BodyDiagrams: Improving Communication of Pain Symptoms through Drawing

Amy Jang  
Computer Science Dept.  
Stanford University  
Stanford, CA 94305 USA  
insunj@cs.stanford.edu

Diana MacLean  
Computer Science Dept.  
Stanford University  
Stanford, CA 94305 USA  
malcdi@stanford.edu

Jeffrey Heer  
Computer Science & Engineering Dept.  
University of Washington  
Seattle, WA 98195-2350  
jheer@uw.edu

ABSTRACT
Thousands of people use the Internet to discuss pain symptoms. While communication between patients and physicians involves both verbal and physical interactions, online discussions of symptoms typically comprise text only. We present BodyDiagrams, an online interface for expressing symptoms via drawings and text. BodyDiagrams augment textual descriptions with pain diagrams drawn over a reference body and annotated with severity and temporal metadata. The resulting diagrams can easily be shared to solicit feedback and advice. We also conduct a two-phase user study to assess BodyDiagrams’ communicative efficacy. In the first phase, users describe pain symptoms using BodyDiagrams and a text-only interface; in the second phase, medical professionals evaluate these descriptions. We find that patients are significantly more confident that their BodyDiagrams will be correctly interpreted, while medical professionals rated BodyDiagrams as significantly more informative than text descriptions. Both groups indicated a preference for using diagrams to communicate physical symptoms in the future.

Author Keywords
Health; pain diagrams; symptom communication; drawing.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: UI

INTRODUCTION
Millions of people use the Internet as a source of medical information. A 2013 Pew Survey on “Health Online” reports that one third of American adults use the Internet as a diagnostic tool [7]. While most users tend to read information contributed by others, thousands also initiate discussions around personal health conditions. A typical format is for a user to compose a textual description of her condition, and then post it on a forum such as MedHelp (www.medhelp.org) or PatientsLikeMe (www.patientslikeme.org) in an effort to garner feedback. Although people discuss a wide range of medical conditions online, here we focus on what we call physically-oriented conditions (POCs): conditions in which physical attributes of symptoms such as location and severity are a differentiating factor in diagnosis or treatment. Examples of POCs include chronic pain, digestive health conditions and musculoskeletal problems.

Text is often an ineffective communication medium for POCs. Consider the following example from MedHelp.com:

“hi i’ve had this awful sharp nervey pain in the middle of my back which shoots round to my ribs on the left side…”

Here the poster is describing back pain. However, the precise location of where the pain starts in the back and how it shoots to the rib is vague. As a result, respondents had a wide range of interpretations for this symptom, and the user had to clarify (again with text) the precise physical location of the pain. Even when users include highly detailed information on symptom location, the resulting description can be confusing:

“the pain is at the right side of my right knee, its situated near the fibula, not indicating that my bone is the one hurting, the pain is distinctive as it only hurts in a manner or form of a thread, the pain doesn’t take up all of the right part of my knee, its just a thread shape pain that seems to extend from the femur down to the fibula.”

Reading this description, the reader must translate the poster’s words into locations on her body to interpret and respond to the information. The overhead entailed in this parsing process might not only damage the reader’s interpretation of the condition, but also affect the type and quality of response given.

In addition to being a poor expression of physical location, text can impose other shortcomings: users may have an insufficient vocabulary for describing symptoms, or poor writing skills. Moreover, a single symptom can be described in several ways without using shared terminology. This last issue may complicate searches for related posts.

In contrast, drawing provides a natural way of directly indicating physical locations. Indeed pain diagrams, such as the ones used in Figure 1, are commonly used in medical practice. In addition to location, symptom properties such as severity or depth also lend themselves to a visual encoding.

We offer two primary research contributions. First, we present BodyDiagrams, an online interface for expressing...
symptoms via drawing and text. BodyDiagrams is based on insights distilled from multiple rounds of design iteration, and synthesizes the benefits of text (flexibility, high resolution) with those of drawing (precise location and directness). With BodyDiagrams, users can visually encode symptom attributes such as severity, frequency and temporal sequence. BodyDiagrams is available at www.mybodydiagram.com.

Second, we conduct a two-phase user study with both patients and medical professionals in which we assess the efficacy of BodyDiagrams as a communication medium compared to plain text. Patients regard their BodyDiagrams descriptions as significantly more interpretable, while medical professionals regard BodyDiagrams as significantly more complete and relevant. We also find ordering effects in description content and quality. Our results help validate our design decisions and provide insights for future deployments, including use of an iterative description process to further improve medical communication.

