

Crowdsourcing Treatments for Low Back Pain

Simo Hosio¹, Jaro Karppinen², Esa-Pekka Takala², Jani Takatalo³,
Jorge Goncalves⁴, Niels van Berkel⁴, Shin'ichi Konomi⁵, Vassilis Kostakos⁴

¹Center for Ubiquitous Computing, University of Oulu, Finland

²Finnish Institute of Occupational Health, Finland, ³Oulu University Hospital, Finland

⁴School of Computing and Information Systems, The University of Melbourne, Australia

⁵Faculty of Arts and Science, Kyushu University, Japan

{simo.hosio, jaro.karppinen, jani.takatalo}@oulu.fi, esa-pekka.takala@ttl.fi,
{jorge.goncalves, niels.van, vassilis.kostakos}@unimelb.edu.au, konomi@artsci.kyushu-u.ac.jp

ABSTRACT

Low back pain (LBP) is a globally common condition with no silver bullet solutions. Further, the lack of therapeutic consensus causes challenges in choosing suitable solutions to try. In this work, we crowdsourced knowledge bases on LBP treatments. The knowledge bases were used to rank and offer best-matching LBP treatments to end users. We collected two knowledge bases: one from clinical professionals and one from non-professionals. Our quantitative analysis revealed that non-professional end users perceived the best treatments by both groups as equally good. However, the worst treatments by non-professionals were clearly seen as inferior to the lowest ranking treatments by professionals. Certain treatments by professionals were also perceived significantly differently by non-professionals and professionals themselves. Professionals found our system handy for self-reflection and for educating new patients, while non-professionals appreciated the reliable decision support that also respected the non-professional opinion.

Author Keywords

Crowdsourcing; Health Information; Decision Support Systems; Information Discovery; User Study.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Crowdsourcing offers means to distribute large volumes of work into smaller, easy to complete chunks to a remote labour force. While the crowdsourcing tasks typically found in online labour markets are fairly simple and repetitive, increasingly also more challenging work is being pitched to

the crowds [18,23] — often by orchestrating the workers by using intermediary collaboration solutions [10]. One example of using crowdsourcing in more complex work is offering real-time health information [3,12].

Our work is also contextualised within the health domain, where people are increasingly seeking help on the Internet [7,16,20]. We focus on a chronic medical condition: low back pain (LBP). LBP has been recently estimated to cause more disability globally than any other condition [14]. It was also responsible for over 20 million disability-adjusted life years in the US alone in 2010, and its age-standardised prevalence was found highest in Western Europe [14]. Further, LBP has no silver bullet solutions to it, and the lack of therapeutic consensus causes a practical problem in choosing treatments to try [26]. Therefore, exploring systems to support people suffering from LBP is warranted.

Crowdsourcing has been identified as helpful in collecting and assessing potential answers to a given problem [6]. In our work, we developed *Back Pain Workshop* — a crowdsourced online system that helps people suffering from LBP answer the question “*what is a good way to treat my low back pain?*”. Our lightweight decision support system stores *structured data* [9] on LBP treatments and aggregates the collected data to rank and offer treatments using the principles behind *wisdom of the crowd* [25].

Using *Back Pain Workshop*, we collected two knowledge bases: one from clinical professionals — medical doctors and physiotherapists specialised in back pain — and one from *non-professionals* (everyone else). This is a crucial aspect of the work. Understanding the patients’ experiential expertise in contrast to the medical experience of professionals is seen as highly desired in the age of unguided advice on social services online [11,16]. The two knowledge bases consist of the assessments for all the collected LBP treatments in regard to four criteria that our domain expert collaborators identified as the most suitable ones to explore. This means collecting a subjective “goodness” value from multiple respondents to each treatment-criterion pair using a value between 1-100. The same scale-based assessment model has been validated as intuitive in capturing subjective knowledge on multiple-choice questions [9,13].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada
© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-5620-6/18/04...\$15.00

Our work answers the following research questions:

RQ1a: Are treatments crowdsourced from professionals perceived differently by professionals themselves and non-professionals?

RQ1b: How do non-professionals perceive treatments crowdsourced from non-professionals in comparison to treatments crowdsourced from professionals?

RQ2: What kind of added value can such crowdsourced knowledge on back pain treatments offer to the two relevant stakeholders: professionals and non-professionals?

Our results show that non-professionals perceive the best treatments by non-professionals to be as high quality as the best treatments by professionals. However, the worst treatments by non-professionals are clearly worse than the worst by professionals, from the non-professional's viewpoint. Further, certain treatments by professionals are clearly rated differently by professionals and non-professionals. Non-professionals find value in the tool offering trustworthy (professional) information that does not dismiss their own opinions. Professionals, on the other hand, can benefit from being able to self-reflect based on the collected data and potentially in discussions with non-professionals in clinical work.

RELATED WORK

Conceptually Back Pain Workshop is located at the intersection of crowdsourced decision support systems and health information discovery. A key distinction between related work and our case is that we allow the users choose the knowledge source: professionals or non-professionals.

Non-professionals Vs. Professionals

Considering the differences of domain experts (professionals) and ordinary contributors is an important issue not only in crowdsourcing [8], but also in the medical domain [11,16]. Professionals with high levels of expertise can be difficult to reach and expensive, while non-experts cannot necessarily provide the needed high-quality output. The role of expert vs. non-expert contributions has been explored in various contexts: *Voyant*, for example, uses non-experts from an online labour market to evaluate visual designs [27]. With *Voyant*, workers were found to provide relevant design feedback, and, at times, the feedback was even seen as a fair substitute for expert-feedback, since the system does not require scheduling meetings or other time-consuming face-to-face events. Therefore, a related key challenge, or opportunity, of crowdsourcing is increasing the quality of the labour produced by non-expert contributors, as domain experts can be expensive or challenging to reach.

Improving the quality of contributions by training and educating the non-expert workforce can be useful in contexts where the requester has continuous access to the same worker. For example, in learning about nutrition,

expert-generated-explanations, *i.e.* short descriptions that experts used to justify the correct answers, were found as highly useful in training non-experts [5]. In addition to experts' feedback, comparing one's own contributions to the comparative frequencies of others' contributions has been shown to rapidly improve workers' expertise in voting-based tasks [21]. Another technique for improving non-expert contributions is using rubrics to augment the assessment task. Yuan *et al.* explored how to enhance the feedback capabilities of non-experts and found that non-experts can produce expert-quality feedback on visual designs, if supported adequately with rubrics [28].

