Journal of Digital Imaging

Impact of Speech Recognition on Radiologist Productivity

Steve G. Langer, PhD

A survey was conducted of radiology practices with productivity data from at least 3 of the following 4 workflows: film with manual transcription, filmless with manual transcription, film with speech recognition, and filmless with speech recognition. Two surveys were submitted to candidate sites. The first was used to ascertain suitable available data for follow-up. The second survey requested data for report turn around times, full-time equivalent (FTE) staffing levels, and report volume. Data were collected and stored in a Microsoft Access database and statistical analysis performed in Excel. Whereas several metrics were used, the normalized figure of reports-per-day/FTE was found to have improved an average of 1.9 (for filmless with speech recognition) and 2.3 (for film with speech recognition) over the film with manual transcription case. Filmless with manual transcription was only 1.4 times the value of the all manual case. At the 10% confidence level, both filmless with manual transcription and film with speech recognition workflows were found to have statistically significant enhanced productivity. Insufficient data exist to show if the fully automated workflow (filmless with speech recognition) offers benefits over 2 previous semi-automated workflows.

KEY WORDS: Picture archiving and communication system, speech recognition, productivity, radiology

AS MEDICAL REIMBURSEMENTS decline in real dollars, solid evidence of potential cost savings and improved service is needed to justify purchasing new health care equipment. Many reports have documented productivity issues with picture archiving and communication system (PACS).[1,2,3,4,5,6] Somewhat fewer studies on speech recognition have examined the cost savings implied by their adoption and service improvement as defined by shortened report turnaround times (RTT).[7,8,9,10] However, it is not clear to what extent a synergistic reaction is possible with adoption of tightly coupled PACS/RIS/speech recognition and how that would affect RTT, although some investigators have hinted at a connection.[11,12]

To have realistic expectations of performance gains possible with speech recognition, it is useful to have a global understanding of the workflow in radiology and the possible points at which automation may be applied. The base case represents a film-based department with human transcription. In that environment, the workflow could look like the following (Radiology Information System [RIS], Electronic Medical Record [EMR]):

I. Film with human transcription
   (a) Patient arrival (technologist)
   (b) Examination performed (technologist)
   (c) Films delivered to radiologist (technologist)
   (d) Report dictated (radiologist)
   (e) Report transcribed on RIS (transcriptionist)
   (f) Report signed on RIS (automatically uploads to the EMR, radiologist)
   (g) Report hardcopy goes into film jacket (file room clerk)

At every step at which data (either visual or textual) are handed from one human to another, there is an opportunity for delay and error. Compare scenario one with the following:

From the Department of Radiology, University of Washington Seattle, Seattle, WA.

Correspondence to: Steve G. Langer, PhD, W2A Mayo, 200 I'st Str SW, Rochester MN 55905; tel: 507-266-4418; fax: 507-266-0884; e-mail: langer.steve@mayo.edu.

Online publication 6 November 2002

doi: 10.1007/s10278-002-0014-7

II. Filmless with human transcription
   (a) Patient arrival (technologist)
   (b) Examination performed (technologist)
   (c) Report dictated (radiologist)
   (d) Report transcribed on RIS (transcriptionist)
   (e) Report signed on RIS (automatically sent to images on PACS, and EMR; radiologist)

Note that in scenario 2, we eliminated exactly 2 steps: the manual delivery of images to the radiologist and attaching a hardcopy report to the film jacket. However, if we replace PACS with a speech recognition system that is coupled to the RIS, we have the following:

III. Film with speech recognition
   (a) Patient arrives (technologist)
   (b) Examination performed (technologist)
   (c) Films delivered to radiologist (technologist)
   (d) Report is dictated, edited, signed, uploaded to RIS and EMR (radiologist)
   (e) Report hardcopy attached to film jacket (file room clerk)

Scenario 3 eliminates an interface between the radiologist and human transcriptionist but adds back the need for a hard copy report and manual film delivery. (It also assumes that the radiologist is self-editing within the speech recognition application and not sending reports off to the transcriptionist/correctionist. As we shall see, this is not necessarily the way the speech applications are used, but this workflow does result in the greatest reduction in human interfaces.) Finally, we have the fusion of PACS with speech recognition:

IV. Filmless with speech recognition
   (a) Patient arrival (technologist)
   (b) Examination performed (technologist)
   (c) Report is dictated, edited, signed, uploaded to RIS, PACS, and EMR (radiologist)

In this scenario 4 complete steps are eliminated from the base scenario as well as the necessity for the radiologist to interact with any human staff. However, this is done at the expense of burdening the radiologist with tasks formerly performed by others. The critical question is, are reduced delays achieved by eliminating human interfaces offset by the additional radiologist workload?

