# Use of Crowdsourcing to Diagnose Surgical Site Infections

Cameron E Gaskill

A thesis
Submitted in partial fulfillment of the
requirements for the degree of:

Masters in Public Health

University of Washington
2016

Committee:
Dr. Charles N Mock
Dr. Heather L Evans

Program Authorized to Offer Degree:

Global Health

University of Washington

## Abstract

Use of Crowdsourcing to Triage Surgical Site Infections

Cameron E Gaskill

Chair of the Supervisory Committee:
Dr. Charles N Mock, MD PhD MPH
Professor
Department of Global Health
Department of Epidemiology
Department of Surgery

Background: Surgeons can use wound photographs to monitor for surgical site infections (SSI). Mobile Health apps that allow patients to securely transmit photos to providers for remote wound surveillance are available, but this new patient generated data stream demands a new monitoring process. Crowdsourcing (employing internet users to rapidly perform discrete tasks) has potential as a scalable method for wound triage. However, the ability of the crowd to assess and triage wound photographs is unknown.

Methods: Ten case scenarios (6 SSI, 4 non-SSI), presented as 4-6 sequential days of surgical wound photographs and symptoms, were administered as a survey through Amazon Mechanical Turk, a global crowdsourcing platform. Participants provided demographics and prior experience with SSI. SSI was defined, but no additional diagnostic training was provided. For each scenario day, they indicated if they felt an SSI was present, level of confidence, and their triage recommendation for treatment or follow-up. Triage appropriateness was defined as escalation of care when SSI was suspected or non-escalation of care in the absence of SSI. SSI and non-SSI cases were analyzed to determine diagnostic accuracy and appropriateness of recommendations.

Results: 1171 participants completed the survey within 6 hours. After quality control for survey completion, data from 993 participants (3311 cases) were analyzed. 530 (53%) of participants were female, mean age was 35.5 years (SD±11.38), and 741 (74.6%) had prior experience with surgical wounds. Overall diagnostic accuracy was 34.6%; 18.1% of SSIs and 59.6% of non-SSI were correctly identified. Personal history of SSI was associated with improved accuracy in both SSI (19.8% v 16.2%,p=0.037) and non-SSI cases (62.9% v 55.8%,p=0.008). Higher levels of confidence were associated with higher accuracy in non-SSI cases (OR 1.34 [1.28-1.43],p<0.001) and lower accuracy in SSI cases (OR 0.84 [0.80-0.86],p<0.001). Triage recommendations were correct in 52.8% of cases; 45.7% of non-SSI cases were over-triaged and 57.6% of SSI cases were under-triaged.

Conclusions: Crowdsourcing without training had poor performance in the detection of SSI using wound photography. More work is required to establish standards for wound image features of SSI to develop triage training tools before crowdsourcing can be employed for SSI surveillance.

# Introduction

Surgical site infections (SSI) are the most common and costly healthcare associated infection in the US.[1] Between 160,000 and 300,000 surgical patients suffer a SSI every year, extending their hospital stays approximately 7 – 11 days and incurring a 2 – 11-fold increase in risk of mortality.[2–6] As post-operative stays shorten, more and more SSI now occur outside of the hospital, leaving the patient to diagnose wound concerns and seek appropriate treatment. Now up to 70% of SSI occur after the patient discharges from the hospital and before regular clinic follow up,[6–9] highlighting the need for improved outpatient wound surveillance.[2,10–12] Active wound surveillance has the potential to reduce SSI rates by up to 40%,[13,14] however no widely accepted system has been established.[12,15–18]

The use of mobile technology to track a patient's health is becoming increasingly popular.[19] The ubiquitous presence of smartphones with high quality cameras has made mobile health (mHealth) outpatient wound monitoring feasible[18,20–23] and allowed patients to share sufficient data for their provider to make meaningful clinical assessments. While patients do self-monitor their wounds, they are unreliable in diagnosing wound problems or infection.[24] Studies testing surgeon assessments of wound photos have shown high specificity, negative predictive values, and inter-rater agreeability, indicating that photography to be a suitable screening mechanism for SSI.[15,16,21,23] Implementation of such a system however, would require medical professionals to review daily images from all of their post-operative patients, a time consuming process that would quickly exhaust the surgeon's resources and time. Alternatively, crowdsource populations may be used to monitor large volumes of surgical wounds; however, the ability of these non-medically trained populations to identify SSI has yet to be evaluated.