Particularly in the medical domain, the role of non-experts' expertise is often seen as complementary [11]. The common conclusion is that peer-to-peer non-expert information systems are not enough to replace the advice of seasoned clinicians [16]. Studies conducted by analysing online forums point out that non-experts can be very resourceful in medical questions and encourage others to see a doctor when necessary [15].

Crowdsourcing Health Information

The role of professionals is seen as increasingly important *e.g.*, in producing and verifying the validity of health information [7]. For instance, Kelley *et al.* used professionals to collectively brainstorm data types that would be meaningful to track in mental health context, and the data types were then used to compare students' experiences to experts' perspectives [17]. Professionals can also greatly benefit from crowdsourcing especially in ideation of new approaches or distributed knowledge collection on specific health problems [4]. Maclean and Heer also demonstrated how the crowds can be harnessed to extract official medical terms – or basically subject descriptors – from online properties containing text authored by non-experts [20]. This highlights the usefulness of non-expert contributions in the medical field, and such descriptors can then be used to help professionals find and read real-world experiences on the identified topics.

Recent work has also highlighted how crowdsourcing can offer a handy means to tap into the ideas and collective knowledge of audiences almost in real-time, making it suitable for assisting people with certain medical problems. As an example, *VizWiz* is a mobile application that allows blind people to quickly post visual problems for online workers to solve in near real-time [3]. A more recent example comes from Hong *et al.*, who explored how to help people with autism in their everyday life by providing crowdsourced social support on coping with different everyday situations [12].

Crowdsourcing Decision Support

Decision support systems are computer-aided mechanisms, historically designed to offer decision support for large organisations [2], and more recently, individual persons as well [24]. Crowdsourcing has also been investigated in the context of decision support [6]. For instance, *TaskGenies*

[19] is a system that allows the crowd to cost-effectively generate concrete action plans for users to follow. Such plans help people to coordinate and complete more tasks daily. Agapie *et al.* introduced *PlanSourcing* as means to come up with specific, achievable plans using the crowd [1]. Their work showed that friends and strangers alike can help create plans that lead to meaningful behavioural change. Another, and arguably the most similar example to our approach, is from Hosio *et al.* who introduced a crowd-based decision support system that pitched pre-defined questions, answers, and evaluation criteria to the crowd [13]. Their users were found to trust the information collected by other users, and that breaking down a question into answers and criteria is conceptually easy to understand for the crowd.

As with any crowdsourcing system [8], reaching the correct type of contributors is a key consideration also when turning the crowd's collective judgement into decision support. Concepts relying on the aggregated, collective knowledge of the audience must ensure that the crowd members do not know who the others are, or how they have assessed an option [25]. In our work, we leverage crowdsourcing to create a very lightweight online DSS that uses both clinical professionals and the general public to collect and assess health (LBP) information. Using the collected information, it then allows users to discover LBP treatments that best match their requirements.

SYSTEM DESIGN

Back Pain Workshop is accessible on a dedicated top-level domain online (<http://kipurihi.org/>), and its main functionality consists of two separate parts: assessing and contributing treatments, and using the collected knowledge to provide treatments that best match the user's needs. End users request treatments by specifying preferred values for criteria describing the treatments. This allows users to conduct a “*what-if analysis*”, *i.e.* experiment with multiple different input (criteria) configurations to explore the output (treatments) of the system [2]. Back Pain Workshop was implemented using standard HTML, JavaScript, and CSS, with PHP and MySQL on the server side.

Assessing and Contributing Treatments

When first visiting Back Pain Workshop, the front page displays a welcome message, information and motivational facts about the project, as well as information on how to get started. The visitor is informed that the purpose of the project is to assess back pain treatments provided by other visitors and to contribute their own treatment ideas. The motivation is based on the fact that most people in Finland at some point suffer from back pain, and that the relevant authorities are interested in “*hearing the visitor's important opinion*”. The front page, depicted in Figure 1, aims to entice visitors to contribute information, and thus it contains no links to the treatment discovery interface of Back Pain Workshop (described later).

Assessing Treatments

The data model storing the assessments is similar to the one developed by Gonçalves *et al.* to assess potential answers for any arbitrary question by breaking down the question into answers, criteria, and their relations [9]. The answers are evaluated in terms of every individual criterion. For example, considering the question “How to treat back pain”, the possible treatments “*Weightlifting exercises*” and “*Ashtanga yoga*” could be assessed using the criteria “*works fast*” and “*monetary cost*”, on a scale from 1 to 100.

The used data model is useful for our purposes, as it can be used for decision support by simply retrieving and mathematically aggregating a given answer's ratings [13]. Second, it is dynamic, *i.e.* new treatments can be added and assessed run-time, and each user can contribute as many ratings as desired. This helps us in crowdsourcing a big picture understanding from several small contributions (as explored recently in [10]). This is important, as LBP has no silver bullet solutions, and the number of solicited treatments – also by non-professionals and commercial operators – is practically endless [26]. Thus, the data model is the cornerstone for facilitating evolving knowledge bases of LBP information that accommodate new treatments and their assessments on the go.

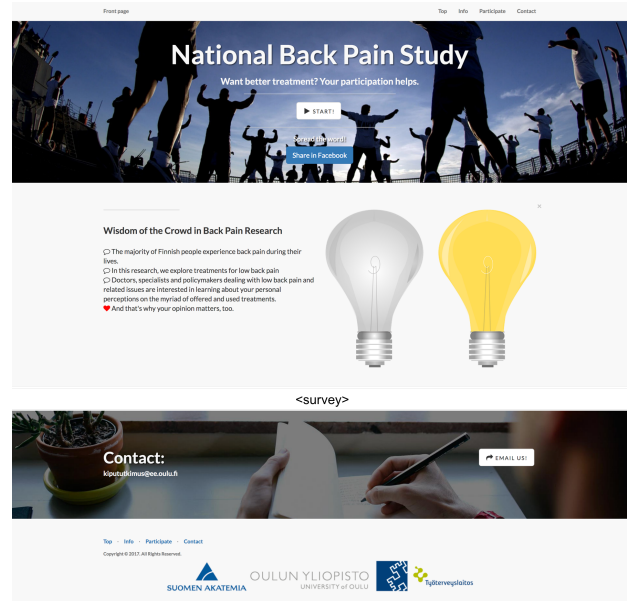


Figure 1. The front page. From top: header area with social sharing and get started button, information area, questionnaire (replaced with a placeholder to save space), contact details and footer with the logos of all collaborators to increase trust and transparency.