To answer the previous question, one must study sites that have longitudinal data for radiologist's productivity across all 4 phases of the workflow previously described. In performing this survey, it also is necessary to discriminate between personal speech recognition products that do not interface with the RIS and those that do. Personal speech recognition systems without a RIS interface (such as ViaVoice or Naturally Speaking) cannot perform the automation synergy described in scenario 2 or 4[5] (ViaVoice, IBM Corporation, White Plains NY; Naturally Speaking, Scansoft, Peabody MA).

It also is necessary to determine the exact workflow that radiologists use within the speech recognition application. The major vendors in radiology speech recognition offer products that can support various workflow types (PowerScribe, Dictaphone, Stratford CT; Talk Technologies, Bensalem PA):

1. Dictate, self correct, sign in sequence (DSS)
2. Dictate, self correct, and sign in batch mode (DSB)
3. Dictate, allow recognition, use transcriptionist correction, then sign (DCS)
4. Dictate, prevent recognition, use transcription based on audio file, sign (DTS)

It is evident that methods 3 and 4 introduce additional human interfaces and delays, which preclude the 2 main advantages to be gained from speech recognition: reduction in transcription personnel and radically decreased report turnaround time. Therefore, a survey should note the predominant workflow in use at a site to accurately compare performance gains/losses among speech recognition users.

METHODS

An ongoing survey of over 40 North American sites has been undertaken to quantify the manpower and RTT efficacy of 4 possible workflow types: film with transcription, film with speech recognition, PACS with transcription, and PACS with speech recognition (as distinct applications or tightly coupled). An initial survey canvassed sites to ascer-

tain suitability for further follow-up. The second survey included *the following parts:*

1. *Film-based with manual transcription*
   Month/year
   Number of reports per day (average over month)
   Full-time equivalent (FTE) transcriptionists for that month
   FTE radiologists for that month
   Report turnaround time (RTTp)
   Report turnaround time finalized (RTTf)
2. *Film-based with speech recognition*
   Month/year
   Number of reports per day (average over month)
   % of reports done by SRec
   FTE transcriptionists for that month
   FTE radiologists for that month
   Number of SRec stations
   SRec vendor
   Workflow type (majority)
   RTTp
   RTTf
3. *PACS-based with manual transcription*
   Month/year
   Number of reports per day (average over month)
   FTE transcriptionists for that month
   FTE radiologists for that month
   Number of PACS workstations
   PACS vendor
   RTTp
   RTTf
4. *PACS-based with speech recognition*
   Month/year
   Number of reports per day (average over month)
   % of reports done by SRec
   FTE transcriptionists for that month
   FTE radiologists for that month
   Number of SRec stations
   SRec vendor
   Workflow type (majority)
   Number of PACS workstations
   PACS vendor
   Is SRec integrated with PACS station (do they auto-sync patient information?)
   RTTp
   RTTf

Results were collected from electronic mail and fax transmissions and compiled in Microsoft's Access database for analysis (Microsoft Corp, Redmond WA). Respondents were asked to complete each data phase for the same (or adjacent) month to reduce the effects of seasonal variation (eg, trauma centers whose workloads increase during the summer from increased recreational accidents).

RESULTS

To analyze the data obtained in the survey, it is useful to examine 3 figures of merit. Obviously, overall report turnaround time (RTT) from examination completion to finalized report is one, and reports per day is another. However, it also is useful to define a figure that helps to discriminate personnel and time efficiency. This is the normalized report productivity, defined as:

$$RP_n = \text{reports per day/total FTE}$$

(where the total FTE is the sum of the transcriptionists and radiologists at the site).

