Why Does It Take Longer to Read Digital Than Film-Screen Screening Mammograms? A Partial Explanation

Tamara Miner Haygood,[1] Jihong Wang,[2] Deanna Lane,[3] Eva Galvan,[4] E. Neely Atkinson,[5] Tanya Stephens,[3] and Gary J. Whitman[3]

Digital screening mammograms (DM) take longer to interpret than film-screen screening mammograms (FSM). We evaluated what part of the process takes long in our reading environment. We selected cases from those for which timed readings had been performed as part of a previous study. Readers were timed as they performed various computer manipulations on groups of DM cases and as they moved the alternator and adjusted lighting and manual shutters for FSM cases. Subtracting manipulation time from the original interpretation times yielded estimated times to reach a decision. Manipulation times for DM ranged from a low of 11 s when four-view DM were simply opened and closed in a 4-on-1 hanging protocol before moving on to the next study to 113.8 s when each view of six-view DM were brought up 1-on-1, enlarged to 100% resolution, and panned through. Manipulation times for groups of FSM ranged from 8.3 to 12.1 s. Estimated decision-making times for DM ranged from 128.0 to 202.2 s, while estimated decision-making time for FSM ranged from 60.9 to 146.3 s. Computer manipulation time partially explains the discrepancy in interaction times between DM and FSM. Radiologists also appear to spend more time looking at DM than at FSM before making a decision.

KEY WORDS: Full-field digital mammography (FFDM), mammography, image manipulation, image interpretation, screening mammography, efficiency

INTRODUCTION

Digital mammography was first approved for use in the USA in 2000[1] and since then has come into use in an increasing number of practices. Approximately 7% of US mammography practices had at least one digital system as of April 2005[1]. That percentage has risen gradually to 43% by October 2008[2]. Digital mammography offers economies in terms of image acquisition time and storage costs compared with film-screen mammography, and the Digital Mammography Imaging Screening Trial demonstrated a small diagnostic benefit in pre- and perimenopausal women with dense breasts[3,4]. Two European studies have recently confirmed a diagnostic advantage for digital mammography[5,6]. An offsetting disadvantage is that digital screening mammograms (DM) take longer to interpret than film-screen screening mammograms (FSM)[1,7]. Previous studies have

[1]From the Department of Diagnostic Radiology, Unit 1273, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030-4009, USA.

[2]From the Department of Imaging Physics, Unit 1352, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030-4009, USA.

[3]From the Department of Diagnostic Radiology, Unit 1350, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030-4009, USA.

[4]From the Yale College (Branford), P.O. Box 200623, New Haven, CT 06520-0623, USA.

[5]From the Department of Biostatistics and Applied Mathematics, Unit 0237, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030-4009, USA.

Correspondence to: Tamara Miner Haygood, Department of Diagnostic Radiology, Unit 1273, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030-4009, USA; tel: +1-713-5638823; fax: +1-713-5636633; e-mail: Tamara.Haygood@mdanderson.org

not, however, evaluated what part or parts of the process of interpreting DM causes the expenditure of extra time. We timed the different physical tasks associated with interpreting DM and FSM to determine how much, if any, of the extra time expended in interpreting DM results from performing the computer manipulations needed to view the images.

MATERIALS AND METHODS

Preceding Experiment

This paper reports an experiment that arose from a previous study[7]. In that study, four readers were timed by one of four trained observers while interpreting screening DM and FSM using their usual methods. Their interpretation times were recorded while reading actual clinical cases. Interpretations were performed without the input of a resident or fellow, and the interpretations constituted the real final interpretation of each study. Institutional Review Board (IRB) approval was obtained for this study, and patient consent was waived. DM were viewed on a Philips Stentor 3.3 (Foster City, CA, USA) workstation, and FSM were prehung by film library staff with comparison films, if any, on a dedicated alternator (Crystal-Viewer, S&S XRay Products, Houston, TX, USA). When comparison studies for DM were on film, these films were also prehung on a CrystalViewer alternator. Reports for both DM and FSM were entered by the radiologist using MagView Mammography Information Management System software (Burtonsville, MD, USA) on a separate computer. All interpretations were performed at times when the workstation or alternator was working properly. All readers were faculty diagnostic radiologists, board certified by the American Board of Radiology, and qualified to practice mammography in accordance with the Mammography Quality Standards Act.