To get started on the site, the visitor fills a personal information questionnaire on the front page. There is no log-in process and we use Web cookies to retain identity across sessions, so the questionnaire needs to be filled only once. The questionnaire contains two mandatory fields: age and gender. We collect other miscellaneous data that may prove useful in the future when analysing the collected data

strictly from a more medical viewpoint. The questionnaire is located under the info section but omitted from Figure 1 to save space, as it is extensive (15 items: dropdowns, radio buttons, input fields on personal pain levels, pain history, surgical history, lifestyle, occupation, education, etc.). We branded the footer of the front page with our own contact details and logos of three partner organisations (University of Oulu, Finland, The Finnish Institute of Occupational Health, and Academy of Finland (funding agency) for transparency and increased trust.

After completing the questionnaire, the area is dynamically replaced with an interactive area to assess treatments, as shown in Figure 2. The area contains a collapsible deck with instructions and a maximum of 12 LBP treatments contributed by earlier users (Figure 2, B). The treatments are shown in order of total ratings given so far, with the topmost treatments having the fewest ratings. This is to balance the amount of ratings each treatment receives, as visitors are likely to start from the top. When a visitor clicks a chosen treatment, its detailed description together with a button to start are shown (Figure 2, C). The button slides open a deck with the available criteria and a slider input for each, to assess how well the criterion describes the treatment, on a 1-100 scale. To ensure precision, we followed two design guidelines for sliders discussed in [22]: no tick marks in the axis, and providing dynamic feedback to show the value in real time as the user slides the handle. The accuracy of the collected data grows as a function of number of contributors, as per theory behind the wisdom of the crowd [25].

Here's how to get started:

- Choose a treatment to assess in the left column.
- Study the details of the treatment and click "assess this" in the right column.
- Send your ratings by clicking the "submit" button.
- You do not need to assess all of the treatments or even all of the individual criteria.
- Use a scale 1-100, according to your own intuition. Generally speaking, higher values mean "more", e.g., more expensive, faster, more effective, etc.
- You have not yet rated this criterion

A

B

C

D

E

Figure 2. Assessing treatments (interface translated to English, treatments in Finnish in this screenshot). A: General instructions. B: Maximum of 12 low back pain treatments to choose from. C: Detailed description of the treatment. D: Deck with sliders to assess the treatment. E: Link to reveal input fields to propose a new criterion to the system.

When assessing treatments, the user can choose to rate any of the 12 simultaneously displayed treatments. Then, the user rates according to the displayed criteria. The user does

not have to use all criteria, but can leave a partial assessment by using only the criteria they want to. After submitting the assessments for a single treatment, a new treatment is fetched from the knowledge base and appended to the list of treatments available for assessment. This process is repeated until there are no more treatments to assess or the visitor simply stops.

Contributing new Treatments

We decided to allow visitors to contribute a new treatment to the system only after first assessing at least five treatments. This way the system allowed visitors to explore the concept and have a reasonable overview of other contributions while attempting to minimise overlap. In our early pilot tests before publishing the deployment, we observed that if we solicited new treatments immediately after the questionnaire, almost all users contributed a treatment very similar to "physical exercise". Further, we note that each visitor at this stage is unaware of the actual crowd assessments, *i.e.*, how well the treatments perform in others' opinion, and thus this design decision was a necessity towards crowdsourcing diverse information.

Treatment Discovery

The treatment discovery interface is accessible on a separate page, and visitors are not required to complete a pre-questionnaire to access it. Visitors are presented with generic information (Figure 3, A) regarding the project and a set of controls to conduct an ad-hoc *what-if analysis* [24] (Figure 3, B) to interactively retrieve treatment suggestions.

Instructions:

- Choose if you wish to browse treatments contributed and assessed by professionals, community or both combined.
- Determine the desired optimal criteria using the slider inputs.
- Bigger values on the scale (1-100) mean "more expensive, faster, more effective, etc."
- You have not yet set a desired value.

A

B

C

Figure 3. Discovering treatments (interface translated to English, treatments in Finnish in this screenshot). A: General instructions. B: Controls to choose data origin and number of treatments to fetch and to define the ideal criteria for the treatments. C: Discovery results with a goodness-of-fit value and the number of ratings that the value is based on.

The controls (Figure 3, B) allow the visitor to determine how many treatments should the system query (3, 5, or 10) and set the information source using radio buttons: professionals, non-professionals (community), or a union of both. Finally, a slider (1-100) is rendered per each criterion in the knowledge base for setting a desired "optimal" treatment that the system then tries to match as best as

possible among the available treatments. The underlying ranking process uses a summed Euclidian distance between the values entered by the user, and the respective mean values in the knowledge base. Thus, smaller values mean better goodness-of-fit. This approach follows the theory behind wisdom of the crowd [25], and has been found as producing accurate knowledge from crowds [13]. The analysis can be repeated as many times as necessary, and by using any subset of the available criteria. After clicking a button to start the match-making process, the system queries the knowledge base and returns the best matching treatments, including information on how many individual ratings is the match-making process based on (Figure 3, C).

Moderation

We decided to handle moderation manually by manipulating the MySQL databases prior to conducting any studies with real end users who discovered treatments. Instead of a community-based moderation approach, we chose to moderate entries on our own, as it was paramount that no harmful entries are offered to visitors.

EXPERIMENT DETAILS

Table 1 offers a brief introduction to the study stages.

Study Stage	Objectives
Collecting treatments from professionals, curating those to form a “gold-standard” set by clinical professionals.	Collect an initial batch of treatment suggestions from verified professionals (mailing lists), and use that batch to curate a list of “gold-standard” treatments by combining similar treatments (together with our medical collaborators).
Professionals assessing the gold-standard treatments	Provide a professional assessment to the curated “gold-standard” treatments (using the same mailing lists) in terms of four criteria defined by our collaborators.
Public call (social media, online) to assess treatments in the system (contributed earlier by professionals) and to collect and assess new treatments.	Provide a non-professional assessment to the treatments inserted by professionals, and also to collect and assess treatments from non-professionals.
Public call (social media, mailing lists, online) to use the discovery interface, now bootstrapped with both knowledge bases.	Collect end-user impressions on the treatment discovery tool (Figure 3) with surveys, from professionals and non-professionals.
Home visits and showcasing the tool at a tech fair in a prototype future hospital.	Collect interview data from real back pain patients as well as from future professionals, to complement the survey data.