RELATED WORK

In everyday medical practice, pain questionnaires with diagrams are sometimes used as self-report tools for patients to communicate with their doctors [12]. In these questionnaires, users select a characteristic word, location, severity and time information to describe their pain.

Researchers have conducted studies on the utility of pain diagrams as screening or diagnostic tools. Results vary, and may depend on the condition being screened for. Jamieson et al. [9] studied whether experts could distinguish between healthy individuals with pain and individuals suffering from chronic pain using only pain diagrams. They found that experts performed only slightly better than random, while a decision model based on the area covered by the drawings achieved 70% accuracy. Fortin et al. [6] achieved no false positives in using pain drawings to screen for sacroiliac joint dysfunction; they were not able to measure false negatives.

Other literature uses pain diagrams to discover and define regions of referred pain cased by joint inflammation. Dreyfuss et al. [4] extracted consistent pain referral patterns from subjects whose thoracic zygapophyseal joints were stimulated via intra-articular injection. A similar but separate study by Dreyfuss et al. [3] found consistent referral patterns stemming from lateral atlanto-axial inflammation, while atlanto-occipital inflammation resulted in varying referral patterns. The authors conclude that pain drawings might be useful for differential diagnosis.

Ghinea et al. [8] propose a GIS-based system for digitizing hand-drawn back pain diagrams, motivated by the desire to search for symptoms on particular bodily regions. However, their system includes no interface for direct user input. Felix et al. [5] conducted a study to determine whether computerized pain drawings by patients undergoing spinal surgery could be used to automatically evaluate recovery. Their preliminary studies supported the feasibility of this approach.

Spyridonis and Ghinea [17] found that patients with spinal cord injuries found 3D diagrams a more expressive medium for describing their pain symptoms than 2D diagrams.

In the HCI research space, several tools have been developed to aid physician interpretation of clinical histories, the most notable being the LifeLines project [14, 15] and Bui’s TimeLine [2]. Both projects integrate HCI principles to visualize patient records for medical professionals. However, they are designed as browsing tools rather than communication aids.

Finally, several visual pain trackers are available as downloadable apps or as online services on sites like MedHelp [11] and Partners Against Pain [13]. WedMD’s visual symptom checker allows users to explore symptoms by clicking affected body parts on a model [18]. However, these tools are limited in expressivity, and are designed either for self-tracking or dispensing information rather than communication.

To our knowledge, Body Diagrams is the first tool developed through researching user needs in this space, and is unique in its synthesis of visual symptom expression with traditional textual communication.

DESIGNING BODYDIAGRAMS

Our overarching goal in developing BodyDiagrams was to improve digital communication of physically-oriented conditions by leveraging natural visual relationships between symptoms and their attributes. To this end, we anchored our process around two design goals:

Support natural symptom-drawing patterns. To facilitate intuitive authoring of POC descriptions, we wanted BodyDiagrams to support drawing and annotation patterns expressed by users drawing symptoms with pen and paper.

Enhance interpretability through a synthesis of increased content and compact representation. To support accurate interpretation of physical symptom descriptions, we sought to incorporate more relevant content and represent it visually.

In this section we discuss our design process, summarizing our pilot study [10] and describing our iterative testing via Amazon’s Mechanical Turk, and present a set of resulting design insights. We then present the BodyDiagrams interface design.

Figure 1: Sample hand-drawn diagrams from our pilot study, drawn on a traditional pen-and-paper pain diagram.
Table 1: Drawing patterns observed in our pilot study.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Example</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td><img src="image1.png" alt="Image" /></td>
<td>rarely utilized; most participants used only one color even when describing a range of symptoms.</td>
</tr>
<tr>
<td>precise marks</td>
<td><img src="image2.png" alt="Image" /></td>
<td>indicates precise symptom location</td>
</tr>
<tr>
<td>regions</td>
<td><img src="image3.png" alt="Image" /></td>
<td>indicates general location of symptom</td>
</tr>
<tr>
<td>text</td>
<td>captions</td>
<td>describe drawn marks</td>
</tr>
<tr>
<td>arrows</td>
<td><img src="image4.png" alt="Image" /></td>
<td>link text annotations to drawn marks</td>
</tr>
<tr>
<td>zoom</td>
<td>call outs with more detail</td>
<td>users drew scaled versions of body parts for higher resolution</td>
</tr>
<tr>
<td>views</td>
<td>side-view, cross-section</td>
<td>users drew body parts in different perspectives</td>
</tr>
<tr>
<td>reference marks</td>
<td>vertebrae, knee cap</td>
<td>users drew body parts as reference “anchors” for the relative location of their symptoms</td>
</tr>
</tbody>
</table>

Table 2: BodyDiagrams encodings for symptom attributes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Severity</th>
<th>Frequency</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td><img src="image5.png" alt="Image" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pilot Study

To better understand how people conceptualize and communicate symptoms, we first conducted a pen-and-paper pilot study. Our pilot study comprised 8 participants (4 female, 4 male, most early 20’s) recruited through an open email call at our university. We provided participants with a paper body diagram template and several colored pens. We prompted them to “describe a physical symptom [she/he] experienced recently”. We interviewed each participant upon completion. Each session took about 45 minutes, and participants received a $10 Amazon Gift Card.

