Crowdsourcing as an Innovative Communication Strategy in Early Melanoma Detection

Robert Gehl, Ph.D.¹, Andy J. King, Ph.D.², Douglas Grossman, M.D., Ph.D.³, and Jakob D. Jensen, Ph.D.¹⁴

¹ Department of Communication, University of Utah
² Department of Communication, University of Illinois
³ Department of Dermatology and Department of Oncological Sciences, University of Utah
⁴ Department of Health Promotion & Education, University of Utah

Author Contact: Robert W. Gehl
The University of Utah
255 S. Central Campus Drive
2515 LNCO
Salt Lake City, UT 84112
robert.gehl@utah.edu
801-581-7313 (phone)
801-585-6255 (fax)
Abstract

**Background:** Skin self-examination (SSE) is the primary method for identifying atypical moles. Unfortunately, past research has shown that SSE has low sensitivity. The current study investigates whether collective effort can improve SSE. Collective effort problem solving, or crowdsourcing, uses the intelligence of a group to make decisions; for example, contestants on game shows use collective effort when they “ask the audience” for help. Contestants can try to answer the question on their own (individual effort) or rely on the opinion of the audience (collective effort). Collective effort can be effective even when a single person has low reliability at a task, as the pattern of the group overcomes the limitations of each individual.

**Methods:** Adults (N = 500) were recruited from a mall in the Midwest. Participants viewed educational pamphlets about SSE and then completed a mole identification task. For the task, participants were asked to circle moles that appeared atypical. Forty mole images were provided; nine of which were clinically diagnosed melanomas.

**Results:** Consistent with past research, individual effort exhibited modest sensitivity (.58) for identifying atypical moles in the mole identification task. As predicted, collective effort overcame the limitations of individual effort. Specifically, a 19% collective effort threshold exhibited superior sensitivity (.90).

**Conclusion:** Modern communication technology facilitates collective effort problem solving. The results of the current study suggest that limitations of SSE can be countered by collective effort, a finding that supports the pursuit of collective effort interventions in early melanoma detection.

**Keywords:** crowdsourcing, melanoma, SSE, screening, collective effort
Crowdsourcing as an Innovative Communication Strategy in Early Melanoma Detection

Detecting all types of skin cancer is an important public health goal; however, the most deadly type, and the focus of most screening efforts, is melanoma. Melanoma incidence has been increasing steadily over the last 30 years (1). Not only is melanoma increasingly common, but it is also very deadly. The five-year survival rate for distant stage melanoma is only 15% with approximately 1 person dying of this illness every 61 minutes in the United States. Five-year survival rates for melanoma improve dramatically if the cancer is caught before it advances to a distant stage. The survival rate is 98% if the cancer has not spread to lymph nodes and 61% if at the regional stage (1).

Melanoma can be detected early via routine clinical examination by a dermatologist (2). Yet, mass screening by dermatologists is neither cost-efficient nor feasible; therefore, routine clinical examination is only recommended for high-risk individuals or those with numerous and/or atypical moles (3). The primary method for identifying the latter is SSE which is a patient-initiated behavior designed to identify atypical moles on the skin. Unfortunately, SSE 1) has low sensitivity for detecting atypical moles (3-6), 2) is only marginally improved by existing educational techniques (7), and 3) is rarely practiced or effective at directing people to clinics (6,8). As a result, SSE is not recommended by most public health organizations (9). Thus, there is a significant need for innovative alternatives that increase the accuracy and impact of SSE.

SSE is typically a solitary (and ineffective) activity; however, evidence suggests that self-examination accuracy increases as family members assist with the task. For example, 86% of melanomas are initially identified during self-examination, often with the assistance of family members and friends (10). Family members and friends help to identify moles in hard to see locations and, perhaps more importantly, provide feedback about the suspicious mole.
In light of this finding, it seems plausible that SSE may be an ideal candidate for collective effort problem solving. Collect effort problem solving uses the intelligence of a group to make decisions; for example, contestants on game shows use collective effort when they “ask the audience” for help. Contestants can try to answer the question on their own (individual effort) or rely on the opinion of the audience (collective effort). Collective effort can be effective even when a single person has low reliability at a task, as the pattern of the group overcomes the limitations of each individual.