Table 1. Summary of the experiment stages in chronological order (from top to bottom).

While Back Pain Workshop is capable of simultaneously facilitating both data collection and offering treatment

discovery on the collected data, we divided the study into two stages. First, we separately harvested data from professionals and non-professionals. Second, we invited users from both groups to explore the treatment discovery feature to provide us insights and feedback about the system and the collected information.

The study we report was conducted in a highly sporadic fashion because there is no crowdsourcing market available where we would be able to recruit clinical professionals on LBP. Further, we wanted to offer this platform to real people suffering from (or interested about) LBP. Thus, we had to engage several different communities and organisations to gain access to the correct audiences. The data collection took place from January 2016 to August 2016.

Crowdsourcing and Assessing Treatments

All data was collected anonymously online and from different sources to avoid participants’ influence on each other and to promote diversity among participants. These are essentially pre-requisites for a smart crowd and help in obtaining high-quality knowledge from crowds [1,25]. The professionals’ knowledge base was collected from clinical professionals (medical doctors and physiotherapists specialised in treating back pain conditions) and the second one from non-professionals (everyone else).

Professionals’ Data Collection and Curation

There is no labour market available that could provide us with input by clinical professionals. Therefore, we contacted a medical doctor from the local university hospital (professor, specialised in LBP research and treatment with decades of experience in the field). Upon agreeing on collaboration, we gained access to a mailing list of clinical professionals. More specifically, the professionals we reached this way were all licenced members of *The Finnish Spinal Health Association* (either physiotherapists and doctors specialised in treating LBP). We bootstrapped our system with five treatments and four criteria, both defined by our domain expert collaborator. A similar approach, taken in [17], was found to result in highly useful data on issues that may be important not only for the professionals directly involved, but also for the domain area at large. Then, we posted a call to participate in a “national back pain study” to the mentioned email-list, promising to share the results later.

After waiting for two weeks to ensure that most of the willing professionals on the email list had their chance to contribute, we asked three other clinical professionals (colleagues) to help curate the collected treatments into a set of “gold standard” treatments, *i.e.*, to combine similar treatments into better-articulated treatment suggestions.

At this point, because the collected treatments were no longer in the system in their original form, we purged the so far accumulated ratings from the knowledge base. Then, the system was bootstrapped with the new, curated treatments

and the same four criteria. As the treatments were reworded and curated, the call was then emailed again to the same list of professionals, but this time adding new treatments of criteria was prevented. To clarify, purging was necessary, because manual curation of the treatments changed the original wordings, and we wanted to collect a new set of clean ratings for every treatment-criterion pair exactly as they were in the system.

Non-professional Data Collection

The data collection from non-professionals was performed by publishing a similar invitation to participate on web and social media properties of *The Finnish Spinal Health Association* and any other back pain related social media properties we could find. This took place after the professionals’ data collection stage had completed. This quickly proved a functional approach, and several people shared the call on their own profiles. Further, a large regional newspaper covered the project in a large 2-page feature story, providing us with mainstream attention.

While usability-wise the system functioned identically as with professionals, now the system was bootstrapped with the curated treatments and criteria from the professionals’ data collection stage. This was never revealed to participants, however, and the instructions simply encouraged participants to “*assess treatments from other participants*”. Thus, participants never knew whether they were assessing treatments originating from professionals or from non-professionals. We see this as important, as we expect people to automatically trust more treatments that are from a trusted origin (professionals). Such skewing of opinions based on authority is unwanted in our scenario that relies on wisdom of the crowd.

While we also solicited new criteria at this stage, these were merely collected and never appeared in the system. The addition of one criterion means that new data must be collected for each treatment in the system, if we want that criterion to be used in data analysis. Thus, we opted to maximise our chances in collecting comprehensive data for the criteria defined by our domain expert collaborator instead of increasing the complexity of the data collection.

End-user Perspective

After the data collection stage, we opened the final stage of the study where we solicited end-users to discover treatments and provide feedback. Again, the call was published in social media and also by emailing prior participants. This stage also took place online, and the visitors landed on a dedicated page containing the interactive discovery interface depicted in Figure 3. The system was modified to prompt a call to take a survey after finishing a minimum of 5 *what-if analyses*. We also collected interview data by 1) visiting the homes of 4 back pain patients, and 2) by participating a tech fair, in a local “prototype future hospital” in Oulu. There, Back Pain Workshop was accepted as one of the showcased “future medical systems”. 29 next-generation clinicians (final year

medical students and physiotherapists) used the system and provided insights on its potential or challenges from the doctors’ perspective. We note that in Finland, final year medical students have already been treating real patients for 2-3 years, well qualifying the users as professionals.

RESULTS

Professionals’ Data Collection Stage

Initially, 65 professionals (37 m, 28 f, age range 29 to 76 years, mean age = 53.0, SD = 11.4) filled in the pre-questionnaire. Of them, 39 contributed a new treatment suggestion to Back Pain Workshop. No new criteria were added by professionals at this point. Then, the collaborating professionals grouped the 39 entries into 12 items based on their thematic similarities, using a shared online Google spread sheet. Table 2 lists the four criteria and the final 12 aggregated treatments. These are translated versions of the titles (we omit the descriptions to save space), presented here to show what type of treatments were being assessed. The versions in our system all contain more thorough explanations, as shown in Figures 2 and 3.

CRITERIA

C1: Cost to the patient	C3: Duration of effect
C2: Speed of effect	C4: Efficiency in general

TREATMENTS

1: Therapeutic exercise	7: Treatment of comorbid conditions
2: Exercise in general	8: Ergonomic advice and adjustments
3: Information, education	9: Biopsychosocial approach
4: Spinal manipulation	10: Stretching
5: Painkillers, drugs	11: Breathing techniques
6: Strength training	12: Psychological treatment

Table 2. Final criteria and treatments contributed by professionals.