Figure 1 shows a selection of pilot study diagrams. We observed several commonalities (summarized in Table 1) as well as procedural patterns, which we discuss below.

Users draw accurately. Both location and shape of drawn marks is intentional and accurate. Users often physically referred to their own body (by touching the area in question) before drawing the corresponding mark on paper.

Users draw, link, describe, and repeat. When recording a symptom, users would first draw it (sometimes utilizing multiple lines or marks), then create a reference line from the mark to white space, and finally describe the symptom in text. Then, they would move on to draw then next symptom. This procedural pattern was uniform across participants.

Users draw the most severe symptom first. Though most users’ symptoms had chronological structure, users drew the most severe symptom first and then “filled in” secondary symptoms, regardless of chronological order.

Users tend not to iterate, but iteration provides better results. Only one user drew a “draft” description before drawing a final one. He noted that while the first drawing contained the relevant information, the second one called out important details and was superior. Other users commented that categorizing symptoms without drafting them first was difficult.

Iterative Design Process

Based on our pilot study observations, we developed a prototype BodyDiagrams interface. We then followed an iterative design process including evaluations with Amazon’s Mechanical Turk (MTurk). Over four major interface revisions, we recruited 10-12 MTurk workers (Turkers), gave them a synthesized pain description, and asked them to replicate it using BodyDiagrams. Turkers then completed a post-task survey and were paid $1.00. By fixing the input condition, we were able to observe the effect of new features on the BodyDiagrams descriptions and compare them across iterations.

Design Considerations

Over the course of iterative design, we distilled a set of key design insights for effective visual communication of symptoms. We summarize these below, highlighting supporting observations from our design process.

DC1: Implement both freehand and regional drawing tools. As drawing accurately with a mouse is difficult, our initial designs favored a purely shape or palette-based annotation approach. In our pilot study, however, users utilized regional marks in concert with free-form marks. The latter were used primarily when the precision of a symptom’s location or shape was key to the description. This behavior was sustained in our online MTurk studies.

DC2: Support the draw, link, describe, repeat cycle. This cycle (described in our pilot study observations) embodies a natural procedural framework for how users develop visual symptom descriptions. Transitioning through the cycle’s stages should be frictionless so as to allow the user to focus her attention on her description. In earlier versions of BodyDiagrams, which supported this process poorly, users reported higher levels of frustration. In later versions, Turkers utilized this feature seamlessly, without prior tutelage.

DC3: Group consecutive marks into a single symptom. It may take several marks to describe one symptom (e.g., a rash or a series of bug bites). We observed that users drew these compound symptoms consecutively before moving to annotate them. A natural way to extract compound symptoms and follow DC2 is to automatically group consecutive marks.

DC4: Structure symptom description input to capture inherent attributes. Formal pain questionnaires, such as the McGill questionnaire [12], often ask patients to describe specific attributes of their pain symptoms. These include severity (typically scored on a horizontal, 1-10 scale [12]), descriptive characteristics (e.g., flickering, burning, radiating), location,
depth and frequency. Table 2 illustrates how we mapped these features onto visual encodings in the annotation box. This was well received by participants: in comparison with their responses from prior iterations, Turkers reported that they felt their BodyDiagrams were a more complete, expressive, and useful complement to the prompt text. Structuring these attributes also resulted in a decrease in the quantity of text used to describe marks (194.5 vs. 226 characters on average).

DC5: Represent symptoms chronologically. In our pilot study we observed that users listed symptoms out of chronological order, starting with the most severe symptom. However, in all our studies, users still indicated chronological relationships between symptoms, sometimes explicitly labeling their annotations with numbers.

DC6: Use a realistic body diagram. In our pilot study we noticed that cartoonish diagrams are both difficult to identify with and also lack important physical detail. This led users to draw in reference lines (one user drew all the vertebrae in his back) as contextual anchors. To facilitate grounding and reference-finding, we introduced two accurate 3D body models (one male, one female) about halfway through our design process. The new diagrams were received positively. Average frustration decreased, and Turkers reported feeling comfortable with the model. One user commented, “I really liked the model, it was easy to see the bones and the muscles”.