This figure effectively is a measure of personnel utilization and normalizes the output of staff to the daily report volume—thus making it possible to compare different institution's productivity. In the table that follows, each of these figures will be normalized to one for the film/manual transcription scenario. For the remaining work scenarios, the values computed will be compared with the unit baseline for film with manual transcription. Data in the table are averages over all respondents and do not necessarily indicate the trend in a single institution. See the associated poster or SCARU presentation for that correlation.

With just over 25% of sites completing the full survey and using the film/transcription workflow as a baseline, the averaged data are shown in Table 1 (numbers in parentheses are raw data).

DISCUSSION

Table 1 shows results averaged metrics (RTT, average reports per day, average FTE transcriptionist, FTE radiologists, and normalized report productivity) over all respondents, hence, should not be used indiscriminately to project trends for a single institution. However, one notes some useful points. First, the adoption of speech recognition by a practice that uses film seems to cut RTT by about 70%. Furthermore, productivity (averaged over all sites) was more than doubled over the film/transcription mode. However, the adoption of PACS without speech recognition reduced RTT by 70% but had a smaller impact on RPn. This finding may be because of the fact that few sites adopting PACS immediately show any reduction in FTE transcription positions. Finally, it is

Table 1. Summary Statistics

Phase	Average RTT	Average Rep/d	Average tran	Average radiolo	RPn
Film/manual trans	1 (48.2 h)	1 (240)	1 (6.3)	1 (12.2)	1 (16.2)
Film/speech recognition	0.32 (15.5)	1.3 (311)	0.73 (4.7)	.84 (10.2)	2.27 (36.7)
PACS/manual trans	0.27 (13.3)	1.03 (248)	.76 (4.8)	.87 (10.6)	1.37 (21.8)
PACS/speech-recognition	0.33 (15.7)	1.3 (310)	.39 (2.5)	1.06 (12.9)	1.89 (30.6)

Note. A paired *t* test also was performed at the 10% confidence limit to establish the significance of the results for RTT. Again, the film/manual transcription phase of workflow was used as a baseline and the *t* values compared with it for each subsequent workflow. The results are summarized in Table 2. The uncertainties in the RTT column represent the 90% confidence limit for the true average RTT for that population. The large uncertainties, particularly for the automated workflows, are a result of the small number of respondents and the resultant large standard deviations. A similar table could be shown for the RPn results, but they track the findings of Table 2 and offer no new insights.

Abbreviations: RTT, Report turnaround time to final signature; rep/d, reports per day. The ''tran'' and ''radiolo'' columns refer to the average number of transcription staff and radiologists on duty during the reporting period. RPn, normalized report productivity. Numbers in parentheses are raw data.

interesting to note that the combination of PACS with speech recognition does not offer the most rapid report turnaround time or even a better RPn then the film/speech-recognition phase. This apparent paradox is explained by noting that 16% of these sites were either not 100% on speech recognition or were using transcription services in a correction mode (DTS). Perversely, the film/speech-recognition group was 100% on speech recognition, but because of one site maintaining the bulk of its transcription group (as a backup), that group's FTE reductions in transcription were not as dramatic as they could have been.

Table 2 displays the statistical significance of the RTT findings at the 10% confidence limits for the various workflow phases versus the film/transcription baseline. Despite the paucity of data points (and the resulting large standard errors), the 10% confidence limits of all 3 alternative workflows either pass or just miss a significant difference from the baseline. One somewhat surprising result is that even though the average RTT fell from an average of 2 days (48 hours) to a value approaching one half of a day, the RTT is still about 15 hours. But there is great variability in this result (as will be seen in the Figs) and those sites who had the best RTT prespeech recognition continued to have the best results after adoption of speech recognition.

In Fig 1, cohort A shows a subset of 4 of the 8 sites that moved through the adoption of PACS and then speech recognition, whereas cohorts B shows the 2 sites that adopted speech recognition first. Site 9 in cohort A shows an extreme improvement in RTT, even in just the initial move from film to PACS. The other sites in this cohort also had RTT improvements in the PACS/transcription phase but not to such a degree.