Finally, after purging and bootstrapping the knowledge base again, we invited professionals to assess the 12 treatments against the four criteria. This time, 28 professionals (19 m, 9 f, age ranging from 29 to 70, mean = 55.8, SD = 10.0) filled in the pre-questionnaire on the front page. Of these, 17 proceeded to rate available treatments, and together they contributed a total of 610 ratings (average of 50.8 per treatment, or 12.7 ratings per treatment-criterion pair). This set of 12 assessed treatments basically formed the professionals’ knowledge base used in the studies in this article. A second knowledge base was collected as follows.

Non-professionals’ Data Collection Stage

288 non-professional participants (63 m, 225 f, age range 23 to 75, mean = 47.2, SD = 11.8) contributed to the non-professional knowledge base. We consider this crowd as relatively knowledgeable about the topic, as only 18 of the entire crowd had never experienced back pain during their lives. Further, at the exact moment of participating, 211 (73%) users were experiencing at least some magnitude of back pain. In total, this crowd contributed 69 new treatment ideas. We manually moderated 13 of these, as their titles or descriptions were gibberish, missing or they were clearly too ambiguous (e.g., “*asdfsdf*”, “*pdf*”, or “*do stuff*”). Thus,

we accepted 56 valid new entries to be added to the set of treatments, for later visitors to assess. We did not receive any derogatory or medically harmful treatment ideas, and thus we did not have to manually intervene to the collection process. Non-professionals also contributed 2 new criteria, but these were not added to the system to ensure sufficient data collection for the four criteria that were curated earlier by professionals. The non-professionals' criteria suggestions were: i) *availability [in your area]*, ii) *duration of the treatment*. The four patients visited consisted of 3 females (25, 25, 60 y/o) and 1 male (31 y/o).

Summary and Data Cleaning

Table 3 presents an overview of the collected data.

	N	Treatments	Criteria	Ratings
Professionals	39 + 17	39, curated to 12	4	610
Non-professionals	288	69, 56 valid	2	8391
Total	305	108, 68 valid (12+56)	6	9001

Table 3. A summary of data collection results in the two stages: professionals, non-professionals and total.

In our analysis, we omit three treatments (and their associated ratings) that did not amass at least 11 ratings for each of the four criteria. This limit of 11 ratings was chosen as all treatment-criterion pairs in the professionals' knowledge base had a minimum of 11 ratings. Thus, in the analysis we focus on the remaining 65 treatments across the four criteria. After data cleaning, a total of 8897 ratings remained in the dataset.

Quantitative Data Analysis

Professionals' and Non-professionals' Assessment of the Professionals' Knowledge Base

To provide a generic overview to the knowledge base harvested from professionals, and more specifically to how the treatments are rated by both groups (professionals and non-professionals), we plot the density of ratings given per each criterion in Figure 5 (x-axis: rating score).

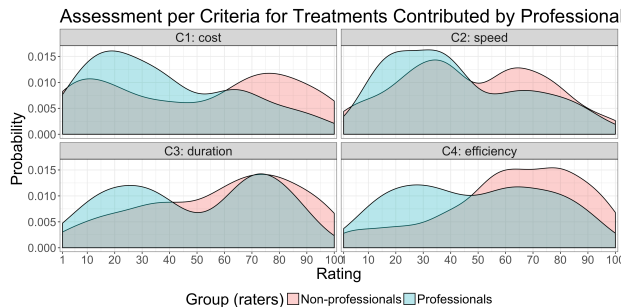


Figure 5. Density of ratings/criteria (professionals' knowledge base).

In Figure 5 we can observe the professionals giving lower ratings across all criteria, but most prominently for criteria

1 and 4 (cost and efficiency). To investigate further, we plot in Figure 6 a breakdown of the treatments in the professionals' knowledge base and provide a statistical analysis that pinpoints the origin of the differences observed in the previous figure. A comparison between the ratings by the two rater groups reveals 10 criterion-treatment pairs where the assessment significantly differs between the groups ($p < 0.05$, Wilcoxon Rank-Sum Test). The 10 pairs are denoted in Figure 6 along with their confidence levels ('***': $p < .01$, '**': $p < .05$). Note that Table 2 can be used to map the Treatment and criteria ids used in the plot to the actual treatments.

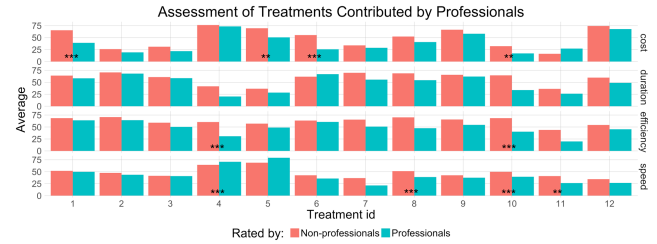


Figure 6. Assessments by professionals and non-professionals to the 12 treatments in the Professionals' knowledge base.

Non-professionals' Assessment on Both Knowledge Bases

First, when comparing the 2 knowledge bases against each other (*i.e.*, 12 treatments by professionals vs. 56 by non-professionals, using the Wilcoxon Rank-Sum Test), we find that non-professionals perceive the treatment batch by professionals as overall more expensive (mean 50.4 vs 44.1), having longer duration of effect (mean 58.5 vs 51.9) and being more efficient (mean 62.2 vs 56.7). However, there was no significant difference found in the overall speed, *i.e.* how fast do they start to work, of the treatments (47.8 vs 48.8).

Next, we focus on contrasting the best and worst treatments in both knowledge bases, to investigate if non-professionals are capable of producing treatments that are perceived comparably to those by professionals. Bottom-3 segments in the non-professional knowledge base were rated overall lower for C1, C3, and C4 ($p < 0.01$, Wilcoxon Rank-Sum Test). Interestingly, all the top-3 treatments in each category show no significant differences, meaning that the highest rated treatments by non-professionals were assessed similarly to those by professionals across all four criteria. A table with all the treatments and comparison between the treatment pairs in the said segments is available online in the auxiliary materials of this manuscript.

Survey Overview

In the online end-user evaluation stage, 43 participants (6 m, 37 f, age ranging from 18 to 64, mean = 40.2, SD = 11.7) filled in the final questionnaire after having conducted a minimum of five what-if analyses using Back Pain Workshop to discover treatments. Out of the 43 participants, 41 had experienced LBP in their lives and 7 identified themselves as clinical professionals on the matter.