DC7: Encourage users to contribute difficult-to-draw information. Some important information, such as medical history, medications and dosages, do not lend themselves to drawing. In early versions of BodyDiagrams, we noticed that Turkers omitted these details or shoehorned them into drawing annotations due to the lack of anywhere else to put them.

DC8: Support detailed zooming. Pilot study users vocalized frustration at not being able to zoom into the diagram, noting that describing symptoms present on small bodily regions, such as a finger, is difficult. Some frustrated users even drew zoomed-in versions of their own.

THE BODYDIAGRAMS INTERFACE

Figure 3 shows a screenshot of BodyDiagrams being used by a patient to describe exercise-related muscle strain. We discuss key features (annotated with letters A-I) and how they instantiate our design considerations.

A) Rotation Panel: Clicking on a thumbnail rotates the main model. Drawings from each view are summarized on each respective panel, preserving context across view switches.

B) Toolbox: The toolbox provides a variety of mechanisms for authoring BodyDiagrams descriptions: undo, drag, pen (freehand) drawing, region drawing, and region fill. Supporting freehand drawing allows users to denote regions and shapes with accuracy when necessary (DC1).
C) Pan and Zoom: Users can pan the model to a region of interest, and zoom to a desired level of detail (DC8).

D) Body Model / Drawing Canvas: Users draw on the model using a mouse or trackpad. We created an anatomically accurate model using images from materials for medical students. The model includes both internal and external bodily reference points for drawing (DC6).

E) Annotation Widget: This widget pops up after a mark is drawn, offering the user a place to input structured symptom metadata (DC4). If the user draws multiple marks before entering data into the widget, these marks are grouped into a single symptom (DC3). Upon clicking Save, the annotation is summarized in the summary box (F). This interaction process supports iterative authoring of symptom descriptions (DC2).

F) Summary Box: A summarized representation of prior annotations. Clicking on this box opens it for editing.

G) Symptom Timeline: Horizontal positioning of symptom summaries on the timeline denotes symptom sequence and duration; vertical positioning matches, as closely as possible, the vertical position of the summary’s corresponding drawn mark. Users can drag symptom summaries left or right to indicate their relative starting time, and resize them horizontally to indicate their duration, supporting chronological ordering of symptoms regardless of the order in which they were drawn (DC5). In the figure shown, the user has indicated that he experiences knee pain which starts before and ends after a muscle spasm on her left leg.

H) Global Textbox: Allows users to provide additional information that may be difficult to draw, such as medical histories and medications used (DC7).

I) Help: Clicking on the help icon overlays a tutorial.

BodyDiagrams is a web application built using Ruby on Rails. The annotation UI is written in JavaScript using D3 [1], with graphics rendered using SVG. A server tracks each session state and persists user data.

User Scenario
We illustrate a sample use of BodyDiagrams through a hypothetical user Mark, who is a composite drawn from real examples. Mark has been experiencing severe back pain and occasional muscle strain in his legs. He is wondering whether these are related, and whether there is anything he can do to alleviate his symptoms.

Upon opening BodyDiagrams and selecting the model gender, Mark sees an overlay tutorial indicating how to get started with the tool (Figure 3a). After reading through it, he clicks on the close button in the right hand corner, and the overlay disappears. Mark starts with his most severe symptom: his back pain. He switches to the posterior view of the model using the rotation panel. Mark touches his back to confirm that the pain occurs precisely between his tailbone and L4. He zooms into the model: the higher resolution allows Mark to precisely locate the regions of interest on the model, because he can see the model’s skeletal structure lightly expressed beneath the skin.
Mark’s pain radiates to the right, and so he uses the pen tool to draw an asymmetric region where it hurts the most (Figure 3b). As soon as he finishes the drawing, the annotation widget pops up in the symptom timeline panel. It is connected via a reference line to his drawing. He fills in additional details about his back pain, clicks save, and sees a summary of his annotation on the timeline.

Next, Mark wants to express his leg pain. As he navigates to the right profile view, the reference lines for his pain symptom fade, and the summary box indicates which view the symptom mark appears on. This time he uses the region tool to define a general area in the upper thigh. He wants to indicate shooting pain down his leg, so with the annotation widget open, he defines another region in the calves area. He now sees that the two drawings are grouped as one symptom; the UI prompts him to ungroup if he wishes (Figure 3c). He wants them to be grouped, and so he proceeds to add details in the annotation widget.