In Fig 2, we note that all sites in both cohorts increased their RPn, but site 36 in cohort A and site 37 in cohort B had the least improvement.

Table 2. Statistical Significance

Phase	Average RTT (h)	T Value	Required *t* at 10%	Pass/Fail
Film/manual trans	48.2 ± 50%	NA	NA	NA
Film/speech recognition	15.5 ± 93%	1.73	1.78	Fail
PACS/manual trans	13.3 ± 119%	2.2	1.76	Pass
PACS/speech recognition	15.7 ± 98%	2.1	1.75	Pass

Note. A significant difference is said to exist (at the 10% confidence limit) if the computed *t* value is greater than or equal to the required value. The previous 2 tables show aggregated results over population statistics and as such should not be used to draw specific predictions about a specific site. To see trends among single institutions, Figures 1 through 5 track RTT, RPn, Average reports/day, average FTE transcriptionists and average FTE radiologists for 2 different paths to PACS/speech recognition. In each figure, cohort A (the first plot) shows the given metric's progression as the site moves from film/manual_transcription to PACS/manual_transcription to PACS/speech_recognition. The second plot (cohort B) shows the progression from film/manual transcription to film/speech recognition and finally to PACS/speech recognition.

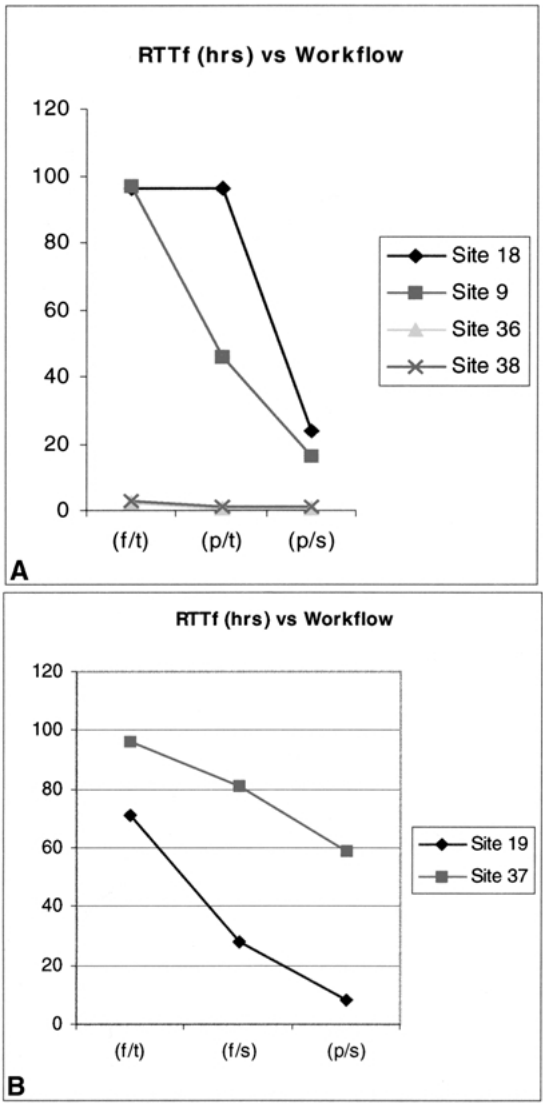


Fig 1. (a) Report turnaround times for Cohort A as a function of workflow as it progresses from film/transcription (f/t), to PACS/transcription (p/t) to PACS/speech recognition (p/s). (For clarity, only 4 of the 8 members of this cohort are shown. They do, however, represent the breadth of results in the survey.) (b) Report turnaround times for Cohort B as the site progresses from f/t to film/speech recognition (f/s) to p/s.

