

Letter to the Editor Re: Voice Recognition Dictation: Radiologist as Transcriptionist and Improvement of Report Workflow and Productivity Using Speech Recognition—A Follow-up Study

James Y. Chen,^{1,2} Paras Lakhani,^{4,5} Nabile M. Safdar,² and Paul G. Nagy³

We were surprised to see without comment in the December issue of JDI two hypothesis-driven research articles with opposite conclusions about the impact of speech recognition on radiologists. Pezzullo et al.¹ strongly favored the use of traditional transcription over speech recognition, whereas Koivikko et al.² determined speech recognition to be superior. With increasing imaging volume each year, identifying the costs and the benefits of either transcription-based or speech recognition workflow is an important contribution. As an interested reader, how can we draw useful conclusions from these two papers and adapt our practices?

One approach to reconcile these seemingly contradictory articles is to analyze their applicability and methodologies. Do limitations or assumptions account for their discrepant conclusions? Pezzullo et al. evaluated an outpatient imaging center in Rhode Island that collaborated with a for-profit imaging corporation and an academic radiology department. Koivikko et al. evaluated a large academic trauma hospital in Scandinavia. The requirements of a private outpatient practice and trauma hospital can be vastly different.

Both studies used report turn-around time (RTT) as a primary metric for comparing speech recognition and transcription. Is this a valid metric? An outpatient center should have a much lower percentage of findings that require urgent reporting. An acute care center requires rapid RTT for examinations that are critical to immediate patient management. Although an outpatient center requires RTT that responds to the needs of current referring physicians and attracts new referrals, the level of urgency is clearly different.

Moreover, research study conditions may not have reflected actual practice realities. In daily usage, speech recognition users often streamline their workflow by using report templates or macros. Transcription users often streamline their reporting by the use of “canned” reports that the radiologist instructs the transcriptionist to modify. Pezzullo et al. did not use templates in their study, suggesting that spine imaging is not conducive to the use of templates. Although it is true that segmental spine analysis is less conducive to templating, techniques and basic findings are templatable in most reports exam indications. For some examinations, the contents of an entire report may be templated. Thus, in a study design excluding the use of templates, the recorded RTT and effort by radiologists are artificially elevated and potentially biased against speech recognition. On the other hand, the transcription system used by Koivikko et al. was based on microcassettes,

¹From the Division of Neuroradiology, University of Pennsylvania, Philadelphia, PA, USA.

²From the Department of Radiology, University of Maryland, College Park, MD, USA.

³From the Department of Radiology; Quality and Informatics Research, University of Maryland, College Park, MD, USA.

⁴From the Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA.

⁵From the University of Maryland, College Park, MD, USA.

Correspondence to: James Y. Chen, Division of Neuroradiology, University of Pennsylvania, Philadelphia, PA, USA; e-mail: james_chen@alumni.brown.edu

Copyright © 2009 by Society for Imaging Informatics in Medicine

Online publication 22 April 2009

doi: 10.1007/s10278-009-9197-5

which must be collected, transported, and rewound, whereas more recent transcription systems may be entirely telephonic or electronic, which would eliminate the manual steps before transcription can begin. It is likely that their study results included increased RTT for transcription compared with what might be expected in many other transcription-based practices. It is therefore difficult to conclude whether others might achieve similar time savings.

The difference in sample sizes and exclusions between the two studies was also so great as to render comparison of their results less useful. One study included only 200 non-urgent cervical and lumbar spine examinations, whereas the other included more than 20,000 radiologic examinations of all types. Because reports may vary widely in complexity and length, a large sample size of varying report and exam types is essential. An ideal study would include large numbers of examinations of all types among multiple institutions to improve its power and generalizability.

One strength of the Pezzullo et al. study was the cost comparison between speech recognition and transcription. In any busy practice environment, report creation time is an important factor. An added transcription time of 2 min per study using speech recognition resulted in 104.5 additional minutes of editing and proofreading reports *per radiologist per day*. When the salaries of radiologists and transcriptionists, reported at \$175/hour and \$16/hour, respectively, were taken into account, the authors calculated approximately \$76,000 in annual savings using conventional transcription, excluding the expected additional financial advantages accrued if radiologists interpreted more studies with the time saved or if fewer radiologists were required.

A major weakness, reducing the applicability and significance of both papers, is that neither addressed limitations in their respective study designs. In addition to the factors we have described, numerous other possible confounding factors may have affected results in both studies, including but not limited to: modalities; speech recognition engines; transcription type; transcription staff size; radiologist age, gender, and experience; and over-read of preliminary results.

Elements in each of these studies are interesting and relevant, but specific flaws (shared by most

studies of workflow) limit their generalizability and usefulness. Inclusion of these studies with contradictory results in a single issue of JDI provided readers the opportunity to identify some of these flaws as seeds for thought and discussion in improving analyses of the relative merits of speech recognition and transcription. In the end, the choice between speech recognition and transcription will depend on some combination of the numerous factors addressed by the two papers, tailored to the specific needs and preferences of each practice.

It should be noted that the application of accepted scientific principles in informatics investigations can be particularly challenging. Randomized, controlled, double-blinded prospective studies may not be feasible in evaluating the real-time effects of information system implementation in the clinical setting. How many departments, for example, would endorse a study design in which PACS is installed for a random half of the department while the remaining half maintained a film-based workflow—possibly to the detriment of timely interpretation and patient care? Because of these practical constraints, informatics investigation designs tend to fall into the category of “quasi-experimental” studies.³ Technology advances relentlessly, and we must study industry innovations within the constraints of enterprise operation and patient care.

The challenge for all of us, especially in a rapidly changing health care environment with looming threats of spending cuts, is to reliably identify cost-effective technologies that truly improve our ability to deliver effective and efficient patient care.

REFERENCES

1. Pezzullo JA, Tung GA, Rogg JM, Davis LM, Brody JM, Mayo-Smith WW: Voice recognition dictation: radiologist as transcriptionist. *J Digit Imaging* 21(4):384–9, 2008
2. Koivikko MP, Kauppinen T, Ahovuo J: Improvement of report workflow and productivity using speech recognition—a follow-up study. *J Digit Imaging* 21(4):378–82, 2008
3. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu JK, Peterson DE, Finkelstein J: The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inform Assoc* 13:16–23, 2006