Indeed, a strength of modern communication technology is that it allows crowds of people to be mobilized to perform tasks typically carried out by a single person – a strategy often referred to as “crowdsourcing” (11). Crowdsourcing was coined in 2006 by Wired editor Jeff Howe (12). Howe defined crowdsourcing as outsourcing tasks to masses of people outside a firm or organization via open calls. This mass of people – the "crowd" – is invited to solve a problem, come up with ideas, or create products.

Crowdsourcing has been explored in various public health and medical domains. For example, by using crowds of people, typically linked up via telecommunications networks and social media, disaster relief efforts can be better coordinated because affected people can report problems easily to centralized relief organizations (13). Researchers can compare "real world" uses of prescription medication to clinical trials and further refine knowledge of treatment efficacy because patients can report their experiences with medicine directly to doctors and pharmaceutical companies (14). Patients can self-diagnose and support one another in online forums (15). Finally, the Systematized Nomenclature of Medicine- Clinical Terms (SNOMED CT) can be further refined by opening up medical documentation to large crowds who can propose connections between sub-concepts (16).
Crowdsourcing can successfully leverage the participatory nature of new communication technology in health domains, but it is not an optimal strategy in all situations. We propose that crowdsourcing is a viable approach when 1) individual effort is underperforming and/or inefficient and 2) collective effort improves accuracy and/or efficiency. For example, NASA successfully employed crowdsourcing by recruiting thousands of lay-users to comb through millions of planetary images distinguishing craters from shadows (17,18). Individual NASA employees could have analyzed each image, but it would have taken decades to complete the task and individual evaluator error would have undermined the results.

To explore the potential of crowdsourcing in early melanoma detection, the current study evaluates whether collective effort outperforms individual effort in the context of SSE. Existing research has revealed that SSE has limited accuracy when carried out by an individual. Currently unknown is whether the limitations of SSE can be countered by collective effort. That is, if a person could “ask the audience” whether a mole was atypical would the audience exhibit superior sensitivity and specificity? If the crowd is more efficacious at detecting atypical moles, then this would support the implementation of crowdsourcing as a possible alternative to SSE.

Methods

Procedure

Data for the present study were collected as part of a larger project testing the efficacy of different educational techniques designed to improve the accuracy of SSE. Participants over the age of 18 were recruited by the research team from a suburban mall located in a mid-sized Midwestern city. Large signs informed mall shoppers about the study, including the incentive ($15 mall gift card) and time required for participation. In total, 500 individuals were recruited into the study. Approximately 1 in 25 people stopped to participate in the study.
Participants first completed a pretest survey, then received printed materials to examine, and finally completed a posttest survey. The printed materials were pamphlets that taught the participants standard SSE techniques (e.g., ABCDEs, Ugly Duckling Sign). In the posttest, participants completed a mole identification task. Forty mole images were used in the task. Mole images were obtained from the company MoleMap (http://www.molemap.co.nz/), as well as Internet sources such as the National Cancer Institute Visuals database (http://visualsonline.cancer.gov/). In total, nine of the forty images were moles clinically identified as melanoma cases. Participants were asked to circle all moles that appeared atypical and potentially cancerous. Following completion of the mole identification task, participants were thanked for their participation and provided compensation. A university institutional research board approved the research protocol, questions, and materials.