Having described both of his symptoms, Mark recalls that he experiences the leg pain only after the back pain starts. The back pain often continues for hours, but the leg pain lasts longer. He drags the summary boxes to indicate their relative start time on the timeline (Figure 3d), and resizes them to indicate their duration. Satisfied, Mark finally uses the general text box to describe the remaining relevant information about his condition and to ask for feedback.

EXPERIMENT 1: PATIENTS
In our first experiment, we sought to evaluate BodyDiagrams’ efficacy as an authoring tool for patients trying to describe physical symptoms online. Based on our design process observations, we developed the following hypotheses:

H1 (Recall): Patients will produce more comprehensive condition descriptions using BodyDiagrams than with text only.

H2 (Resolution): Patients will describe symptoms in finer detail using BodyDiagrams than with text only.

H3 (Effectiveness): Patients will have higher confidence in their BodyDiagrams descriptions being correctly interpreted by readers than their text only descriptions.

H4 (Preference): Patients will express a preference to use BodyDiagrams to describe physical conditions in the future.

Participants
We recruited 21 participants (7 male, 14 female) via an open email call at our university, asking them to sign up for the study only if they had experienced physically-oriented symptoms which they might seek advice about on an online health forum. To encourage sufficiently complex conditions, we gave musculoskeletal and digestive system disorders as example conditions, and sprained ankles and bug bites as non-examples. We informed participants that the study would take place entirely online, be anonymous, and that their contribution would be shared with medical professionals. Subjects received a $20 gift card for participation.

Methods
The experiment followed a within-subjects design with interface presentation order randomized. Upon beginning the study participants were presented with the following prompt:

Please recall a set of physical symptoms that you have recently experienced. Imagine that you want to solicit feedback from people who have experienced a similar condition. On the next page, please describe your symptoms - and any related information - with the hypothetical goal of sharing this description with an online community in order to get feedback.

After reading the prompt, subjects used either the BodyDiagrams or text-only interface. After completing their pain description, subjects were given the same prompt and asked to describe their symptoms again using the alternative interface. Finally, they completed a survey. Throughout the study and recruitment phases, we referred to the interfaces as “text-only” and “drawing+text” in order to avoid branding effects.

Quantifying information and detail overlap
To quantify information overlap between BodyDiagrams and plain text descriptions, we read through each BodyDiagram and identified any piece of information (symptom, medical history item, etc.) that was missing from the corresponding text description, and vice versa. For example, if a user mentioned taking Tylenol in their text description, but there was no mention of Tylenol in their BodyDiagram, we counted this as an additional piece of information in the text description.

We repeated this process for detail asymmetry, considering only shared information between the descriptions (so as not to double-count information overlap). For example, if both descriptions mentioned stomach pain, but only the BodyDiagrams description mentioned the pain duration, we counted this an an extra detail in the BodyDiagrams description. We disregarded the quality of detail in this analysis. For example, if both descriptions mentioned the physical location of a symptom, we did not credit BodyDiagrams with extra detail simply because it renders such details in higher resolution.

Results
Three users described conditions that were not physically-oriented, leaving 18 participants (6 male, 12 female). Unless otherwise noted, all survey responses were on a 5-point Likert scale (1=Strongly Disagree, 5=Strongly Agree), with significance determined by Wilcoxon signed-rank tests (test statistic in tables denoted V) with paired samples, or Wilcoxon rank-sum tests otherwise (test statistic in tables denoted W).

BodyDiagrams rated as more likely to be correctly interpreted
Users were significantly more confident that readers will correctly interpret both their individual symptoms as well as their overall condition using their BodyDiagrams description (Table 3). Ordering effects are not significant in this result.

Each description comprises some exclusive information
As shown in Table 4, both description types comprise, on average, one piece of information not mentioned in their counterpart description. While information quantity differences were not significant, we noticed that medical histories and
treatments are more often included in text descriptions, while symptoms are more often included in BodyDiagrams.

**BodyDiagrams descriptions are more detailed**

On average, BodyDiagrams descriptions contain 0.9 more detailed facts than their plain text counterparts (Table 4), a statistically significant difference. Despite this, in self-report data users were neutral about which description was more detailed (mean=3.2, median=3.0, sd=1.0 on a 5-point Likert scale where 1=PT and 5=BD). Segmenting users into those who prefer BodyDiagrams (rating ≥ 4, 9/18 subjects) shows no significant preference (p = 0.5, one-tailed binomial test).

**BodyDiagrams take more time and contain less text**

Description length (in characters) and time to completion are shown in Table 7. We replaced one outlying value (1,347 seconds in the text condition) with the group mean; we believe the subject temporarily abandoned the task. On average, users took about 100 seconds longer using BodyDiagrams, and wrote about 110 fewer characters.