The four readers' average experience in interpreting FSM at the beginning of the previous experiment[7] was 10 years, with a range of 1 to 16 years. The four readers averaged 14 months (range 9 to 18 months) of experience in interpreting DM. All had 4 months experience with DM on the Stentor workstation. The average number of screening mammograms interpreted by the four readers in the preceding year was 1,254 (range 947 to 1,564). Three of the four readers limit their practice to breast imaging. The fourth has a more varied practice. The study began in August 2006 and ended in May 2007.

The observer recorded the time required to interpret each case and enter the report into MagView. Recorded times were taken from the Central Time Zone setting on the official US government time website (www.time.gov). This program was kept running in the corner of the computer monitor on which MagView ran, so the time was always visible to the observer. A program built into the data collection spreadsheet later calculated the elapsed time.

Present Experiment

For the study presented here, our IRB granted exemption from approval and required no informed consent. We chose sets of cases from among our previously timed interpretations. These cases were all bilateral screening mammograms without implants originally interpreted as Breast Imaging Reporting and Data System (BIRADS) category 1 or 2 (negative or benign, respectively). We excluded BIRADS category 0 cases, examinations which had previously been excluded from the earlier study due to an unusual cause of prolonged interpretation time and cases for which the radiologist had personally selected and hung films. Our intention was to compare the time required for the physical tasks associated with interpretation of DM and FSM with the known original interpretation time for relatively basic simple interpretations. Thus, the exclusion criteria were intended to separate out complex atypical cases. Once these exclusions were made, we did not allow the original interpretation time to enter further into the selection or rejection of a case. Before being included, all cases were viewed by one investigator to exclude any that had findings that, even though qualifying as BIRADS 2, might serve as a distracter. One DM case that otherwise would have been included was rejected when its barcode, used to bring up the images on the computer, was found to work inconsistently (Barcodes were originally used for both FSM and DM to access the appropriate report template in MagView, and both at original interpretation and in this study, barcodes were used to bring DM studies up on the

Table 1. Characteristics of Groups of Studies

Group	DM or FSM	Views	Number of studies in the group	Comparison studies	Readers for timed original interpretation
A	DM	4	10	Yes	1 and 2
B	DM	6	10	Yes	1 and 2
C	DM	6	10	Yes	3 and 4
D	FSM	4	8	No	1 and 3[b]
E	FSM	6	8	No	1, 2, and 4[b]
F	FSM	4	8	Yes	1,2,3, and 4[b]
G	FSM	6	7	Yes	1,2,3, and 4[b]
H[a]	FSM	4 or 6	8	Yes or no	1,2, and 4[b]

[a]Group H studies were mixed for hanging as might occur in a clinical situation if the number of films in a study and the presence or absence of comparisons are not considered factors in deciding the order at hanging
[b]Studies interpreted by specific readers were deliberately sorted for groups A, B, and C. We did not deliberately sort by reader for groups D through H

workstation). We grouped the examinations according to whether they had included four or six views and according to the presence or absence of comparison studies. DM studies were also grouped according to the viewing method of the original interpreting radiologist. These methods were known not merely by self-reporting but also by direct observation during the previous study [7] (though not discussed in resulting paper). Readers 1 and 2 used similar methods to one another, and readers 3 and 4 also used similar methods to one another. A description of the characteristics of each group of studies is included in Table 1.

Radiologists who performed the physical tasks described in this report are the same individuals who performed the original interpretations. The physical tasks that were timed in this study were the same ones needed to perform the original interpretations. Readers 1 and 2 performed various physical tasks with the DM images on one of the two Stentor workstations used for original interpretation and timing (Table 2). Readers 3 and 4 repeated those physical tasks that related to bringing the images up to 1-to-1 display and panning through them. FSM studies were hung in the same way that they had originally been hung for interpretation, and

Table 2. Timed Manipulation of DM

	Group	Manipulations arranged by increasing complexity[a]	Reader	Manipulation time (s)	Decision-making time (s)
1	A (DM, 4 views with comparisons)	1	1	11.0	177.9
2	A	1	2	13.2	175.7
3	B (DM, 6 views, with comparisons)	2	1	16.1	199.1
4	B	2	2	13.0	202.2
5	A	3	1	19.3	169.6
6	A	3	2	21.9	167.0
7	B	4	1	27.1	188.1
8	B	4	2	24.7	190.5
9	C (DM, 6 views, with comparisons)	5	1	113.8	128.0
10	C	5	2	82.8	159.0
11	C	5	3	66.0	175.8
12	C	5	4	82.8	159.0