Participants

Average participant age was 36 years old (\(M = 36.14, SD = 14.22\)), ranging from 18 to 80 years old. Participants were more likely to be female (57.2%, \(n = 286\)), white (73.8%, \(n = 369\)), and at least a high school graduate (92.8%, \(n = 461\)). Most participants were either single (38.4%, \(n = 192\)) or married (41.6%, \(n = 208\)), with fewer participants identifying as being divorced, widowed, separated, or living with a long-term partner. Skin cancer risk was measured using the brief skin cancer risk assessment test (BRAT) (19). BRAT estimates classified just over half of participants as low risk (54.8%, \(n = 274\)), a third at moderate risk (34.8%, \(n = 174\)), and a small proportion at high risk (10.2%, \(n = 51\)). Three participants failed to complete the mole identification task.
Statistical Methods

Participants completed the mole identification task by circling the moles they believed to be atypical and potentially cancerous. Melanomas classified as atypical were true positives (TP). Typical moles classified as typical were true negatives (TN). Typical moles classified as atypical were considered false positives (FP), and false negatives (FN) were atypical moles that were not identified as atypical.

After classifying individual participant effort into units of true/false positives/negatives, four proportional scores were calculated: sensitivity, specificity, positive predictive value, and negative predictive value. Sensitivity represents the proportion of true positives divided by the sum of true positives and false negatives: TP/(TP+FN). Specificity is the number of true negatives divided by the sum of true negatives and false positives: TN/(TN+FP). Positive predictive value (PPV) is calculated by dividing the true positives by the sum of the true positives and false positives: TP/(TP+FP); while negative predictive value (NPV) is calculated by dividing the true negatives by the sum of the true negatives and false positives: TN/(TN+FP) (20-22). These four proportions represent individual effort for the identification of atypical/typical moles. Average individual effort refers to the mean performance of an individual in the sample.

Following the calculation of individual effort scores, collective effort scores were calculated. Collective effort considers the pattern across all users (e.g., 35% of participants think a particular mole is atypical). For instance, imagine that 100 people were asked to examine 5 mole images (1 of which was a clinically diagnosed melanoma). In this hypothetical situation, most people (65%) incorrectly classify the melanoma as typical (i.e., a false negative), a response that yields a low average individual effort score. Collective effort ignores the limitations of
individual effort and considers the pattern of the group (see Table 1). In this case, the pattern of the group is revealing as 35% of people did score the melanoma as atypical (i.e., a true positive) which is a relatively high number compared to the rest of the moles. A reasonable person would be concerned about mole #5 in Table 1, as the score is unusual. Thus, this hypothetical scenario illustrates the potential of collective effort to successfully overcome the limitations of individual ability in SSE.

The pattern across all users is useful information; for example, contestants on a game show “ask the audience” for help and then interpret what the audience response means. But research can aid users of collective effort data by identifying meaningful thresholds or cut-off points. In the current data (see Table 2), the majority of people viewed most moles as typical (Median % scored atypical = 16.25%). In fact, for 60% of the mole images fewer than 19% of raters were concerned. Given that, we examined 19% as a threshold for melanoma identification. For comparative purposes, we also examined 65% as a collective effort threshold. That threshold was selected as the majority of the moles (85%) were identified as atypical by fewer than 65% of people.

**Results**

The goal of the current study was to determine if crowdsourcing—collective effort—was more efficient than individual effort at identifying atypical moles (see Table 3). In the current study, individual effort correctly identified 58% of melanomas (sensitivity), and correctly classified 81% of moles as not melanoma (specificity). For moles identified as atypical, 49% were melanoma (PPV). For moles identified as normal, 87% were not melanoma (NPV). These numbers are consistent with those found in past evaluations of SSE accuracy. A review by the
U.S. Preventive Services Task Force found that SSE sensitivity ranged from 58-75% and specificity ranged from 62-98% (9).

Collective effort scores were calculated using the methods and thresholds specified in the statistical methods section. Using the 19% threshold, collective effort correctly identified 90% of melanomas, and correctly classified 72% of non-melanomas. For moles identified as atypical, 50% were melanoma. For moles identified as normal, 96% were not melanoma.

Using the 65% threshold, collective effort correctly identified 67% of melanomas and correctly identified 100% of non-melanomas. For moles identified as atypical, 100% were melanoma. For moles identified as not atypical, 91% were not melanomas.