**Second descriptions are rated more complete and relevant**

Table 7 shows users’ estimates of information asymmetry in each description. While the difference is not significant for either interface, comparison of results by interface presentation order shows a significant effect: users report that their second description contains more information than their first. Users also reported agreement on whether each description contains all the relevant information about their condition, as shown in Table 8. Again, there is no significant difference between interfaces, but a significant ordering effect is present.

**Users prefer BodyDiagrams descriptions**

Users expressed a preference for using the drawing+text interface in the future (mean=3.8, median=4.0, sd=0.9). Segmenting users into those who prefer BodyDiagrams (rating ≥ 4, 13/18 subjects) and others, we find a significant preference for BodyDiagrams (p < 0.05, one-tailed binomial test).

**Qualitative User Feedback**

In commenting on the strengths and weaknesses of each interface, reported advantages of text include its expressiveness, flexibility, familiarity, and the fact that it can pressure users to be precise. Cited advantages of BodyDiagrams include the ability to pinpoint or highlight regions of concern (“in my opinion, the drawing+text interface has a clear advantage over the text-only interface because sometimes it is hard to describe and locate a specific part/point of human body.”), that it provides a template for symptom details (“It helped me think about each distinct symptom, the timing of the symptoms, and about measuring those symptoms in ways I wouldn’t have otherwise.”), and that it facilitates comparisons (“It is easier to compare my pain to others by looking at a picture instead of having to read everyone’s comments.”). However, some users did not feel that the benefits of text were fully integrated into BodyDiagrams. As one user said, “If we could combine the two, I would vastly prefer that. I also would like a way to aggregate all the data and information so that I can compare data points between at least 3 people to gain a better understanding of my physical symptoms and what I can do to relieve the pain.” Another stated, “I think it will be really useful to have the drawing interface in addition to the text interface, so long as the information is not lost due to the ‘easiness’ of providing the drawings.”
Discussion

H1 (Recall): Patients will produce more comprehensive descriptions of their condition using BodyDiagrams than with text only. We indeed find that users provide more details when using BodyDiagrams—roughly one piece of information per condition. Our results also indicate an ordering effect, as users perceive improvements in their description after refining it. This finding agrees with our pilot study observation that iteration provides better results. As users include different types of information in each interface, it may be helpful to cue users to include information they are likely to omit.

H2 (Resolution): Patients will describe symptoms in finer detail using BodyDiagrams than with text only. Users included significantly more detail in their BodyDiagrams than in plain text descriptions (Table 4). However, differences in users’ opinions regarding detail were not significant. We attribute the increase in detail to BodyDiagrams prompting users to record attributes like location, severity and frequency.

H3 (Effectiveness): Patients will have higher confidence in their BodyDiagrams descriptions being correctly interpreted by readers than their text only descriptions. Users are significantly more confident in the interpretability of their BodyDiagrams (Table 3) than their text only descriptions.

H4 (Preference): Users will prefer to use BodyDiagrams for describing physically-oriented conditions. In addition to commenting positively on BodyDiagrams, users significantly preferred to use a drawing + text based interface in the future for communicating physically-oriented conditions.

EXPERIMENT 2: MEDICAL PROFESSIONALS

In our second experiment, we evaluated BodyDiagrams for communicating patients’ descriptions to medical professionals online. We posed the following hypotheses:

H5 (Interpretation): Professionals will have a clearer understanding of conditions described with BodyDiagrams than conditions described with plain text.

H6 (Confidence): Professionals will feel more confident in their advice given in response to BodyDiagram descriptions than to plain text descriptions.

H7 (Content): The information content of professionals’ responses to BodyDiagram descriptions will be higher than that given in response to plain text descriptions.

Participants

We recruited 10 medical professionals (5 licensed doctors, 3 registered nurses, 1 community pharmacist and 1 4th-year med student) through oDesk (www.odesk.com). We screened subjects for frequent patient interaction experience and English skills. Subjects were paid $5 for each assessment.

Methods

Although a direct comparison between BodyDiagrams and plain text descriptions was desirable for this experiment, a within-subjects design presented several difficulties. Foremost was the problem of information transfer: regardless of the order in which experts read corresponding text descriptions and BodyDiagrams, they would have strictly more information about the patient’s condition upon reading the second description. We were concerned that this would impede subjects’ ability to compare the interfaces objectively. Alternatively, we might have presented experts with non-corresponding BodyDiagram and plain text descriptions. In this case, however, we were concerned that subjects would not have a clear basis for comparison.