[a]Manipulations arranged by increasing complexity: 1 = swipe the barcode, open the study with 4-on-1 display mode, go on to the next study; 2 = in addition to 1 above, bring the two additional images up in the 4-on-1 display mode; 3 = in addition to 1 above, double mouse click on each image of 4 to bring it up in fit-screen display mode, double click again to return to 4-on-1 display mode; 4 = in addition to 2 above, bring the two additional images up in the 4-on-1 display mode then in fit-screen display mode; and 5 = after each of six views was brought up in fit-screen display mode on monitor, a right mouse click and selection from two subsequent drop-down menus enlarged the image to 1-to-1 display mode, then the mouse was used to drag the image until all parts of it had passed across the monitor screen. In addition to the above, for all sets, the readers also clicked once with the mouse on the control monitor to determine the full number of images in the study and moved paperwork

readers 1 and 2 moved the alternator and adjusted the lighting for each case in the same way that they would do if interpreting the study.

It was not possible to recreate exactly the setting in which these FSM studies had originally been interpreted because the studies preceding or succeeding them on the day of interpretation would not necessarily be included in this study. For this reason, we performed separate timings on several groups of mammograms, sorting them variously by number of films and the presence or absence of comparisons, so that a larger or smaller number of lighting adjustments would be needed to ready each case for reading. Cases were considered ready to read when the alternator panel had been moved, lights were turned on or off as necessary to illuminate only the panels with films, and the manual light shutters had been moved to accommodate the size of the films. The alternator used was separate from but of the same type as the alternator on which films had originally been interpreted for the first timed reading study.

Before each of these physical tasks were timed, readers were instructed in the particular tasks to be performed and were allowed opportunities to ask questions and to practice until they felt comfort-

able with the tasks. This training was performed for both DM and FSM timing.

For each group of studies, we created patient jackets, and a sheet of paper was placed in the front of the jacket to simulate the request form that would be placed for actual clinical reading. For DM studies, the examination was called up using a barcode on this sheet of paper, just as we do in clinical reading. For both DM and FSM studies, the reader removed the sheet of paper, placed it to the side in a stack, and then shifted the jacket out of the way to another stack in between each study. This maneuver simulated the paper processing routines that would have been performed when the studies were originally interpreted and timed.

All readers limited themselves to performing the physical tasks necessary to bring the DM computer images up for reading or to make the FSM images ready to read and to shift the paper and jackets. Readers did not attempt to interpret or examine the studies and understood that all images they were viewing had already been interpreted and found to be free of any lesion requiring additional evaluation. Timing of the physical tasks was performed when the computer and alternator were operating correctly. A trained observer, who had no other

*Obtained by averaging the averages of all readers performing the relevant group of manipulations
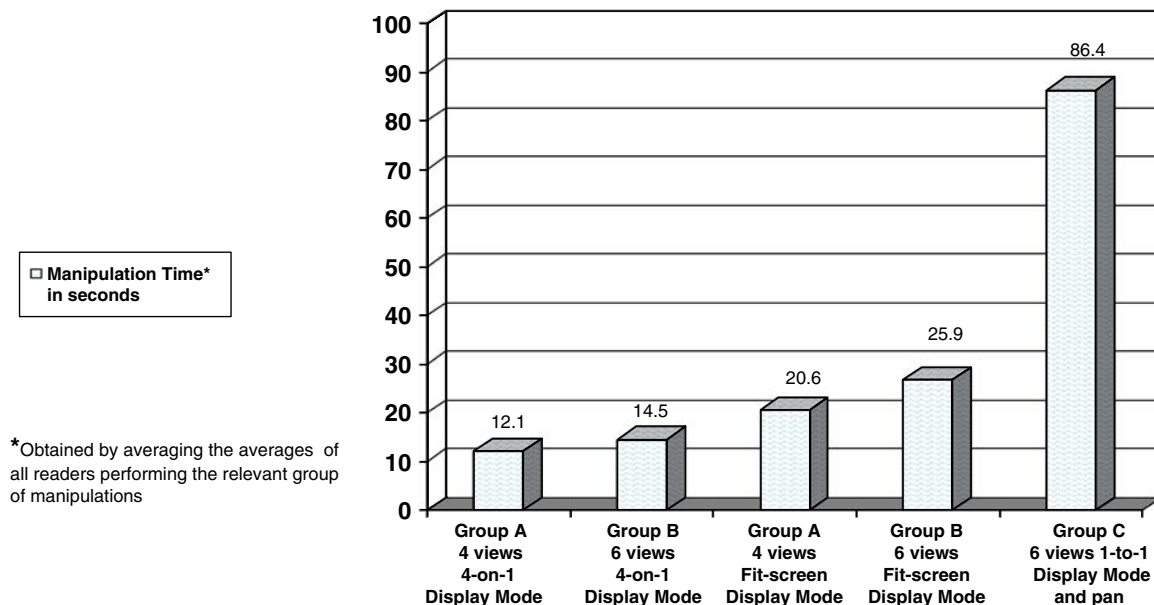
Fig 1. Average manipulation times for DM groups, arranged by increasing complexity of manipulation. Time expended in computer manipulation increased gradually with the complexity of the operation except for a large sudden increase when the image was brought to 1-to-1 display mode and panned through. The 86.4-s value for manipulations with panning reflects an average of the time used by all four readers. The average value for readers 3 and 4 is 74.4 s.

