Characterizing the range of simulated prostate abnormalities palpable by digital rectal examination

Leigh A. Baumgart, Gregory J. Gerling *, Ellen J. Bass

Department of Systems and Information Engineering, University of Virginia, 151 Engineer’s Way PO Box 400747, Charlottesville, VA 22904 USA

1. Introduction

Prostate cancer has a high incidence rate (one in six for men in the U.S.), with an estimated 192,280 new cases in 2009. When diagnosed in an early and less aggressive stage, the five-year survival rate approaches 100% [1]. To promote early detection, the American Cancer Society advises that screening via the digital rectal examination (DRE) and prostate specific antigen (PSA) blood test be conducted concurrently. The DRE is important because the PSA tends to both over diagnose (65–75% of findings reported as false positives for PSA greater than 4.0 ng/l [2]) and miss cancers (15.2% of findings reported as false negatives for PSA less than 4.0 ng/l [3]). Although the DRE plays an integral role in early detection and is a skill clinicians are expected to learn, the perceptible limits surrounding this exam are unknown. Therefore, there is no basis from which to set reasonable expectations about clinical performance or to develop appropriate training.

When conducting a DRE, the clinician’s task is to detect hard nodules that vary in size, depth and hardness or prostate enlargement that varies in volume change and stiffness. The former typically relate to carcinoma, the latter signal benign prostatic hyperplasia (BPH) or prostatitis [4]. The size, depth, and hardness of nodules and relative stiffness of a given prostate contribute to the perceptible range of abnormalities. At present, neither the thresholds of absolute detection nor variance in ability between examiners have been identified.

In contrast, palpable limits have been studied in terms of the clinical (CBE) and self (BSE) breast exams [5]. In two studies with rubber-like materials, abnormality size emerged as the major dimension affecting the detection of lumps [6,7]. In general, larger lumps in more shallow positions pose the least difficulty. However, simulated lumps as small as 3.0 mm diameter were detectable when embedded in breast-like materials (which is an order of magnitude more pliant than prostate tissue) [9–10]. Aside from the lump size findings, abnormality depth and hardness appear to have a minimal impact, whereas the stiffness of surrounding tissue may decrease one’s ability to detect deeper lumps [11].

Hall et al. have shown that training on silicone models effectively increases exam performance on natural breast tissue [12]. Most of their training, and that prescribed by others for use with silicone models [13–17], takes place at the level of hands-on
skills. There is a focus on tactile skills because knowledge of disease and attitudes about domain are not strongly related to proficiency [18] and basic rules for diagnosing abnormalities as certain diseases are not difficult to master [19]. One major prerequisite for improving hands-on skills, however, is knowing the limits of tactile sensation, in this case with respect to the DRE.

When characterizing the limits on performance, the DRE differs from the breast exam in several key ways. First, a rectal wall is positioned between the finger and prostate, in addition to a glove and lubricant. Second, the clinician uses a single finger and is more constrained in search movement. Third, the prostate is a stiffer gland with less volume. Fourth, the exam typically takes place in less than 30 s in contrast to 2 min for the breast exam [7,20]. With these differences in mind, one common factor is that clinician performance in both exams does benefit from training.

In this work, the overall goal is to determine the perceptible limits of simulated abnormalities of various size, depth, and hardness within substrates of different stiffness when the examiner is constrained as with a DRE. In addition to determining the thresholds of absolute detection, we seek to determine the degree of variance in ability between examiners.

2. Methods

To analyze the limits of tactile perception in the DRE, we conducted a human-subjects experiment with 18 participants, using simulated prostates where abnormality size, depth, and hardness were varied within substrates of two stiffness levels. The objectives were to determine (1) the size of abnormalities detectable above 75% of the time at three discrete depths, (2) how substrate stiffness impacts the detectability of size/depth combinations, (3) if changes in abnormality hardness (consistent in objectives 1–2) impact detectability over size/depth combinations, (4) if some abnormalities require a minimum hardness be consistently detected, and (5) the variance in ability between participants.

2.1. Apparatus

An apparatus was built specifically for this study. The apparatus utilized silicone-elastomers to simulate the feel of prostate tissue and a rectal wall and employed a computer and electronics to control polyethylene balloons that simulated abnormalities. The computer also monitored the water pressure in the balloons and force on sensors embedded in the simulated tissue. The apparatus design is similar to that described in Ref. [19].