We instead used a between-subjects design with interface as the independent variable. Each professional evaluated several patient descriptions using only one of the interfaces (BodyDiagrams or text). 18 BodyDiagrams and 18 text descriptions from Phase 1 were each evaluated by 2 medical professionals. Upon starting an assessment, participants were shown a patient’s description and asked to “respond with advice [they] would give this patient” in an adjacent text box. After each assessment, participants completed a post-task survey.

Results

Survey responses were on a 5-point Likert scale (1=Strongly Disagree, 5=Strongly Agree). To account for non-normal responses and patient variability, we analyzed responses using a non-parametric Friedman rank sum test, in which plain text vs. BodyDiagrams was the condition and patients the block.

BodyDiagrams rated as more focused than plain text

Tables 9 and 10 show subjects’ ratings of completeness and relevance of the descriptions that they evaluated. Subjects evaluating plain text descriptions felt significantly more strongly that their descriptions were both lacking relevant information and contained extraneous information.

Subjects are confident in their advice, regardless of interface

We asked subjects to report their confidence in overall diagram interpretation, individual symptom interpretation, diagnosis given, correctness of advice given, and helpfulness of advice given. Differences between interface conditions were not significant. Scores were uniformly high, suggesting that medical professionals are highly confident in their interpretations and contributions regardless of presentation interface.

No time or length effects

Both subject groups reported reading the description twice on average before responding, and referring back to the description about twice while writing. Differences in response time

Table 9: Response to the question “I felt that this description was missing information relevant to the user’s condition.”

<table>
<thead>
<tr>
<th>Description</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>IQR</th>
<th>χ²(2)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD</td>
<td>2.40</td>
<td>1.30</td>
<td>2</td>
<td>1.50</td>
<td>12.25</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>PT</td>
<td>3.40</td>
<td>1.10</td>
<td>4</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10: Responses to “I felt that this description contained information that was not relevant to the user’s condition.”

<table>
<thead>
<tr>
<th>Description</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>IQR</th>
<th>χ²(2)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD</td>
<td>1.60</td>
<td>0.50</td>
<td>2</td>
<td>0.00</td>
<td>4.57</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>PT</td>
<td>2.00</td>
<td>0.80</td>
<td>2</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
are not significant (mean = 1857 seconds (PT) vs. 1181 (BD), median = 1135 vs. 1054, $\chi^2 = 0.9, p = 0.34$); neither are differences in response length (mean = 591 characters (PT) vs. 999 (BD), median = 749 vs. 457, $\chi^2 = 2, p = 0.16$).

Response Content
We analyzed the content of medical professionals’ responses in order to get a sense of typical response components. Overall, subjects gave detailed responses in which they typically discussed possible causes of the symptoms, and presented potential diagnoses, both often conditional on an aspect of the user’s health omitted from the description (e.g., smoking history). In addition, they often gave suggestions for alleviating symptoms and advice on whether to seek proper medical attention. Finally, medical professionals often utilized repair dialogue to indicate missing description components. Repair dialog is characterized by both direct questions such as “how many hours do you work?” and indirect comments such as “if you have been sick recently then ...”, implying that inclusion of such information would have been helpful.

To measure content difference between conditions, we manually coded the number of questions (mean = 1.0 (PT) vs. 0.8 (BD), median = 0 vs. 0, $\chi^2 = 0.06, p = 0.81$), pieces of repair dialogue (mean = 1.7 (PT) vs. 1.5 (BD), median = 2 vs. 0, $\chi^2 = 0.0, p = 1.0$) and suggestions given (mean = 4 (PT) vs. 3.6 (BD), median = 3 vs. 3, $\chi^2 = 0.25, p = 0.62$) for each response. None of these differences are significant.

Qualitative Feedback
We asked medical professionals who had evaluated BodyDiagrams descriptions to submit free-form feedback. All participants responded positively. One user wrote: “[BodyDiagrams] clearly points out the affected site of patient so because of this, diagnosis becomes very clear. Now the diagrammatic description helps a lot because the risk of misunderstood medical conditions is minimized.” Another indicated that she felt more empathy towards the patients, as “[BodyDiagrams] helps you to easily imagine your patient...”. Furthermore, users said that BodyDiagrams provide more detailed descriptions: “The image + description concept is great as it gives medical professionals more detailed knowledge about the patient’s condition.” Finally, one user compared BodyDiagrams to text directly, saying “I think, it is far better than [a] text description. The way of showing the pain area and symptom timeline was most useful to me.”

Suggestions for improvement included incorporating a general patient information box for data like age, BMI, and weight, as well as incorporating existing assessment tools such as the PQRST (Provocation Quality Region and radiation Severity Time) [16] method for pain assessment.