The primary purpose of this study was to evaluate the ability of non-medically trained people to identify SSI through monitoring of serial wound photographs. Secondary goals included evaluating participant's subsequent confidence in SSI diagnosis, triage aptitude, and timeliness of diagnosis. We hypothesize that non-medically trained people will provide suitable diagnostic sensitivity and specificity to be used for future SSI screening efforts.

# Methods

We administered three web-based surveys containing patient scenarios with daily symptoms and photographs of wounds. Scenarios were presented as if a family member or friend recovering from surgery had called the participant with report of symptoms and an accompanying photograph. Based on this information, participants were asked to assess for the presence of a SSI. Participants were also asked to report their confidence in this assessment, and their advice to the patient on seeking treatment (i.e., management/ triage recommendation). This study was approved by the University of Washington Institutional Review Board and electronic consent was obtained from all participants.

## Participants

Survey participants were recruited via Amazon's Mechanical Turk (mTurk) platform (www.mturk.com). mTurk is a crowdsourcing Internet marketplace, consisting of a large group of users who complete un-automatable tasks for payment, such as surveys, transcription, or data processing. Human Intelligence Tasks (HIT) were created directing participants to our study consent, study instructions, and a subsequent survey. General demographics of mTurk users are well described.[25,26] For our purposes, we collected additional information on participant's age, sex, race, education level, and experience with surgery as a patient or caregiver. Participants were compensated $2.00 for completion of the survey. Participants were not permitted to repeat the same survey more than once.

## Survey Design

Three surveys were created, each with 3-4 unique patient scenarios. Scenarios were comprised of data extracted from a surgical wound photo database of hospitalized post-operative patients derived from a prior prospective cohort study dataset.[27] This database included daily photographs and logged symptoms of surgical wounds, including assessments of SSI using CDC criteria.[28] Clinical data and photographs were reviewed by physician and surgeon study personnel to ensure scenarios clearly represented the diagnosed outcome. Of the 10 total cases, 6 cases depicted wounds with SSI and 4 did not. Each scenario introduced a patient, their age, basic medical history, limited vital signs, and the procedure they had undergone. Wound photographs were displayed in a chronologic order.

Symptoms, including presence of fevers and wound drainage quality and quantity were listed with each daily photograph. Each case scenario contained between 4 and 6 days of information. (Table 1) This survey design was meant to replicate the data available to be shared by a person recovering at home after surgery. Participants were asked to indicate if the person pictured had a “surgical site infection” (yes/no) with each photograph, their confidence in this diagnosis (scale of 0-10, 0= “least confident” to 10= “most confident”), and their recommendation to the patient based on this assessment. Management recommendation options ranged from “Look at the wound again tomorrow” to “Go to the nearest Emergency Department” (Figure 1). Respondents could not go back to review previous photos or change prior answers based on additional data. After an SSI diagnosis was made, that particular scenario ended and the next scenario was started. (Figure 2) Surveys included simple quality-control questions to ensure participants were attentively completing the survey and providing quality survey responses. Incorrect answers to these quality-control questions resulted in that participant’s responses being discarded.