The online end-user questionnaire revealed that both user groups found the combined knowledge of professionals and non-professionals as the key in our system. When explicitly asked to choose one of the three knowledge sources as the most *interesting* one, 30 of the 43 survey respondents chose the combined one. Overall, the combined knowledge base was also seen as the most trusted one (albeit not statistically significantly more than the professionals' one).

Key Qualitative Insights

The following qualitative analysis considers combined survey and interview results from 1) the end-user questionnaires, 2) patient home visits, and 3) the future hospital tech fair. Two of the paper's authors first collaboratively reduced the result set to items that relate to the original research questions (value to stakeholders, perceptions by stakeholders). Then, the same two authors independently identified emerging themes in the result set. Finally, the results were inserted in a shared spreadsheet online, and the authors collaborated to find and refine the largest common themes.

Complementing Professionals' Expertise and Treatments

The end users of the decision support tool – both professionals and non-professionals – agreed that the inclusion of the non-professional knowledge clearly adds value to the system. Particularly in the non-professionals' opinion the voice of their own community was deemed valuable: *"Sometimes, practical treatments and authentic experiences can help much better than advice from professionals"* (Female, 29).

Professionals made no such direct claims about the validity of non-professional advice but on several occasions stated that they would be happy to give the tool to their patients before a face-to-face consultation. That way, patients would have ideas from both professionals and their own non-professional peers to discuss, and they could together agree on a treatment experiment that has scientific or practical validity to it (is a part of national recommendations, or the doctor personally agrees with it). The patient would also feel that she is given attention to: *"Much better than forums and actually I think this could be offered on a tablet before going to see for example a doctor to show what their options are"* (Female professional, at the hospital tech fair). Particularly in regard to giving attention to patients, it was indeed fairly alarming in our collaborators' opinion to notice that *"taking patients more seriously"* was rated as the most rapidly working treatment by non-professionals (see auxiliary materials online).

Pitfalls and Roadblocks to Adoption

One particularly troublesome issue identified by several participants is how to discover the system online in the first place. One of the interviewed professionals summarized it: *"By far the biggest problem is making people find it online"* (Male professional, tech fair). We agree on this, and are currently working to tackle this problem. This is however, far from trivial. The battle for attention online is fierce, with

both commercial actors as well as established public health information outlets competing for the audiences.

The second limitation identified by some of our users was the system's suitability (or lack thereof) for serious back pain cases. A comment from a non-professional participant, *"When one is completely burnt out with pain, using tools like this is pointless"* (Female, 43), is a testimony to the harsh reality that many back pain patients experience: Sometimes the treatments just do not help. Another popular, shared sentiment among participants was about finding nothing new among the treatments, implying that they have over the years tried and heard about everything already. In our context, this suggests that Back Pain Workshop is best suited to early-stage information discovery, with patients that do not yet have excessive, long-term first-hand experience with the burdensome condition.

DISCUSSION

The potential of crowdsourcing has long been acknowledged in public health applications [4] and – relevant to our work – in decision support as well [6]. In our research, we set to collect and analyse LBP treatments from two distinct stakeholder groups. We did this via means of crowdsourcing and to help discover best-matching LBP treatments. We found this approach as highly compelling, since it provides clear benefits for both of the groups, and particularly non-professionals valued being able to tap into the knowledge of professionals while their own opinions were respected as well.

A typical way to aggregate crowdsourced contributions to draw reliable conclusions is assigning "trustworthiness scores" for the contributors [8]. We wanted to offer the intended end-users (non-professionals) a clear means to choose whose judgement to trust, as especially in the medical domain the separation of professionals and non-professionals can be seen as beneficial [3,12,17]. Other systems exploring distinct groups have also found that the added diversity tends to lead into higher number of novel ideas [1].

Another differentiating aspect in our work is reusability of the knowledge bases. Crowdsourcing decision support in the medical domain typically focuses on providing (near) real-time support (e.g., [3,12]). However, the crowds' role in collecting reusable knowledge bases has also been recently acknowledged [6]. A knowledge base that contains numerous options to choose from, contributed by both professionals and non-professionals was indeed seen handy in the case of LBP, where the lack of therapeutic consensus causes trouble in choosing treatments [26]. Unlike in forums where information is hidden inside threads, Back Pain Workshop offers a clean interface to simply determine an ideal treatment and, as a result, returns a list of best matches from the knowledge base.

Contrasting the Knowledge Bases (RQ1)

First, while the end-users clearly trusted the professionals (Figure 6), they found the combination of both as the overall most *interesting* aspect in the work (30 out of 43 survey respondents). The design choice was further praised in the surveys and interviews, where including both viewpoints was seen as highly valuable. In the opinion of our own medical collaborators, for example, the ratings data is a precious outcome for two reasons: they can easily find out how the treatments work “out there”, and they can self-reflect when their own estimates differ from those of the collective professional community.

Related work suggests that with adequate support the output of non-professionals can be improved to match that of the professionals in many cases [5,21,28]. In our study, the best treatments by non-professionals were assessed very similarly to the best by professionals. We find this remarkable, given how the professionals’ treatments were carefully curated by a group of seasoned LBP professionals with decades of practical knowledge from the field. The low-performing treatments by non-professionals, however, were clearly assessed lower than the worst professionals’ treatments (Table 4).

Crowdsourcing literature suggests that laymen’s contributions, when augmented and filtered in novel ways, can often replace the need for experts [27]. In our case, however, the professionals were quick to point that the treatments by non-professionals need to be scientifically validated before being even considered to be included as treatment suggestions in any kind of national authoritative guidelines. This type of manoeuvring between medical liability and practical help has been also identified earlier as one of the key challenges in medical peer-support systems [15]. There must exist a clear division between official treatments by clinical professionals and our crowdsourced treatment database accessible via Back Pain Workshop.

In addition to simply providing “good treatments”, the non-professionals’ knowledge base is valuable in other ways as well. Based on an informal content analysis, the treatments in the non-professionals’ knowledge base were not as medically exactly articulated in nature as in the professionals’ one. They also at times included typos, bad grammar, and upper-case text. However, textual non-professional descriptions on which treatments seem to work can be used by professionals in learning about the real-world experiences of potential patients [20].

