology secretaries involved in the preliminary study expressed their belief that speech recognition was not as productive for shorter jobs as for longer jobs. Examining mdt/hsw against job duration for speech recognition, it was found that, for endocrinology jobs with a dictation time longer than 3 minutes, productivity was significantly ($p = 0.021$) better than that for all job lengths combined. However, among psychiatry transcriptionists, productivity with long notes using speech recognition compared with standard transcription decreased.

Discussion

The introduction of new technology into a complex existing process is challenging. The benefit is sometimes as dependent on the process as it is on the technology that supports it. Speech recognition is a new technology that promises to reduce the costs of creating documents for the medical record. In a preliminary study of a small sample of physicians and secretaries, we found encouraging evidence that speech recognition would improve the efficiency of our documentation process. We sought to verify this preliminary finding by including all members of two medical specialties in a randomized, controlled study. We found that, overall, secretaries and transcriptionists using speech recognition could complete less dictation than transcriptionists using standard transcription. In endocrinology, secretaries were only 87.3% as productive using speech recognition compared with standard transcription; the psychiatry transcriptionists were 63.3% as productive, and the psychiatry secretaries were 55.8% as productive. In all three groups, however, the author and secretary or transcriptionist had a significant influence on productivity; therefore, technology was not the only variable that needed to be considered.

Although overall statistics showed no productivity benefit for speech recognition, a reasonable variable for subset analysis was the dictating author. If it were possible to identify a subset for whom this technology was helpful, significant savings could still be attained. Unfortunately, despite the fact that productivity with use of speech recognition depended on the author, the secretaries were unable to identify a subset of authors for whom the technology could improve productivity. This remains an area for further investigation.

A more successful approach was to identify a subset of secretaries for whom speech recognition would improve productivity. When secretaries who processed only 15 mdt/hsw were analyzed, productivity with speech recognition improved in psychiatry and surpassed that for standard transcription in endocrinology. Further investigation may clarify whether this subset should be targeted for use of this technology. It is unlikely that the small improvement of 4.6% in productivity seen in our endocrinology secretaries would justify the expense of the speech-recognition hardware and software support in our environment.

Another way to improve the use of speech recognition could be to restrict it to certain types of clinical notes. We found that speech recognition appeared to improve productivity for endocrinology consultation notes (work type 3) compared with standard transcription, but this was not statistically significant. We also found that, with use of speech-recognition, the productivity for supervisory notes (work type 14) was significantly lower than that for other work types. Because these were short notes, we evaluated whether dictation length could be used to determine which documents should be processed by speech recognition. In endocrinology, the productivity of secretaries using speech recognition for dictation jobs longer than 3 minutes improved but was not better than productivity with the use of standard transcription. In psychiatry, there was no improvement.

After analysis of our results, we sought to identify aspects of the implementation and use of speech recognition in our setting that could be changed, leading to improved productivity in the future. First, we considered training time. In our preliminary study and in the current study, we found no significant change in secretarial productivity over time. We assumed this meant that the 6 hours of training and 4 weeks of practice before the study resulted in peak competence at the study onset. However, because the five endocrinology secretaries in the preliminary study had used speech recognition intermittently dur-

ing the 12-month interval between the preliminary study and the current study, we compared their current speech-recognition productivity with that of the other endocrinology secretaries in the current study. The productivity of these five secretaries was significantly ($p < 0.001$) better than that of the others but was still slightly worse ($p = 0.28$) than that with standard transcription. Therefore, a longer training and practice period with the editor of the speech-recognition product may improve its chance of success.

We next evaluated the effect of speech-recognition accuracy on productivity. According to the Voice Recognition Accuracy Standard (National Institute of Standards and Technology formula), accuracy is calculated as 100 minus the word error rate. The word error rate is defined as the number of incorrect word substitutions, deletions, and insertions multiplied by 100 and divided by the true number of words. For each endocrinology author, one dictation job was randomly selected. The average recognition accuracy was 84.5% (range: 55–95%). The Pearson correlation coefficients for recognition accuracy and productivity were 0/29 ($p = 0.16$) for the endocrinology secretaries and 0.06 ($p = 0.43$) for the psychiatry transcriptionists. Thus there was no association between speech recognition accuracy and productivity in our study and, on the basis of our statistical analysis, we could not count on significant improvement from a better tool. Although we did not calculate recognition accuracies in the preliminary study, we could compare the productivity results of five authors and five secretaries who were in both studies. Between the preliminary study and the current study, the average secretarial productivity decreased for four of the five authors. Because we are aware of no process changes that occurred between the two studies and there was one change in the speech-recognition software, we assume that changes in the speech-recognition tools may have accounted for some of the worsening but that the remainder came from selection bias in the preliminary study.

Finally, we considered whether work flow accounted for the lack of productivity benefit found with speech recognition. Secretaries and transcriptionists used a special editor to move quickly through a speech-recognized document to correct errors. They then had to enter this text into a structured note. This was done using an interface that put sections into the correct locations in our Clinical Notes application. When text was transferred into the appropriate note sections without error, the process was fast; however, on occasion text would not be put into the appropriate section and a subsequent editing step was required. For speech recognition, 80.5% of the time to complete a note was spent in the speech-recognition editor, 1% in the interface tool that transferred edited text to the proper section, and 18.5% in the Clinical Notes application, where documents were reviewed, reedited, and saved. In addition, a document containing instructions such as "Paste my previous social history" required work in the Clinical Notes application and additional edits. Thus, the burden of structured notes and the inclusion of instructions in the dictation job undermined productivity. We plan to investigate the possibility that avoiding the use of instructions in the dictation job and having the secretary edit in only one application will improve productivity.

Conclusions

In summary, we could not find an overall benefit in the use of speech recognition in our clinical documentation process. We did find that secretarial and transcriptionist productivity was better in certain subsets. It was better for secretaries who were slower and, perhaps, for longer dictation jobs. Although there were significant differences in secretarial productivity between authors, we did not find a way to identify those authors with whom the technology could be used. It is possible that physician dictation training would increase productivity. Many users of speech-recognition software find that it is necessary to modify their style to get good recognition. This happens naturally when the clinician interacts directly with speech-recognition software on a workstation, but when the software is used as a secretarial aid there is no incentive and no feedback for clinicians to modify their style. It is also possible that longer training and a single editor work flow would allow secretaries to use the technology more productively. Because we did not find overall benefit, we did not perform a cost-benefit analysis to establish the business case for introducing speech-recognition technology.

References ■

1. Zafar A, Overhage JM, McDonald CJ. Continuous speech recognition for clinicians. J Am Med Inform Assoc 1999;6:195–204.
2. Devine EG, Gaehde SA, Curtis AC. Comparative evaluation of three continuous speech recognition software packages in the generation of medical reports. J Am Med Inform Assoc 2000;7:462–468.
3. Rosenthal DF, Bos JM, Sokolowski RA, et al. A voice-enabled, structured medical reporting system. J Am Med Inform Assoc 1997;4:436–441.
4. Lowes R. Bits and bytes. Med Econ 2000;77:35.
5. Borowitz SM. Computer-based speech recognition as an alternative to medical transcription. J Am Med Inform Assoc 2001;8:101–102.