role in this study, timed the manipulations using a time stamp drawing on the internal clock of a laptop computer. This study was performed during July 2007.

RESULTS

Data may be more clearly understood if we define a few terms. "Manipulation time" means the average time expended by a reader in performance of a particular set of physical tasks for a particular group of mammograms. "Original interpretation time" means the average time expended by the readers at the time they interpreted the studies in the relevant group in conjunction with our first study[7]. "Decision-making time" means the time remaining when manipulation time is subtracted from original interpretation time. "4-on-1 display mode" means the image display method in which four images are displayed on one monitor. "Fit screen display mode" means the image display method in which the data from a single image are rebinned to fit on one 5-megapixel monitor. "1-to-1 display mode" means the image display method in which the image is magnified beyond the physical size of the surface of the 5-megapixel monitor so that each pixel on the monitor corresponds to and displays 1 pixel at acquisition. Depending on manufacturers' terminology, this display mode may also be called "full resolution" or "100% resolution."

Table 2 summarizes manipulation times and original interpretation times of DM studies considered in this investigation. Manipulation times gradually increased with increasing complexity of the physical tasks (Fig. 1) except for a sharp increase for panning through the six 1-to-1 display mode images. The average manipulation time for all four readers for this set of manipulations was 86.4 s. Manipulation time for readers 3 and 4, who routinely perform this maneuver in clinical interpretation, averaged 74.4 s.

Based on observations made during our original study (unpublished data), our readers differed in their usual methods of interpretation. Readers 1 and 2 typically brought each image up in the fit-screen display mode and selectively zoomed in on specific areas of interest using the track wheel in the middle of the mouse (Fig. 2a, b). Readers 3 and 4 typically brought each image up in the fit-screen

display mode then used a right mouse click and selection of options from two subsequent drop-down menus to shift to 1-to-1 display mode and then panned through the image (Fig. 3; The panning is necessary for this style of reading because the resulting image is larger than the monitor. We do not have a feature that allows individual quadrants of the image to be enlarged one by one with a mouse click). Therefore, lines 7 and 8 and lines 11 and 12 of Table 2 illustrate the physical tasks that would actually have been performed by the particular readers in originally interpreting the examinations. 1-to-1 display mode and panning (lines 11 and 12) took nearly three times as long as bringing each image of the study up in fit-screen display mode. The original interpretation time for the studies for which 1-to-1 display mode and panning were performed as part of the original interpretation was also longer than for the studies for which 1-to-1 display mode and panning were not performed.

Table 3 summarizes the manipulation times and original interpretation times of FSM studies considered in this investigation. Manipulation times tended gradually to increase as the number of adjustments of light panels increased, but there were only 3.8 s difference between our fastest and slowest manipulation times.

DM and FSM manipulation times are similar when comparing FSM manipulations for our most complex group of studies, group H, and DM manipulations for our simplest computer manipulations, opening and closing a four-view set of DM studies without additional magnification. Bringing each image of a six-view DM study up in 1-to-1 display mode and panning through each image took more than seven times as long as the FSM manipulations for group H studies.