Value to Stakeholders (RQ2)

The most valuable aspect of our work for professionals was being able to contrast their own knowledge to that of non-professionals. Indeed, several treatments by professionals were perceived differently by non-professionals than by professionals themselves (10 treatment-criterion pairs with statistically significant differences, depicted in Figure 6). This, as was evident in our discussion with our collaborators, is highly important knowledge for the

professionals, as they learn to “calibrate” their own expectations to the experiences of the non-professionals (*i.e.*, their potential patients) out there. In addition, Back Pain Workshop was seen useful especially for discussing different treatment options and their validity for patients that have not yet sought help for their LBP before. The same opinion was shared by non-professionals, who on many occasions stated that the treatments that the system can provide are not very helpful if they already have a rich history with clinical care for their condition.

Furthermore, non-professionals found the information repository trustworthy, and useful in particular for users with less-severe cases of LBP. Another interesting finding among non-professionals is how they believe that the tool will inform professionals. There was a clearly observable tendency to think that professionals are somehow ignoring the non-professional opinion, and that the professionals simply cannot always know all the “tricks” that non-professionals have discovered on their own. This is much in line with related work by Hartzler and Pratt [11], who conclude that non-professionals’ role is a complementary one, and that non-professionals are often able to provide information that clinicians cannot. Indeed, some of the professionals who participated in our study supported this viewpoint and brought up that LBP is a non-trivial problem with no silver bullet solutions.

Design and Development Opportunities

Mitigating gaming attempts or finding ways to improve the quality of aggregated contributions is one of the key challenges in crowdsourcing in general [8]. In our case, gaming constitutes one of the future challenges to tackle: how to prevent people from inserting treatments that benefit themselves (*e.g.*, commercially), as already noted by one professional in our interviews? The granular data model [9] is key here, and for example simple approaches like removing outliers inserted to the system during the same time window can be explored, to begin with. Additionally, community-based flagging of treatments as well as some kind of metric about scientific validity of the treatments – even if implemented as just one among the assessment criteria – could be helpful in weeding out generally undesired results.

Another related issue is the number of similar treatments in the system. Users could be offered some type of NLP-based solution that helps them find potentially overlapping ideas and then improve upon them rather than contribute a similar one, much like the concept explored in [19].

As stated earlier, we carefully surveyed all respondents in the data collection stages. However, we did not leverage the collected data in the treatment discovery by allowing users to more granularly filter the information source. We acknowledge that personal factors of the respondents can affect their opinions and data. The current straightforward approach of treating each respondent as equal is straightforward and functional, but not necessarily optimal.

Thus, further developing and leveraging surveys to create more granular controls for choosing the knowledge source (Figure 3) would help people to discover even more personalised and trustworthy treatments (*e.g.*, based on knowledge collected only from people suffering from similar type and magnitude of back pain).

Another direction to take with the work is exploring its generalisability across other health problems. We point, however, that *e.g.*, contacting all the professionals needed to bootstrap the professionals' knowledge base was a daunting job, and automating this is far from trivial. Also, coming up with unambiguous criteria mechanism (users have differing understanding of criteria) needs to be explored. Developing the tool into a mobile version that professionals can use together with their first-time patients, but only after first donating their own wisdom, could be one solution to begin with.

As an interesting final consideration for future technical development, the system shows promise to be a feedback channel in general for people with LBP. In the beginning of the study we created a dedicated email alias for the project, and included it on the front page (Figure 1), for users to send comments and questions to. We received 7 emails, among which were 2 from professionals asking to be involved in discussing the results in more depth, and 3 from non-professionals to personally thank us for the work we are doing for the community and to provide detailed and lengthy descriptions of their struggles with the national health care system and their own LBP condition. Yet, the system is admittedly hard to find online, as indicated as the biggest problem by professionals.

Limitations

One of the most prominent limitations in our real-world case study was the initial seeding of the system with professionals' contributions when collecting data from non-professionals. Non-professionals who saw the most likely very "popular" treatments were discouraged to suggest the same treatments as their own contributions. This was, however, a strict necessity as we needed to collect the assessment from non-professionals to professionals' contributions.

Further, due to counterbalancing the shown treatments, this issue affected only the first few non-professionals who contributed new treatments, and for example in Table 4 we can indeed observe that non-professionals did suggest similar treatments to those by professionals. Then, the origin of treatments in the two knowledge bases most likely contain minor overlap. Although we attempted to ensure that only non-professionals contributed to the non-professional knowledge base, the calls were public and spread organically in social media. Thus, it is possible that some professionals also saw the call and contributed to the "incorrect" knowledge base.

We also did not account for personal differences of the users who contributed the knowledge bases and how that affects the data. Some users, due to their personal life situation, may have provided extremely low or high ratings for certain treatments. Currently, our design attempts to mitigate these types of errors by relying on wisdom of the crowd, but other mechanisms, such as deploying surveys and increasing the granularity of the treatment discovery interface, need to be considered as well (as discussed among the future development opportunities).

Finally, while the system design is highly dynamic and accommodates adding new treatments and criteria runtime, we did not include the user-contributed criteria. This was due to rapidly increasing data collection burden: new data must be collected for all treatments for the new criterion. The real issue to overcome here is securing a steady flow of new contributors to the system.

CONCLUSION

Back Pain Workshop is a crowd-powered online system that lets users discover LBP treatments that originate from two distinct crowds: professionals and non-professionals. Our results indicated that, according to non-professionals' assessment, the highest-rated treatments crowdsourced from non-professionals are of comparable quality to the highest-rated treatments from professionals. However, in their assessment, the worst treatments from non-professionals were perceived clearly as worse than the worst from non-professionals. Professionals, on the other hand, perceived many of their own suggestions to be of lower quality than what non-professionals think of the same treatments. Clinical professionals are, thus, more critical of the presented "gold-standard" solutions.

Back Pain Workshop is not only about data: It was seen as a reliable information source that contributed especially by bundling the non-professional opinion in the same tool that offers professional advice. The clinical professionals found value in self-reflection, because the general public disagreed with their assessment of certain treatments. They also found potential in the tool in clinical use with new patients with no extensive experience with back pain.

We contribute to crowdsourcing literature with a real-world case study that built diverse, reusable knowledge bases on a relevant and burdensome global issue: low back pain. In the future, we shall focus on improving this particular deployment and experimenting with other medical conditions.

ACKNOWLEDGEMENTS

This work is partially funded by the Academy of Finland (Grants 276786-AWARE, 286386-CPDSS, 285459-iSCIENCE, 304925-CARE), the European Commission (Grant 6AIKA-A71143-AKAI), and Marie Skłodowska-Curie Actions (645706-GRAGE). We also acknowledge all the participants who were kind enough to share their stories and pains with us.