Twenty-three simulated prostates, 30 mm diameter and 20 mm tall, were mounted to a round platform that could be rotated so that the prostate under test was located beneath the examiner’s finger. The idealized cylindrical prostates did not include the surface undulations or an overall walnut shape, although the size was roughly the same as an actual prostate [19]. The platform containing the simulated prostates was housed within a structure that restricted access to and view of the simulated prostates. The examiner inserted his or her finger through an opening in the structure that was built of silicone-elastomer to mimic the rectal wall. The opening was angled at approximately 110° from the participant.

Each simulated prostate included a single polyethylene balloon embedded at one of three depths: 5, 10, and 15 mm. Balloons of seven volumes were used: 20, 40, 80, 200, 470, 1060, and 1770 mm³ that correspond to diameters of 4.0, 5.0, 7.5, 10.0, 15.0, 17.0, and 20.0 mm, respectively. Preliminary studies found that balloons of 3.0 mm diameter were inconsistently detected in stiffer simulated prostates and were not included. The balloons were inflated with water, thereby controlling hardness. Balloons could be inflated to be hard, like a rock, but were not detectable when deflated. In this study, three hardnesses were used: 23, 27, and 31 durometers, type Shore A. These fall within the range used for simulated breast tumors [6,7]. Water pressure sensors (Honeywell, SenSym Pressure Sensor, Model SX100DD4) monitored the water pressure over time, which was logged by the computer.

In addition to factors of depth, size, and abnormality hardness, simulated prostates of two stiffness levels were used: 21 and 82 kPa. These have been evaluated via compression tests and fall into the measured range of prostate stiffness [mean elastic modulus = 44.20 kPa, SD = 25.89 kPa [8]]. These stiffness values also fall in line with those deemed “realistic” in a subjective study with resident physicians and nurse practitioner students [19]. Located in the backing of each prostate were four, laterally spaced pressure sensors (Flexiforce 0–1 lb, Tekscan, South Boston, MA) which logged the examiner’s finger pressure over the simulated prostate.

2.2. Participants

Ten male and eight female participants (mean age = 20.4 years, SD = 1.38) were enrolled in the human-subjects study, approved by the IRB at the University of Virginia. No participant had prior clinical experience. A questionnaire also indicated that no participant had any remarkable prior experience working with his or her hands.

2.3. Experimental design

Using a modified version of the psychophysical method of constant stimuli [21], participants palpated the simulated prostates to determine the presence or absence of abnormalities. Typically the method of constant stimuli employs stimulus and blank trials presented in a randomized fashion where all stimulus combinations are presented an equal number of times. However, in the version we employed, we made three modifications to reduce participant fatigue. First, from all possible combinations (abnormality size, depth, hardness and substrate stiffness) only a subset of stimulus combinations were presented to participants (e.g., size 4 mm was used at 5 and 10 mm depth but not 15 mm depth). Pilot testing was conducted to remove combinations that were detectable 0% or 100% of the time. Second, the number of times that each stimulus combination was presented varied from two to four times depending on the difficulty of detecting the abnormality in the pilot study. Specifically, from the chosen subset of abnormalities, the most difficult to detect were presented four times, while the easiest to detect were presented two times. Third, due to hardware and time limitations, participants were presented with stimuli and blanks in one of six pre-determined random orders. Table 1 shows all stimulus combinations used and the number of times each was presented per participant in the experiment.

2.4. Procedure

Every participant participated in two experimental sessions, held on separate days for 90 min each. In session 1, each participant completed a 5 min pre-test questionnaire, a 5 min hands-on practice, and an 80 min hands-on experiment. During session 2, each participant completed a 5 min hands-on practice, an 80 min hands-on experiment and a 5 min post-test questionnaire. During sessions 1 and 2, participants palpated 96 simulated prostates, half of which contained an abnormality (the balloons were not inflated for the other half). Four participants returned for session 3, which was a 5 min hands-on practice and a 45 min hands-on experiment. Session 3
3.2. Effect of substrate stiffness on abnormality detection

Scenarios across participants were combined to determine the size of abnormalities consistently detectable (above 75% of the time for all participants combined) at the three discrete depths for two levels of prostate stiffness. Deeper abnormalities were more difficult to detect for all abnormality sizes. For the stiffer substrate (82 kPa), the abnormality size threshold was 5 mm diameter at 5 mm depth, 10 mm diameter at 10 mm depth, and 17 mm diameter at 15 mm depth (Fig. 1). For the more pliant substrate (21 kPa), the abnormality size threshold was 4 mm diameter at 5 mm depth, 5 mm diameter at 10 mm depth, and 15 mm diameter at 15 mm depth (Fig. 2). Therefore, both increases to substrate stiffness and abnormality depth make similarly sized abnormalities harder to detect.