DISCUSSION

It is fairly well established that full-field digital screening mammography is at least diagnostically equivalent[8–11] and probably better[3–6] than film-screen screening mammography. It is also fairly well established that it takes significantly longer to read DM than FSM studies[1,7,12]. Why it takes longer and what might be done to bring interpretation times more in line with those for film-screen screening mammography is not so well
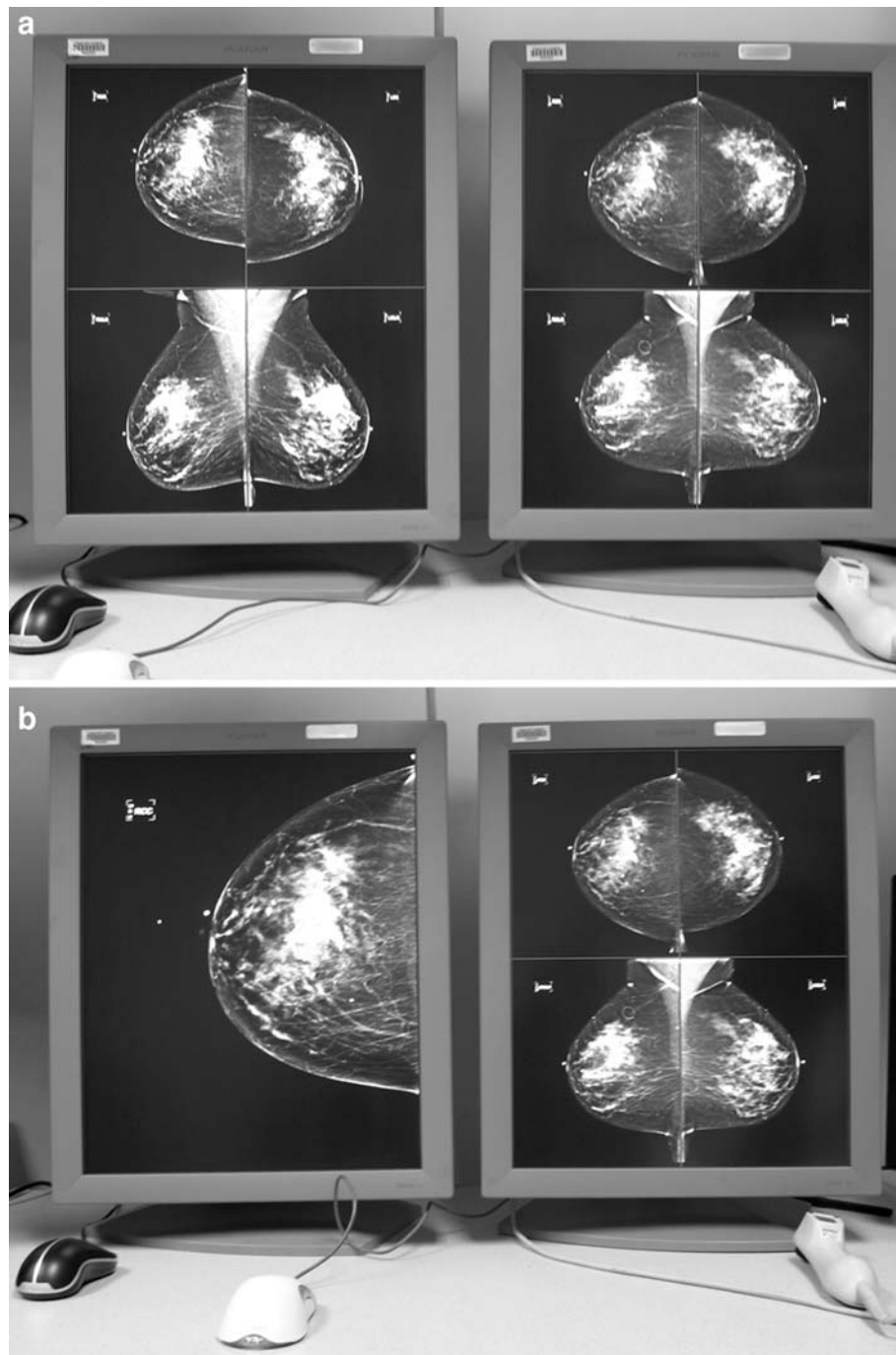
Fig 2. a Side-by-side monitors displaying a mammogram with comparisons. According to our tested hanging protocol (and the one actually used clinically by three of our