Thus, the 19% collective effort threshold appeared to yield optimal sensitivity compared to other strategies. To determine if the observed differences between individual and collective effort were statistically significant, z-tests comparing proportions were calculated. There were substantive and statistically significant differences between individual effort sensitivity and collective effort sensitivity at the 19% threshold, $z = 12.34$, $p < .001$, as well as differences between a 19% threshold and 65% threshold, $z = 9.18$, $p < .001$, and differences between individual estimate sensitivity and a 65% collective effort threshold, $z = 2.94$, $p = .002$.

Proportions for all other dimensions—specificity, PPV, and NPV—were all significantly different as well with one exception. There was no difference between the PPV for individual effort and the 19% collective threshold.

**Discussion**

In summary, though specificity was lower (i.e., a higher false-positive rate), all other benchmarks favored collective effort. Notably, a 19% collective effort threshold was considerably more sensitive than individual effort at detecting melanomas. In this case,
sensitivity is the more important component as a false negative equates to missed melanoma whereas a false positive presumably leads the individual to schedule an appointment with a dermatologist for a clinical examination.

From a broader perspective, the results of the current study suggest that the limitations of SSE can be countered by a collective effort approach. Thus, the next step in this research is the implementation and evaluation of collective effort or crowdsourcing interventions. For example, crowdsourcing could be implemented via a web-based interface that allowed individuals to post images of their moles to receive communal feedback. Individuals could post their own images or post images with the assistance of a portable camera system. The latter could be introduced to underserved populations via portable healthcare units, such as those utilized by public health nurses (23). For example, rural populations are less likely to have access to dermatologists or healthcare professionals trained in dermoscopy, a service gap that increases melanoma mortality rates in this population (24). Yet, rural populations increasingly have access to portable healthcare units. Such a system could provide users with a more reliable means for managing their own care, encourage innovative telemedicine efforts, and nudge users toward action (25).

Yet the promise of crowdsourcing needs to be weighed against past failures. The initial hype around crowdsourcing is now wearing off, because crowdsourcing is not a magic bullet, nor is it well defined (26). For example, Lichtenthaler and Ernst found that crowdsourcing projects like InnoCentive have not lived up to the high expectations of much of the early literature (27). Chanal and Caron-Fasan have produced a rare postmortem of a failed crowdsourcing site, CrowdSpirit, a project created to allow crowds to design, test, and produce products; they found that early conceptual problems undermined this project (28). Jeff Howe himself experienced
failure with a crowdsourced journalism project, Assignment Zero, which failed to create an organizational structure to help volunteers do the work of journalism (12).

Past failures have identified several obstacles that can undermine the success of crowdsourcing interventions. For example, crowdsourcing often fails to achieve meaningful results because of problems related to scale. Scale refers to several factors including the size of the sample material the crowd will work with (e.g., the number of mole images), the size of the crowd, and the amount of material produced by the crowd (e.g., the amount and type of feedback). Raymond argued that a successful crowdsourced project needs a seedbed of material for users to work with (29). The Mars Clickworkers project (30), ReCaptcha (31), and The GoldCorp Challenge (11) all started with a huge database of objects for users to analyze. For example, in the case of Mars Clickworkers there were three years' worth of high-resolution images – over 83,000 in all – for the crowd to scour for craters (32).

Scale also refers to the size of the crowd. Scholars of open source and crowdsourced projects have frequently noted that a bigger user base translates to greater success (33). With more participation by diverse people, better solutions, ideas, and products emerge. The effects of crowd size can be seen within crowdsourced projects: Wikipedia entries with many participants tend to suffer vandalism for far shorter periods than less-trafficked articles (34). In addition, depending on the project, if a large crowd is attracted, they have to be managed, which requires staffing. 12

Finally, even with a large initial seedbed and an adequate crowd, there is the issue of vetting results. Scale can become a challenge after the crowd weighs in: what is to be done with the mass of data and ideas the crowd produces (33)? Often this judgment falls on the sponsor of the project. For example, in crowdsourcing data for disaster relief, verification of crowdsourced
geographic data and the elimination of fraud require much work on the part of relief
organizations (13). In design and idea production contests, someone has to act as judge, and with
more entries, this work is harder (12).