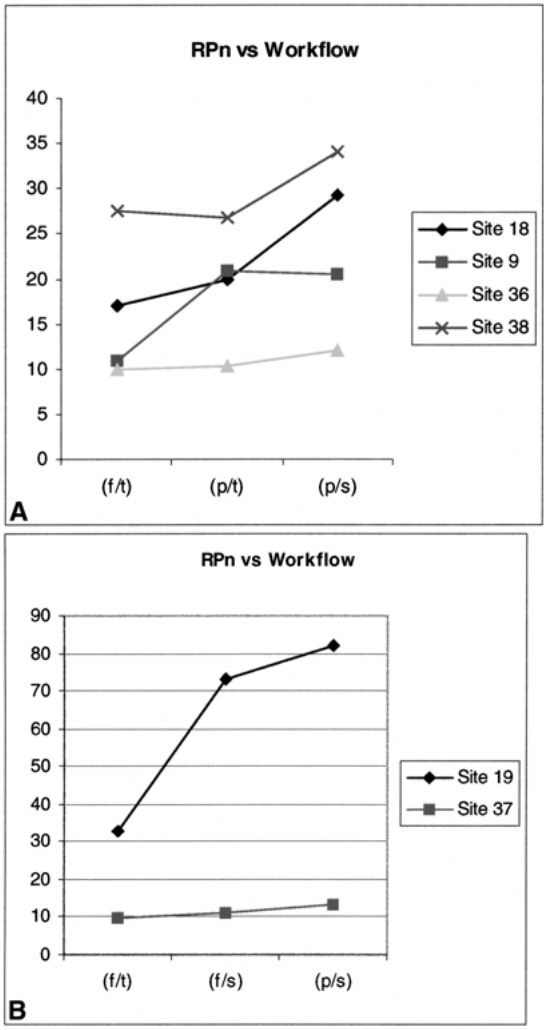
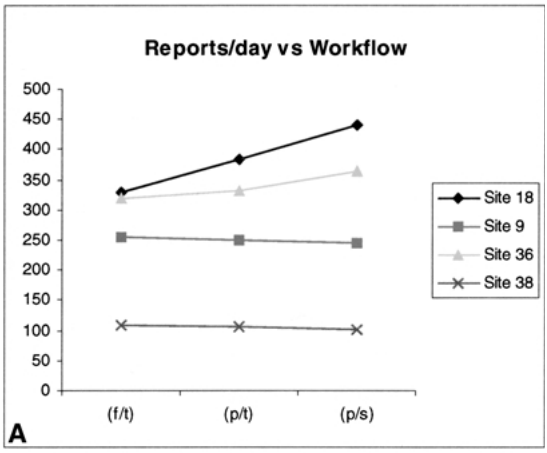


Fig 2. (a) Normalized Report Productivity (RPn) by PACS early adopters. (b) Same metric tracked by speech recognition early adopters. Both groups ultimately arrive at combined PACS/speech recognition.

Why? Site 36 used speech recognition for only 70% of all reports. Site 37 used speech recognition for 80% of reports but also used the DCS workflow, which reintroduces signing delays. It will be instructive to follow the remaining metrics for these 2 sites versus the others.

Figure 3 shows the growth in daily report volume over the time covered by the survey (most respondents reported data longitudinally over time as they introduced the new technologies). The exceptions are sites 9 and 38 in cohort A. Site 9 actually was flat over the survey period, whereas, 38 reported data from 3 different areas (using 3 different workflows) over the same month. However, the relationship between examination volumes did not necessarily correlate with radiologist staffing levels as will be seen.

In Fig 4, cohort A does not show an appreciable decrease in FTE transcription in the second phase at most sites. This seems odd, because the RPn values increased in the PACS/



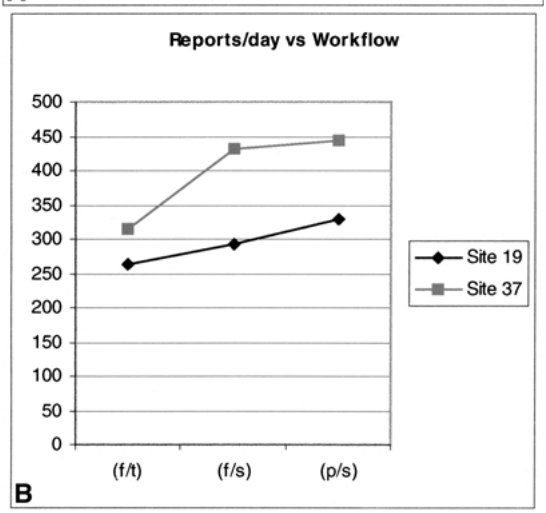
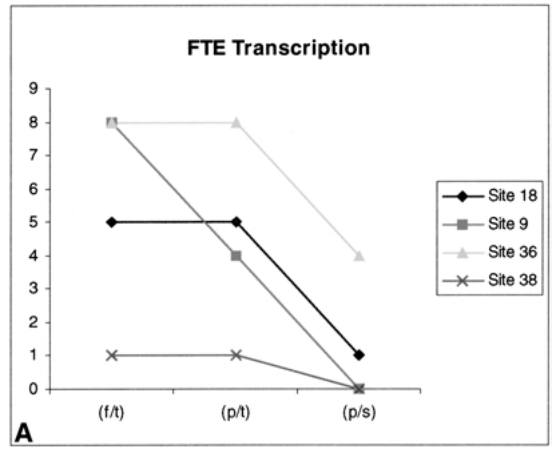