four readers), the current study appears on the left monitor and the comparison study appears on the right monitor, both in 4-on-1 display mode. Craniocaudal (CC) views are on the *top*, and mediolateral oblique views are on the *bottom*. **b** As the first step to any further magnification, the reader performs a double mouse click on one image, and that image (in this case, the right CC) is enlarged to fit-screen display mode.
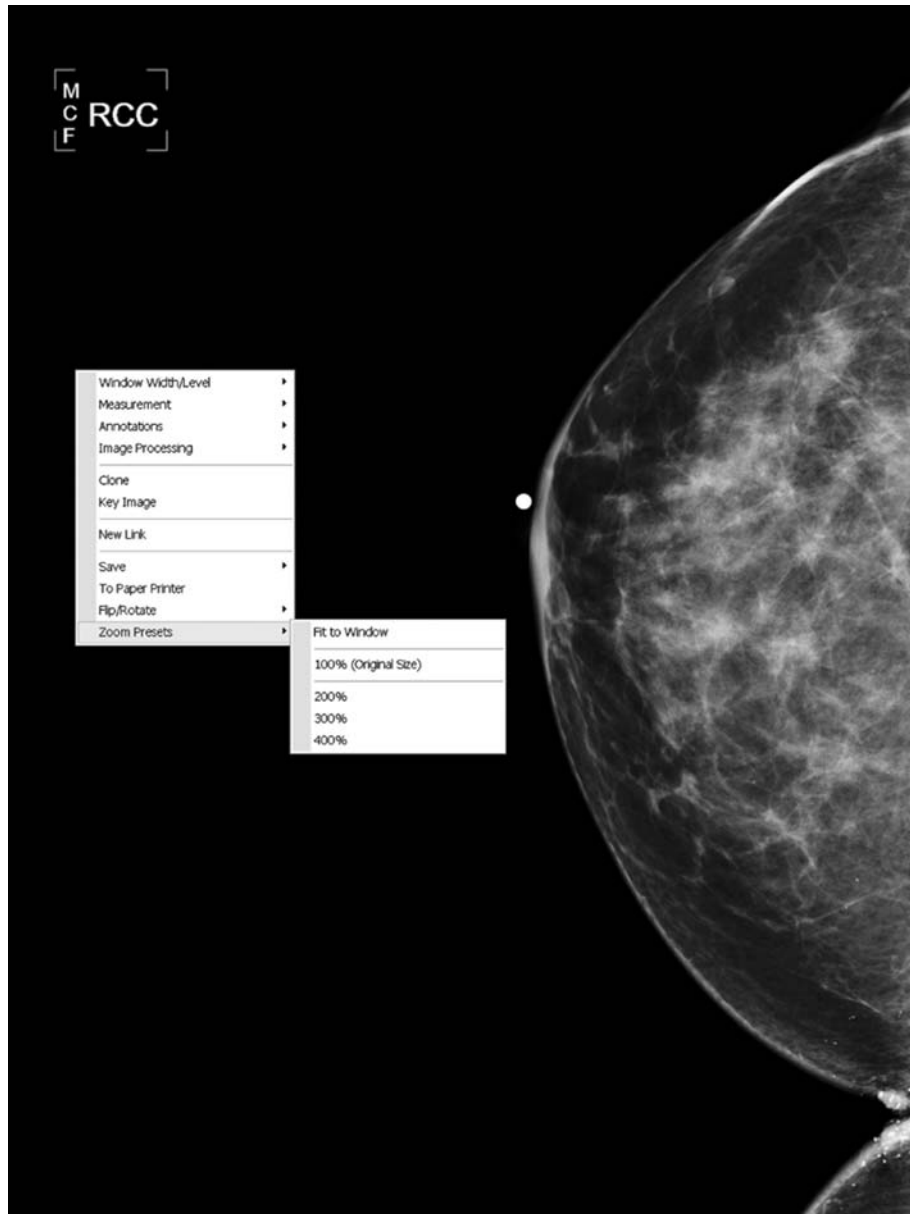
Fig 3. To shift from fit-screen display mode to 1-to-1 display mode, the reader right clicks on the image and then selects "zoom presets" and "100% (original size)" from two drop-down menus.