Thus, it is clear that there are fundamental challenges for any crowdsourcing project
related to scale: 1) having enough material for a crowd to work with; 2) recruiting and retaining a
large enough crowd to do the work; and 3) being able to vet the results, especially when the
crowd produces a mass of data or feedback.

Future work will also need to address challenges identified by past SSE research. For
example, researchers have noted that SSE is undermined by a failure to completely scan all parts
of the body (35,36). Users might be willing to upload a photo of a single mole, but that could
ultimately prove suboptimal if they are failing to monitor other parts of their body. Similarly,
Grossman and colleagues have spent the last decade studying the relationship between various
forms of mole imagery and melanoma screening accuracy (37). Their research to date suggests
that accuracy is improved by the addition of regional photographs so that individual lesions can
be viewed in the context of other lesions. For example, if a patient has a suspicious lesion on
his/her arm, then it is useful to have a photograph of the lesion (up close) and the arm (i.e., the
region). Regional photographs provide context and allow for the possible identification of
melanoma arising de novo (i.e., from normal-appearing skin).

Despite these challenges, crowdsourcing’s potential is alluring, especially for researchers
and practitioners interested in improving early melanoma detection. SSE is a visual identification
task, and some of the most effective crowdsourcing interventions have focused on the processing
of visual information (17). In fact, there is evidence that individuals are already utilizing quasi-
collective effort approaches to identify atypical moles. Past studies have found that people rely
on family members and friends when performing SSEs (10), and a search of the Internet reveals numerous people posting mole images to health forums and asking for feedback.
References


   
   


   
   2009:418–419.


Table 1.
Hypothetical Collective Effort Data

<table>
<thead>
<tr>
<th>Mole</th>
<th>% scored atypical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mole #1 (Typical)</td>
<td>3%</td>
</tr>
<tr>
<td>Mole #2 (Typical)</td>
<td>5%</td>
</tr>
<tr>
<td>Mole #3 (Typical)</td>
<td>8%</td>
</tr>
<tr>
<td>Mole #4 (Typical)</td>
<td>4%</td>
</tr>
<tr>
<td>Mole #5 (Atypical)</td>
<td>35%</td>
</tr>
</tbody>
</table>

Note. The hypothetical data in this table are meant to illustrate the concept of collective effort. Moles #1 – 4 are all typical whereas mole #5 is atypical (i.e., clinically diagnosed melanoma). Collective effort considers the pattern across all users rather than the average individual ability of a user. Thus, even though 65% of the hypothetical participants incorrectly classified the atypical mole as typical (low individual ability), the pattern of response across all users still identifies the atypical mole because a relatively large number of people scored it as atypical.
Table 2. Collective Effort Data for Mole Identification Task