Discussion
H5 (Interpretation): Professionals will have a clearer understanding of conditions described with BodyDiagrams. While confidence in interpretation is no different between interfaces, qualitative feedback from professionals supports the claim that BodyDiagrams aids condition interpretation. These comments specifically mention the ease of identifying symptoms’ physical locations and the symptom timeline view as helpful. In addition, subjects rated BodyDiagrams descriptions as significantly more complete and relevant (Tables 9 and 10).

H6 (Confidence): Professionals will feel more confident in their advice for conditions described using BodyDiagrams. We found that professionals have high confidence in their advice regardless of presentation interface. One possible explanation is that subjects were reluctant to admit uncertainty. Another is that the interface does not affect confidence of response, because experts account for uncertainty in the advice that they give. In either case, these results underscore the importance of patients creating complete and accurate descriptions of their conditions.

H7 (Content): The information content will be higher in professionals’ responses to BodyDiagrams descriptions. We found no differences in content between professionals’ evaluations of BodyDiagrams and those of plain text descriptions.

DISCUSSION AND FUTURE WORK
Our results demonstrate BodyDiagrams’ perceived efficacy as a communication tool from the perspective of both patients (Experiment 1) and medical professionals (Experiment 2). In this section, we revisit our original design goals and discuss how these results relate to our design considerations.

Support natural symptom-drawing patterns. The ease with which patients expressed their symptoms using BodyDiagrams suggests that supporting freehand drawing (DC1) and structuring input flow via the draw-link-describe-repeat pattern (DC2) enhances BodyDiagram’s ease of use. Moreover, the introduction of grouping multiple consecutive marks into single symptoms (DC3) triggered a notable decline in users’ frustration levels with the tool. However, supporting natural drawing behaviors does not necessarily entail mimicking their precise pen-and-paper mechanisms. Our symptom timeline (based on DC5), for example, leverages affordances of a digital interface (easy repositioning and resizing) to render chronology in a manner that would be difficult to draw on paper. Similarly, after incorporating a realistic diagram (DC6) users stopped spending time drawing in reference lines on the diagram. To this end, we propose that natural drawing behaviors be used as guidelines, but that implementations of symptom-drawing interfaces should extend interactions beyond those available using pen-and-paper.

Enhance interpretability through a synthesis of increased content and compact representation. Medical professionals rated BodyDiagrams descriptions are more informative and relevant than their textual counterparts. BodyDiagrams naturally capture physical content and context that is difficult to record and interpret via text. In addition, acquiring inherent symptom attributes via structured input boxes (DC4) helps to standardize symptom descriptions and capture information that might otherwise have been omitted. However, there is a tradeoff between structuring data collection and maintaining expressiveness. While we felt that symptom input was only lightly structured (and included a free-form text field), one participant didn’t like the symptom depth categories; another “wanted there to be an ‘inner’ area, vs. skin/joint/muscle,
for [her] symptoms (abdominal pain).” Moreover, while we anticipated that users would write difficult-to-draw information in the general text box (based on DC7), in our opinion this feature was underutilized. Prior to a large-scale deployment of BodyDiagrams, additional thought should be given to structuring and standardizing information on the one hand, and maintaining flexibility on the other. Nonetheless, our experiments find that BodyDiagrams lead to more detailed patient descriptions that are rated more informative by medical professionals, suggesting that BodyDiagrams achieves a productive balance of compactness and descriptiveness.

Limitations and Future Work
High priorities for future work include better synthesizing text input into BodyDiagrams (despite the inclusion of the general text box, users still noted that the diagram distracted them from entering other information), incorporating more realistic and racially-neutral body models, and the introduction of iterative symptom description into the interface workflow. While our results showed general user preference for BodyDiagrams, differences in content quantity and quality point to strong ordering effects: second descriptions are almost always better. To this end, we propose that facilitating easy symptom “drafting” prior to final authoring would result in significantly better descriptions. Moreover, we are eager to evaluate BodyDiagrams in a collaborative environment, with back-and-forth communications between readers and patients.

We also plan to deploy BodyDiagrams on mobile devices that support drawing via touch input. Drawing with a mouse can be difficult; we believe that touch may provide a better input modality. Finally, BodyDiagrams can be used for visual indexing of symptoms, enabling effective aggregation and search over common pain patterns.

ACKNOWLEDGMENTS
We thank the subjects who volunteered to participate in our experiments, and the Division of Clinical Anatomy at Stanford University for providing 3D body models. This research was supported by Sloan Foundation Fellowship.

REFERENCES