REFERENCES

1. Elena Agapie, Lucas Colusso, Sean A. Munson and Gary Hsieh. 2016. PlanSourcing: Generating Behavior Change Plans with Friends and Crowds. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, ACM, 119-133. <https://doi.org/10.1145/2818048.2819943>
2. Steven Alter. 1982. *Decision support systems: Current practice and continuing challenges*. Reading, Massachusetts: Addison-Wesley Publishing Co., 1980, 316 pp. *Behavioral Science*, 27 (1). 91-92. <https://doi.org/10.1002/bs.3830270109>
3. Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White and Tom Yeh. 2010. VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, ACM, 333-342. <https://doi.org/10.1145/1866029.1866080>
4. D. C. Brabham, K. M. Ribisl, T. R. Kirchner and J. M. Bernhardt. 2014. Crowdsourcing applications for public health. *Am J Prev Med*, 46 (2). 179-187. <https://doi.org/10.1016/j.amepre.2013.10.016>
5. Marissa Burgermaster, Krzysztof Z. Gajos, Patricia Davidson and Lena Mamykina. 2017. The Role of Explanations in Casual Observational Learning about Nutrition. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, 4097-4145. <https://doi.org/10.1145/3025453.3025874>
6. Chao-Min Chiu, Ting-Peng Liang and Efraim Turban. 2014. What can crowdsourcing do for decision support? *Decision Support Systems*, 65. 40-49. <https://doi.org/10.1016/j.dss.2014.05.010>
7. R. J. W. Cline and K. M. Haynes. 2001. Consumer health information seeking on the Internet: the state of the art. *Health Education Research*, 16 (6). 671-692. <https://doi.org/10.1093/her/16.6.671>
8. Anhai Doan, Raghu Ramakrishnan and Alon Y. Halevy. 2011. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54 (4). 86-96. <https://doi.org/10.1145/1924421.1924442>
9. Jorge Goncalves, Simo Hosio and Vassilis Kostakos. 2017. Eliciting Structured Knowledge from Situated Crowd Markets. *ACM Transactions on Internet Technology*, 17 (2), article 14. <https://doi.org/10.1145/3007900>
10. Nathan Hahn, Joseph Chang, Ji Eun Kim and Aniket Kittur. 2016. The Knowledge Accelerator: Big Picture Thinking in Small Pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 2258-2270. <https://doi.org/10.1145/2858036.2858364>
11. A. Hartzler and W. Pratt. 2011. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. *Journal of Medical Internet Research*, 13 (3). e62. <https://doi.org/10.2196/jmir.1728>
12. Hwajung Hong, Eric Gilbert, Gregory D. Abowd and Rosa I. Arriaga. 2015. In-group Questions and Out-group Answers: Crowdsourcing Daily Living Advice for Individuals with Autism. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 777-786. <https://doi.org/10.1145/2702123.2702402>
13. Simo Hosio, Jorge Goncalves, Theodoros Anagnostopoulos and Vassilis Kostakos. 2016. Leveraging Wisdom of the Crowd for Decision Support. In *British Human Computer Interaction Conference*, article 38.
14. D. Hoy, L. March, P. Brooks, F. Blyth, A. Woolf, C. Bain, G. Williams, E. Smith, T. Vos, J. Barendregt, C. Murray, R. Burstein and R. Buchbinder. 2014. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis*, 73 (6). 968-974. <https://doi.org/10.1136/annrheumdis-2013-204428>
15. Jina Huh. 2015. Clinical Questions in Online Health Communities: The Case of See your doctor Threads. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 1488-1499. <https://doi.org/10.1145/2675133.2675259>
16. Jina Huh, Wanda Pratt, Jina Huh and Wanda Pratt. 2014. Weaving clinical expertise in online health communities. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 1355-1364. <https://doi.org/10.1145/2556288.2557293>
17. Christina Kelley, Bongshin Lee and Lauren Wilcox. 2017. Self-tracking for Mental Wellness: Understanding Expert Perspectives and Student Experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, 629-641. <https://doi.org/10.1145/3025453.3025750>
18. Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*, ACM, 1301-1318. <https://doi.org/10.1145/2441776.2441923>
19. Nicolas Kokkalis, Thomas Köhn, Johannes Huebner, Moontae Lee, Florian Schulze and Scott R. Klemmer. 2013. TaskGenies: Automatically Providing Action Plans Helps People Complete Tasks. *ACM Trans.*

- Comput.-Hum. Interact.*, 20 (5). 27 21-27 25.
<https://doi.org/10.1145/2513560>
20. D. L. MacLean and J. Heer. 2013. Identifying medical terms in patient-authored text: a crowdsourcing-based approach. *J Am Med Inform Assoc*, 20 (6). 1120-1127.
<https://doi.org/10.1136/amiajnl-2012-001110>
 21. Lena Mamykina, Thomas N. Smyth, Jill P. Dimond and Krzysztof Z. Gajos. 2016. Learning From the Crowd: Observational Learning in Crowdsourcing Communities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 2635-2644.
<https://doi.org/10.1145/2858036.2858560>
 22. Justin Matejka, Michael Glueck, Tovi Grossman and George Fitzmaurice. Year. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 5421-5432.
<https://doi.org/10.1145/2858036.2858063>
 23. Marion K. Poetz and Martin Schreier. 2012. The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *Journal of Product Innovation Management*, 29 (2). 245-256. <https://doi.org/10.1111/j.1540-5885.2011.00893.x>
 24. Daniel J. Power and Ramesh Sharda. 2007. Model-driven decision support systems: Concepts and research directions. *Decision Support Systems*, 43 (3). 1044-1061. <https://doi.org/10.1016/j.dss.2005.05.030>
 25. James Surowiecki. 2005. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Anchor.
 26. Pauline Savigny Underwood, Watson Paul and Martin. 2009. Early management of persistent non-specific low back pain: summary of NICE guidance.
<https://doi.org/10.1136/bmj.b1805>
 27. Anbang Xu, Shih-Wen Huang and Brian Bailey. Year. Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM, 1433-1444. <https://doi.org/10.1145/2531602.2531604>
 28. Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P. Dow and Bjorn Hartmann. Year. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, ACM, 1005-1017.
<https://doi.org/10.1145/2818048.2819953>