<table>
<thead>
<tr>
<th>Mole</th>
<th>% scored atypical</th>
<th>Mole</th>
<th>% scored atypical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mole #1 (Typical)</td>
<td>1.40%</td>
<td>Mole #21 (Typical)</td>
<td>17.80%</td>
</tr>
<tr>
<td>Mole #2 (Typical)</td>
<td>2.90%</td>
<td>Mole #22 (Typical)</td>
<td>18.30%</td>
</tr>
<tr>
<td>Mole #3 (Typical)</td>
<td>3.50%</td>
<td>Mole #23 (Typical)</td>
<td>23.10%</td>
</tr>
<tr>
<td>Mole #4 (Typical)</td>
<td>4.80%</td>
<td>Mole #24 (Typical)</td>
<td>25.20%</td>
</tr>
<tr>
<td>Mole #5 (Typical)</td>
<td>5.20%</td>
<td>Mole #25 (Typical)</td>
<td>25.40%</td>
</tr>
<tr>
<td>Mole #6 (Typical)</td>
<td>5.20%</td>
<td>Mole #26 (Typical)</td>
<td>28.50%</td>
</tr>
<tr>
<td>Mole #7 (Typical)</td>
<td>6.10%</td>
<td>Mole #27 (Typical)</td>
<td>38.80%</td>
</tr>
<tr>
<td>Mole #8 (Typical)</td>
<td>6.40%</td>
<td>Mole #28 (Typical)</td>
<td>60.70%</td>
</tr>
<tr>
<td>Mole #9 (Typical)</td>
<td>6.60%</td>
<td>Mole #29 (Typical)</td>
<td>60.90%</td>
</tr>
<tr>
<td>Mole #10 (Typical)</td>
<td>7.40%</td>
<td>Mole #30 (Typical)</td>
<td>61.00%</td>
</tr>
<tr>
<td>Mole #11 (Typical)</td>
<td>7.60%</td>
<td>Mole #31 (Typical)</td>
<td>64.10%</td>
</tr>
<tr>
<td>Mole #12 (Typical)</td>
<td>9.50%</td>
<td>Mole #32 (Atypical)</td>
<td>11.00%</td>
</tr>
<tr>
<td>Mole #13 (Typical)</td>
<td>10.30%</td>
<td>Mole #33 (Atypical)</td>
<td>19.00%</td>
</tr>
<tr>
<td>Mole #14 (Typical)</td>
<td>11.20%</td>
<td>Mole #34 (Atypical)</td>
<td>20.50%</td>
</tr>
<tr>
<td>Mole #15 (Typical)</td>
<td>12.20%</td>
<td>Mole #35 (Atypical)</td>
<td>68.60%</td>
</tr>
<tr>
<td>Mole #16 (Typical)</td>
<td>13.00%</td>
<td>Mole #36 (Atypical)</td>
<td>70.00%</td>
</tr>
<tr>
<td>Mole #17 (Typical)</td>
<td>13.70%</td>
<td>Mole #37 (Atypical)</td>
<td>74.60%</td>
</tr>
<tr>
<td>Mole #18 (Typical)</td>
<td>15.50%</td>
<td>Mole #38 (Atypical)</td>
<td>77.00%</td>
</tr>
<tr>
<td>Mole #19 (Typical)</td>
<td>16.10%</td>
<td>Mole #39 (Atypical)</td>
<td>85.00%</td>
</tr>
<tr>
<td>Mole #20 (Typical)</td>
<td>16.40%</td>
<td>Mole #40 (Atypical)</td>
<td>92.60%</td>
</tr>
</tbody>
</table>

Note. Actual data from the mole identification task study. Atypical moles are clinically diagnosed melanomas. The percent of participants that scored a mole as atypical is listed under % scored atypical. Typical/atypical moles are presented here in numerical order based on % scored atypical. Actual mole images were presented to participants in a more random order.
Table 3. Comparing Individual SSE Performance to Collective Effort Performance

<table>
<thead>
<tr>
<th></th>
<th>Individual Effort (average)</th>
<th>Collective Effort – 19% Threshold</th>
<th>Collective Effort – 65% Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>.58</td>
<td>.90</td>
<td>.67</td>
</tr>
<tr>
<td>Specificity</td>
<td>.81</td>
<td>.72</td>
<td>1.00</td>
</tr>
<tr>
<td>PPV</td>
<td>.49</td>
<td>.50</td>
<td>1.00</td>
</tr>
<tr>
<td>NPV</td>
<td>.87</td>
<td>.96</td>
<td>.91</td>
</tr>
</tbody>
</table>

Note. \(N = 497\). Individual effort is the average ability of a single user to detect an atypical mole. A 19% threshold means that moles are only considered atypical if at least 19% of the group deems them to be.

PPV = Positive Predictive Value    NPV = Negative Predictive Value