3.3. Effect of abnormality hardness on abnormality detection

To investigate the effect of abnormality hardness (held constant in objectives 1–2) on detectability, we analyzed specific size/depth combinations that were expected to minimally surpass the 75% threshold when abnormality hardness was decreased (see circled data points in Fig. 1 and Fig. 2). Fig. 3 illustrates the impact of varying the hardness of an abnormality that is 10 mm diameter and 10 mm depth, for both 21 and 82 kPa substrates. At the lowest abnormality hardness (23 durometers, type Shore A) the detectability rate in the stiffer substrate (82 kPa) drops close to the 75% threshold from the 100% detectable threshold when its hardness is increased. In contrast, the detectability rate in the more pliant substrate (21 kPa) was unaffected by changes to abnormality hardness.

3.1. Absolute threshold – abnormality size, depth, and substrate stiffness combinations

Tables across participants were combined to determine the size of abnormalities consistently detectable (above 75% of the time for all participants combined) at the three discrete depths for two levels of prostate stiffness. Deeper abnormalities were more difficult to detect for all abnormality sizes. For the stiffer substrate (82 kPa), the abnormality size threshold was 5 mm diameter at 5 mm depth, 10 mm diameter at 10 mm depth, and 17 mm diameter at 15 mm depth (Fig. 1). For the more pliant substrate (21 kPa), the abnormality size threshold was 4 mm diameter at 5 mm depth, 5 mm diameter at 10 mm depth, and 15 mm diameter at 15 mm depth (Fig. 2). Therefore, both increases to substrate stiffness and abnormality depth make similarly sized abnormalities harder to detect.

3.2. Effect of substrate stiffness on abnormality detection

The detection rate for larger abnormalities in stiffer substrates (Fig. 1) is similar to smaller abnormalities in more pliant substrates (Fig. 2), given equal depth. For example, at a depth of 15 mm, 17 mm abnormalities in the 82 kPa substrate were detected at approximately the same percentage (80%) as 10 mm abnormalities in the 21 kPa substrate (74%).

Table 1

<table>
<thead>
<tr>
<th>Abnormality dimensions</th>
<th>Number of presentations at 82 kPa stiffness</th>
<th>Number of presentations at 21 kPa stiffness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (diameter in mm)</td>
<td>Depth (mm) Hardness (durometers, Shore A)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5  31  4  2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10 31 2 4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5  23, 27, 31 3 3, 3 3 3, 3, 2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10 23, 27, 31 3 3, 3 3, 3</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>5  31 3 2</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>10 31 3 2</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>15 31 3 2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5  31 2 3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10 23, 27, 31 4 4, 4, 4 4, 4, 2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15 23, 27, 31 3 3 3 3, 3, 2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>10 31 3 3</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>15 23, 27, 31 4 4, 4, 4 4, 4, 2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>17 31 3 3</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>10 31 3 3</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>15 31 3 3</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>15 31 3 3</td>
<td></td>
</tr>
</tbody>
</table>

* Stimulus combinations only presented in Session 3 and examined by four participants.
size/depth combinations in both substrates. In stiffer substrates (two top plots – 82 kPa), the decrease in detection rate with changes in the abnormality hardness is more evident. The detection rate decreased by 38% (Fig. 4a) and by 42% (Fig. 4b) when the hardness of the two abnormalities decreased from 31 to 23 durometers. In contrast for the 21 kPa simulated prostates, the detection rate exhibited a more gradual decrease (by 10% (Fig. 4c) and by 14% (Fig. 4d)) when abnormality hardness was decreased.

3.5. Variance in detection performance between examiners

Detection varied across stimulus combinations (abnormality size, depth, hardness and substrate stiffness) encountered by each participant and also between participants. The hit rate for each participant was defined as the number of abnormalities detected by the participant divided by the total number of abnormalities presented to that participant. Participants’ hit rates ranged from 95.8% to 48.5% (mean = 74.0%, SD = 13.0%), as indicated by the solid points in Fig. 5. Three (out of 18) participants were below one standard deviation of the mean hit rate. The distribution of participants’ hit rates is shown in Fig. 6.

Furthermore, all stimulus combinations that were detected more than 90% of the time (from Figs. 1 and 2) were classified as “easy” stimulus combinations. All stimulus combinations that were detected less than 90% of the time were classified as “difficult” stimulus combinations. Participants’ hit rates for “easy” stimulus combinations are indicated by the crosses (+) in Fig. 5, whereas hit rates for “difficult” stimulus combinations are indicated by the exes (x). The mean hit rate for all participants decreased (96.9% to 56.4%) while the standard deviation in hit rate increased (4.8% to 21.4%) from the “easy” to “difficult” stimulus combinations. Fig. 7 shows the hit rate distributions for these subsets of stimulus combinations.