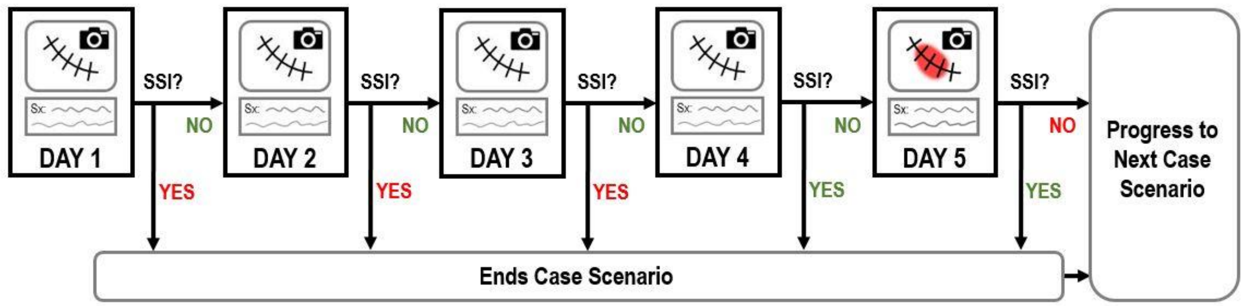
**Table 1: Survey Case Scenarios**

| Case Number | Age | Gender | Wound Classification | Overweight | Smoking | DM | Fevers | Wound drainage |
|---|---|---|---|---|---|---|---|---|
| SSI Cases | | | | | | | | |
| 1 (247) | 64 | M | Clean-cont | + | - | - | + | serous |
| 2 (304) | 71 | F | Clean-cont | - | + | - | + | - |
| 3 (346) | 69 | M | Clean-cont | - | + | - | - | - |
| 4 (201) | 55 | F | Clean-cont | - | + | - | + | purulent |
| 5 (412) | 34 | F | Clean | - | - | - | + | - |
| 6 (299) | 69 | M | Clean-cont | - | - | - | - | serous |
| | | | | | | | | |
| Non-SSI Cases | | | | | | | | |
| 1 (75) | 79 | M | Clean | + | - | + | - | - |
| 2 (853) | 45 | M | Clean-cont | - | - | - | - | serous |
| 3 (65) | 78 | M | Clean-cont | - | - | - | + | - |
| 4 (370) | 53 | F | Clean | + | - | - | - | serous |

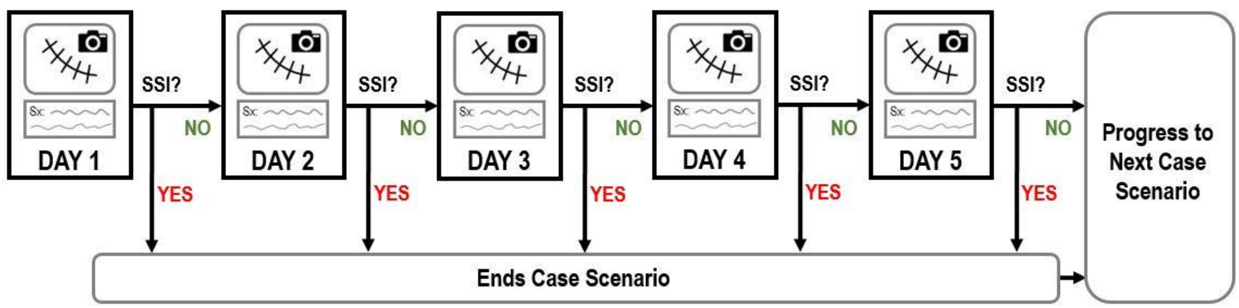
**Figure 1: Example Survey**


**Figure 2: Survey Case Scenario Design.** Green “No” and “Yes” indicate correct responses while, red “No” and “Yes” indicate incorrect responses.

**SSI Cases**




**Non-SSI Cases**




## Data Collection

Surveys were administered and data managed using REDCap electronic data capture tools hosted at the University of Washington. REDCap (redcap.iths.org) is a secure, web-based application designed to support data capture for research studies.[29] Three study HITs, each directing to one of the three surveys, were launched midweek to optimize response rate. Eligible mTurk users were selected and opened each HIT through the Amazon mTurk platform. HITs provided a brief description of the survey task before linking to the REDCap survey. The REDCap survey presented a participant consent form before reviewing survey instructions and proceeding to the patient case scenarios. No other participant recruitment was utilized. Surveys were launched at various times throughout the week to capture a variety of participants. Data was downloaded and converted for analysis in Stata (Stata v14.1, StataCorp College Station, TX).