known. The mere introduction of soft-copy viewing alone may not be the explanation for longer interpretation time. Herrmann et al.[13] compared manipulation time for reading digital chest radiographs with time necessary to shift from study to study on an alternator and found that computer manipulations were quicker. Repeated magnification of the images, however, will increase the manipulation time. Hemminger[14] found that computer-based zooming and panning through an image was more cumbersome and slower than using a magnifying glass and moving it and one's eyes around a film-screen image and that extra time spent panning correlated with longer interpretation times.

We have compared the manipulation time required to do the repetitious physical tasks used in our reading environment for interpretation of FSM with those used to interpret DM. Though many different combinations of physical tasks

Table 3. Timed Manipulation of FSM

Group	Reader	Manipulation time (s)	Decision-making time (s)
D (FSM, 4 views, no comparisons)	1	10.3	60.9
D	2	8.3	62.9
E (FSM, 6 views, no comparisons)	1	8.9	100.7
E	2	8.3	101.3
F (FSM, 4 views, with comparisons)	1	9.7	70.6
F	2	8.5	71.8
G (FSM, 6 views, with comparisons)	1	11.0	146.3
G	2	11.0	146.3
H (FSM, Mixture of 4 and 6 views, with and without comparisons)	1	11.8	86.9
H	2	12.1	86.5

The manipulations in each case were to move the alternator panel, adjust the lighting, and move paperwork

could have been tested, we tested combinations that our readers actually used in the previous timed study. If one were to limit one's computer manipulations to bringing the digital images up in 4-on-1 display mode, the manipulation time for DM would approximate that for FSM. This, however, would require the rebinning of data from an image that is already larger than our available monitors into just one fourth of the monitor, and though we have not tested attempts to read only from this type of image, our subjective impression is that it would be difficult or impossible to achieve an acceptable cancer-detection rate without further enlargement of the images.

All other manipulations beyond bringing the study images up in 4-on-1 display mode increase the manipulation time beyond that of FSM, with a slow steady increase in time expended as the complexity of the manipulations (short of panning) increases. With any combination of computer manipulations, therefore, at least some of the time difference for interpretation of DM compared with FSM can be explained by the time expended in performing the most basic physical tasks of manipulating the computer interface. If one subtracts the manipulation time from original interpretation time, one arrives then at an estimate of the time used for the more cerebral parts of image interpretation and report entering, the decision-making time. The estimated decision-making time is also greater for DM than for FSM, so our radiologists apparently spent more time studying the DM than the FSM images independent of the time they spent performing the basic computer manipulations.

There is a sharp increase in manipulation time with 1-to-1 display mode and panning. Surprising-

ly, however, decision-making time drops for studies interpreted with panning. The reason for this is not clear from our data, but we suspect it may be because radiologists who habitually pan through the 1-to-1 display mode image may then be content to make their decisions with less frequent use of ancillary tools. For example, there may be less desire to zoom selectively on specific areas of interest or less tendency to go back and forth from one image to another. This would be an interesting avenue for future research. This suggestion would then also imply that some of the decision-making time for DM interpretations is actually also taken up with computer manipulations. If a radiologist brought each image of a six-view DM up on the monitor in fit-screen display mode and examined each image then wanted to go back and look at one of the images a second time, this would entail additional time spent both in visual inspection of the image and in manipulation of the computer to bring the image up. At the same time, of course, some decision making occurs simultaneously with computer manipulation. While one image is being brought up in fit-screen display mode, for example, the reader can be thinking about what was seen on the 4-on-1 display mode.

Despite the drop in decision-making time, original interpretation time remains higher for studies interpreted with 1-to-1 display mode and panning. This suggests both a need to evaluate the diagnostic efficacy of various interpretation methods and the potential value of higher-resolution monitors such as 8- or 9-megapixel monitors which would allow all the information gained during original acquisition of each view of the screening mammogram to be displayed on the monitor at once, without panning.

Fig 4. Photograph of the control panel of the alternator. All control buttons are clustered in the middle third of the table. The button marked *next* moves both the top and bottom film holders to the next panel and stops them there.

Our study was performed using specific equipment, and different results might have been obtained with different equipment. The Crystal-Viewer alternator has several features that tend to facilitate quick transition from case to case. Cases are arranged on large drums that rotate to move the panels from side to side, unwinding cases from one side and wrapping them back up on the other side. This arrangement makes for fairly quick movement among adjacent cases. The alternator also has a button that allows smooth coordinated movement of both the top and bottom panels from one case to the next with the push of that single button. Light controls are in physical proximity to each other and to the button for moving the cases, minimizing hand motion (Fig. 4).