Fig 3. (a) Daily report volumes for the PACS first adopters. (b) Report volumes for speech recognition first adopters. Most sites enjoyed some examination volume increases over the 2 to 3 years covered by the longitudinal survey. The exceptions are sites that reported on all workflows during the same time frame (caused by coincident differing workflows in 3 distinct locations).



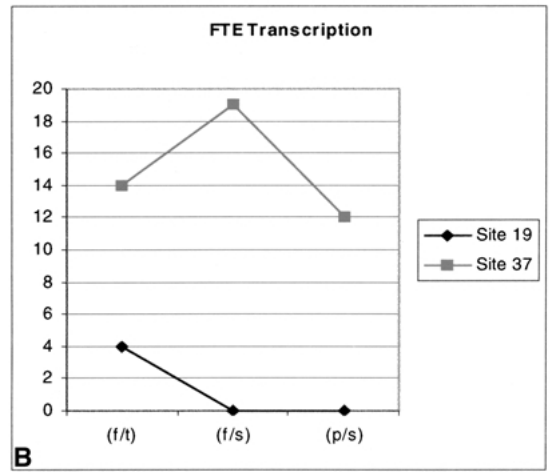


Fig 4. The average numbers of FTE transcriptionists by PACS early adopters (a) and speech early adopters (b). Note that Cohort A does not enjoy much in the way of FTE reductions in transcription in the first phase of automation. But, paradoxically, neither did Cohort B's site 37 because of their high use of DCS workflow in the f/s phase.

transcription phase; it would seem that some slight reductions could be made. Cohort B is even more interesting. Our problematic site 37 even increased, temporarily, the number of FTE transcriptionists because of their use of a DCS workflow.

Figure 5 allows us to examine the answer, at least partially, to our initial question, "Do the additional tasks placed on radiologists with adopting speech recognition result in measurable increased radiologist workload or staffing?" From Table 1 and this figure, most site's staffing of radiologists remained flat or decreased – *even when the daily examination volume increased.* The exceptions are site 36 from cohort A and site 37 from cohort B. Why are they different? Recall that they also are the sites that had the least improved RPn because they either had less then 100% speech recognition utilization or used DCS or DTS workflows (or both).

Thus, from the current data, it would appear that with moderate examination volume increases (even up to 30%/year) it is possible to maintain radiology staffing and decrease transcription staff with adoption of PACS and speech recognition. What we have not answered, however, is, are the radiologists working longer hours to accomplish this, or are they more efficient? To determine this, one would need more invasive survey that tracked the actual hours