## Study Definitions

The day of SSI diagnosis was recorded in the original dataset as the day that treatment for SSI was begun (either the day antibiotics were started or the surgical wound was reopened, whichever occurred first); this served as our gold standard for SSI diagnosis. Participant diagnosis of SSI was considered correct if it was indicated on the day or day before the gold-standard diagnosis. An incorrect diagnosis of SSI was defined as a diagnosis made prior to those two days ("early") in an SSI scenario, or a diagnosis at any time in a non-SSI scenario. In regards to management, under-triage was defined as failure to give a recommendation to seek additional evaluation in clinic or the nearest Emergency Department for cases of SSI. Over-triage of non-SSI cases or early escalation of care of SSI cases was defined as not receiving a recommendation to look at the wound tomorrow or call the primary care provider for advice.

## Analysis

Participant demographics were summarized using counts and percentages for categorical data, and means and standard deviations for normally-distributed continuous data. Overall and subgroup diagnostic accuracy were represented using percentages of correct responses. Diagnostic accuracy, confidence, and management recommendations were compared by scenario type (SSI vs non-SSI) using Student's t-test and ANOVA. Association between confidence level and accuracy was performed using univariate regression. P-values of 0.05 or less were considered significant. All analysis was completed using Stata v14.1.

## Results

A total of 1171 participants consented and completed the surveys. One hundred seventy-eight (15.2%) participants incorrectly answered quality-control questions resulting in their survey responses being excluded, leaving 993 participants and 3311 cases for further analysis. Survey completion times averaged 91.6 seconds (SD±89.4) per participant, and data collection was completed within 6 hours of survey launch. Fifty-three percent (520) of participants were female, with an average age of 35.5 years (SD±11.38). Four hundred fifty-two (45.5%) of the participants had graduated college and 109 (11%) had advanced degrees. Approximately half of participants reported experience as a surgical patient (533, 53.7%), or as a caregiver to a family member or friend while recovering from surgery (526, 56.4%). Six percent (60) were healthcare workers. (Table 2).

**Table 2: Demographics**

| | |
|---|---|
| Total participants | 993 |
| Mean time to complete survey (seconds) | 91.6 (SD±89.4) |
| | |
| Mean age (years) | 35.5 (SD±11.38) |
| Gender (% female) | 53.4% |
| Education level | |
| Some high school | 1.2% |
| High school graduate | 42.3% |
| College graduate | 45.5% |
| Master's degree | 8.2% |
| Doctoral degree | 2.8% |
| Race | |
| Asian | 12.6% |
| Black | 6.8% |
| Caucasian/ White | 71.4% |
| Hawaiian/ Pacific Islander | 0.1% |
| Hispanic/ Latino | 7.8% |
| Native America/ First Nation | 1.2% |
| Declined to answer | 0.2% |
| | |
| Surgery experience | |
| Personal history of surgery | 53.7% |
| Has cared for family member/friend recovering from surgery | 56.4% |
| Personal history of SSI | 5.5% |
| Has cared for family member/friend with SSI | 21.1% |
| | |
| Healthcare Professionals | 6.0% |
| Physician | 0.5% |
| Registered nurse | 1.1% |
| Physician assistant | 0.2% |

| | |
|---|---|
| Nurse Practitioner | 0.1% |
| Student | 1.1% |
| Other | 3.0% |

Overall diagnostic accuracy for correctly identifying SSI was 34.6%. In SSI cases, 18.1% made a correct and timely diagnosis. 39% diagnosed SSI too early, and 42.9% failed to diagnose SSI throughout the scenario. In non-SSI cases, 40.4% incorrectly diagnosed an SSI. Subgroup analysis revealed statistically higher diagnostic accuracy for participants with a personal history of surgery for both SSI and non-SSI scenarios and history of having cared for a family member or friend recovering from surgery in non-SSI scenarios. However, no difference in diagnostic accuracy was observed between education or being a healthcare professional. (Table 3)