The Pearson’s product-moment correlation model was used to determine that there was no significant correlation between participants’ hit rates and false alarm rates ($r = -0.174, p = 0.489$) for all stimulus combinations (Fig. 8).

4. Discussion

This study sought to characterize the range of prostate abnormalities that are palpable via digital rectal examination. Among the four factors (abnormality size, depth, hardness, and substrate stiffness), our overall finding was that at a depth of 5 mm, the majority of abnormalities can be detected

![Fig. 3. Psychophysical functions for the detectability of abnormalities 10 mm diameter at 10 mm depth as a function of abnormality hardness for prostate stiffness levels of 82 and 21 kPa. Standard error bars are shown around the data points.](image1)

![Fig. 4. Psychophysical functions for the detectability of four different size/depth combinations in two prostate stiffness levels as a function of abnormality hardness: (a) 5 mm diameter abnormality at a 10 mm depth embedded within prostate of stiffness 82 kPa, (b) 15 mm diameter abnormality at a 15 mm depth embedded within prostate of stiffness 82 kPa, (c) 5 mm diameter abnormality at a 5 mm depth embedded within prostate of stiffness 21 kPa, (d) 10 mm diameter abnormality at a 15 mm depth embedded within prostate of stiffness 21 kPa. Standard error bars are shown around the data points.](image2)
independent of prostate stiffness or abnormality hardness or size. However, abnormalities located more deeply must be larger to be consistently detected. This is especially pertinent for the stiffer substrates, where abnormalities must be 10–17 mm diameter to be consistently detected at depths 10–15 mm. In more pliant substrates, detectable abnormalities at depths 10–15 mm are smaller (5–10 mm diameter). These results may aid in the development of training regimens so that scenarios utilize reasonable ranges per the four factors. Additionally the between examiner variance may also have implications in the clinic for a starting point in placing expectations on detection performance.

In comparison to simulated breast lumps, we found that the size of prostate abnormalities had be larger to be consistently detectable (5 mm compared to 3 mm) [6,7]. Additionally, our study found that depth played a larger role in abnormality detection. These findings are most likely due to the differences in substrate stiffness (21 or 82 kPa compared to 1–5 kPa). Compared to size and depth, the hardness of an abnormality did not play as large a role, similar to results related to simulated breast tissue [6]. While hardness was not as prominent a factor overall, it did emerge as relevant when substrate stiffness and abnormality hardness were varied for the more deeply embedded (5–15 mm diameter and 10–15 mm depth) abnormalities. Detecting the hardness of an abnormality is most pertinent to detecting changes in prostate stiffness due BPH or prostititus, and therefore is relevant to larger and more deeply embedded tumors.

When we considered the performance of individual examiners, we uncovered that there were clearly a set of “easy” and “difficult” stimulus combinations. More effective training may need to be developed that reduces the number of “easy” training scenarios that yield high levels of detection performance and little variance between examiners. It is clearly the “difficult” scenarios that yield a lower level of detection performance and greater variability.

This experiment utilized simulated prostates in a simulated environment. The simulated prostates did not include fibrous tissue, varying surface texture, or undulations of the surface. Additionally, the experimental environment lacked any of the typical patient–clinician interactions. Ideally, future studies will augment our findings by using in vivo human prostate tissue with known abnormalities of various size, depth, and hardness. As a first
approximation, however, this study afforded a repeatable and controlled method and the results are valuable for gathering data on thresholds of detection and variance between examiners, similar to that done for the breast exam.

In relation, when analyzing the individual examiners, we also identified a low false alarm rate. This is likely because the simulated prostates did not include non-homogenous, fibrous tissue. The introduction of such elements would likely increase detection thresholds. That said, our false alarm finding does not indicate that better detectors are more willing to answer “yes” to improve their performance. Finally, participants were required to use the defined search technique. Other work indicates that search technique plays a role in detection performance [22] and should inform investigations, similar to that presented in here, to further quantify the impact of technique on improving the detection of smaller and deeper tumors.

Conflict of interest

None.

Acknowledgments

The authors would like to acknowledge the support and clinical guidance of Marcus L. Martin, M.D. (School of Medicine, University of Virginia) and Reba Moyer Childress, SN, FNP, APRN-BC (School of Nursing, University of Virginia). The research described was supported in part by Grant Number T15LM009462 from the National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.

Sources of support: (1) The Congressionally Directed Medical Research Program administered by the Department of the Army and (2) The National Library of Medicine T15 training grant: A Systems Engineering Focus on Medical Informatics

References