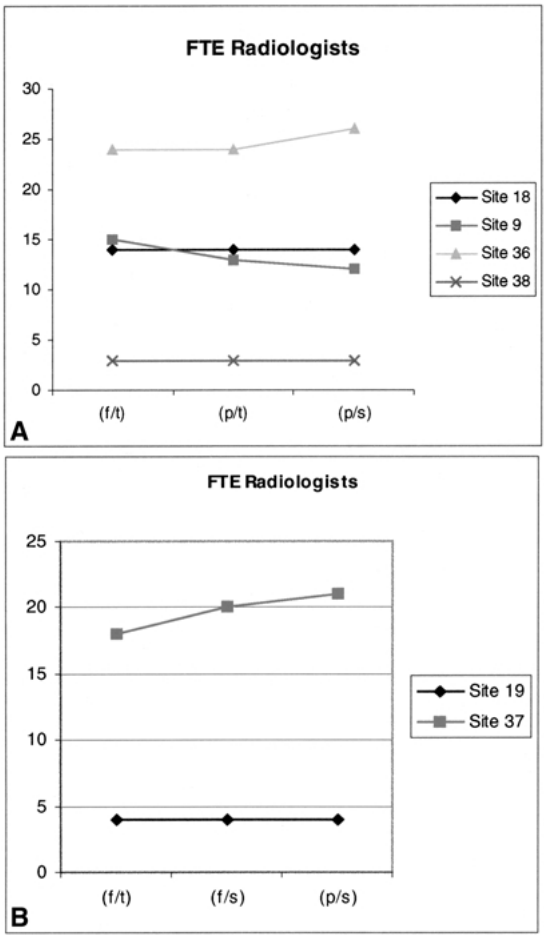


Fig 5. (a) FTE trends in radiologists for groups that adopted PACS first versus those who adopted speech recognition first (b). Note that even though most sites had increased examination volumes, the only ones who increased their FTE radiologists (sites 36 and 37) were those who had less then 100% speech recognition utilization or used DCS or DTS workflows.

worked per day for each radiologist and the time points for signing off on reports. Whereas it may be possible to get this level of RIS detail from one's home institution—it is unlikely that a general survey of national health care providers will be able to get significant compliance.

CONCLUSIONS

The adoption of either PACS or speech recognition or both improve RTT and RPn. This report has sought to break out the effects of each technology both separately and when applied in concert. The length of time for these technologies to pay for themselves depend on to the utilization and workflow used—because this determines the possible FTE reductions in transcription services. If one desires maximum improved RTT for minimum cost, it would appear that adopting speech recognition before PACS would be advantageous based on typical PACS installation costs and FTE transcription reductions. Finally, it appears that with efficient use of PACS and speech recognition, it is possible to moderately increase examination volumes while maintaining radiologist staff levels.

ACKNOWLEDGMENTS

The author would like to thank Talk Technologies for a grant in partial support of this work.

REFERENCES

1. Bryan S, Weatherburn G, Watkins J, et al: Radiology report times: Impact of PACS. Am J Roentgen 170:1153-1159, 1998

2. Krupinski EA, Lund PJ: Differences in time to interpretation between PACS workstations and film. Acad Radiol 4:177-182, 1997

3. Reiner BI, Siegel EL, Hooper FJ, et al: Effect of film-based verses filmless operation on the productivity of CT technologists. Radiology 207:481-485, 1998

4. Langer SG, Wang J: An evaluation of ten digital image review workstations. CJ Digit Imaging 10:65-78, 1997

5. Mattern CW, King BF, Hangiandreou NJ, et al: Electronic imaging impact on image and report turnaround times. J Digit Imaging 12:155-159, 1999

6. Mehta A, Dreyer K, Boland G, et al: Do picture archiving and communication systems improve report turnaround times? J Digit Imaging 13:105-107, 2000 (Suppl 1)

7. Zick RG, Olson J: Voice recognition software versus a traditional transcription service for physician charting in the ED. Am J Emerg Med 19:295-298, 2001

8. Gale B, Safriel Y, Lukban A, et al: Radiology report production times: voice recognition vs. transcription. Radiol Manage 23:23-27, 2001

9. Lemme PJ, Morin RL: The implementation of speech recognition in an electronic radiology practice. J Digit Imaging 13:153-154, 2000 (suppl 1)

10. Ramaswamy MR, Chaljub G, Esch O, et al: Continuous speech recognition in MR imaging reporting: Advantages, disadvantages, and impact. Am J Roentgenol 174:617-622, 2000

11. Pavlicek W, Muhm J, Collins J, et al: Quality of service improvements from coupling the digital chest unit with integrated speech recognition, information, and PACS. J Digit Imaging 12:191-197, 1999

12. Langer SG: Radiology speech recogintion: Workflow, integration and productivity issues. Curr Prob Diagnostic Radiol 31:95-104, 2002