Our computer system also has several features that facilitate transition from case to case. We keep all patient studies on-line at all times in two mirror-image storage systems, each of which grows by approximately 1 TB per month. If a defect in the primary storage system necessitates a query of a portion of the secondary storage system, this may delay presentation of comparison views by a second or two, but ordinarily, all images are essentially instantaneously available. The lookup table we use is the Digital Imaging and Communications in Medicine Grayscale Standard Display Function. The network has a 10-GB backbone. The computer used for the workstation is a Dell Precision Workstation 690 with 4-GB random access memory and 2.66-GHz processor speed.

Our study relies on background data related to the interpretation speeds of specific individual radiologists, and radiologists are well known to vary in their speed of interpretation of exams[1,14–16]. Therefore, some differences in results might also be obtained with different readers. The original interpretation time also includes the time for entering reports into MagView. Other methods of reporting may be either faster or slower.

Our manipulation times may also have reflected variations from time to time in the operating speed of the equipment, particularly the computers. Although our intention was to perform both our original interpretations and our timing of physical tasks when there was no noticeable slowing of computer speed, small differences in computer speed from day to day or minute to minute could have contributed to our other observed variations, particularly as there was no consistent trend among individual readers regarding speed. The individual who was a little faster with one set of manipulations was not necessarily faster with the next.

Several sets of timing of physical tasks were performed with six-view DM mammograms. This is because technologists at our institution obtain more than four views about 75% of the time when performing digital screening mammography, so for

us, a six-view DM is fairly typical of what we are presented to read. This is not true either for our FSM studies, approximately 21% of which have more than four views, nor is it true for DM studies in all practices[1], and one might expect a proportionate decrease in manipulation time for four-view DM studies compared with figures we have given for six-view studies.

There are other ways in which our DM and FSM groups are not replicas of one another. For DM, we made the assumption that the time required to prepare a case for interpretation was independent of the type of case that had preceded it. A four-view mammogram with a comparison study was going to take a certain amount of time to come up on the monitor and then be adjusted to the display desired for interpretation (or the display being tested at that moment). That time might vary a bit from case to case depending on factors described above including the operating speed of the computer at that moment and the smoothness with which the reader performed the necessary physical manipulations of the mouse and paperwork, but it was not going to vary depending on what sort of image had been looked at previously. The situation is quite different for FSM. A reader can go quickly through a group of studies if they each have the same number of images on the same size film and all either have or do not have comparison studies, because there is no need to adjust the shutters or turn lights on and off between cases. When the cases vary in size and number of films and in the presence or absence of comparisons, speed slows down because of the need to adjust the shutters and flip the lights on and off between cases. Therefore, group H was the model for the most complex grouping of FSM cases we could devise, in which there was considerable variation between studies and lots of shutter adjusting and light flipping were needed. For DM, complexity relates not to the nature of the preceding exam but to what one does with the study at hand; therefore, the complexity equivalent of group H FSM studies was for DM exams the timing session in which we brought six-view mammograms up in 1-to-1 display mode and then panned through them.

Finally, this experiment reports time required for performing, under artificial conditions, specific sets of stylized physical tasks that would be performed in interpretation of the relevant studies. The experimental environment itself may have affected the outcome in some ways. For example, for both DM and FSM, we had the readers manipulating the actual patient images interpreted in the original study. Despite directions to readers to stick with the specified physical tasks and despite our attempts to exclude eye-catching abnormalities, it is possible that readers spent a second or two now and then eyeing the pictures. They are, after all, radiologists and that is what radiologists do. A second example is that we find clinically that FSM studies, though prehung for us in a fairly organized fashion, often need a tug or two to get them just right before interpretation. Readers did no tugging or adjusting of film position in this study, and this may have lowered the manipulation times for FSM groups compared with what would be done at actual interpretation. It is also likely that in actual interpretation, radiologists would often have performed some additional manipulation of DM images.

CONCLUSIONS

For both DM and FSM, the time expended to manipulate images increases with the complexity of the manipulations, but the increase is particularly sharp when panning through 1-to-1 display mode DM images is included. Manipulation time for FSM case groups was shorter than for DM case groups, with it taking more than seven times longer to bring six images up in fit-screen display mode and then to 1-to-1 display mode and pan through them than to move alternator panels and adjust lights for the most complex set of FSM cases, one in which separate light adjustments were needed for every case.

Computer manipulation time partially explains the discrepancy in interpretation times between DM and FSM. Radiologists also appear to spend more time looking at DMs than at FSMs before making a decision.

ACKNOWLEDGMENTS