Incorrect diagnosis in SSI cases were made with more confidence than correct diagnosis (7.3 vs 5.8, p<0.001). In non-SSI cases, confidence levels were higher in those correctly determining no infection was present (7.2 vs 5.8, p<0.001). Subgroup analysis indicated that participants with more experience (personal history of surgery, history as a caregiver, or medical professional) reported higher confidence levels across both case types, regardless of their diagnosis being correct or incorrect (Table 4a). Increasing confidence levels were not associated with improved diagnostic accuracy in the SSI scenarios (OR 0.84 95%CI 0.80-0.86, p<0.001) however increased confidence was associated with improved accuracy in the non-SSI scenarios (OR 1.34 95%CI 1.28-1.43, p<0.001). (Table 4b-c and Figures 3a-b)

**Table 3: Diagnostic accuracy; overall and among subgroups**

| | **SSI scenarios** | | **Non-SSI scenarios** | |
|---|---|---|---|---|
| Overall | 18.1% | | 59.6% | |
| | | | | |
| Education level | | | | |
| Some high school | 25.0% | | 55.6% | |
| High school graduate | 18.7% | | 58.4% | |
| College graduate | 17.7% | | 61.9% | |
| Master's degree | 14.7% | | 57.9% | |
| Doctoral degree | 23.6% | p=0.522 | 47.4% | p=0.374 |
| | | | | |
| Surgery experience | | | | |
| Personal history of surgery | 19.8% | | 62.9% | |
| No history of surgery | 16.2% | p=0.037 | 55.8% | p=0.008 |
| Has cared for family member/friend recovering from surgery | 18.5% | | 61.2% | |
| No history as a caregiver | 17.6% | p=0.591 | 57.4% | p=0.165 |
| | | | | |
| Healthcare professional | 19.0% | | 52.4% | |
| Non-healthcare professional | 18.1% | p=0.806 | 60.1% | p=0.166 |

**Table 4a: Reported confidence levels; correct vs incorrect responses (mean on scale 0-10)**

| | **SSI scenarios** | | | **Non-SSI scenarios** | | |
|---|---|---|---|---|---|---|
| | Correct Diagnosis | Incorrect Diagnosis | | Correct Diagnosis | Incorrect Diagnosis | |
| Overall | 5.8 (SD±2.1) | 7.3 (SD±2.2) | p<0.001 | 7.2 (SD±2.1) | 5.8 (SD±2.1) | p<0.001 |
| | | | | | | |
| Surgery experience | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Personal history of surgery | 5.8 (SD±2.1) | 6.8 (SD±2.2) | p<0.001 | 7.4 (SD±2.0) | 5.7 (SD±2.1) | p<0.001 |
| No history of surgery | 5.8 (SD±2.0) | 6.5 (SD±2.3) | p=0.001 | 7.1 (SD±2.2) | 5.9 (SD±2.1) | p<0.001 |
| | | | | | | |
| Has cared for family member/friend recovering from surgery | 5.5 (SD±1.9) | 6.8 (SD±2.2) | p<0.001 | 7.5 (SD±1.9) | 5.8 (SD±2.1) | p<0.001 |
| No history as a caregiver | 6.2 (SD±2.2) | 6.5 (SD±2.7) | p=0.146 | 6.9 (SD±2.3) | 5.8 (SD±2.2) | p<0.001 |
| | | | | | | |
| Healthcare professional | 6.1 (SD±1.9) | 7.2 (SD±2.0) | p=0.018 | 8.2 (SD±1.5) | 6.5 (SD±2.1) | p<0.001 |
| Non healthcare professional | 5.8 (SD±2.1) | 6.6 (SD±2.2) | p<0.001 | 7.2 (SD±2.1) | 5.8 (SD±2.1) | p<0.001 |

**Table 4b: Accuracy by reported confidence**

| Reported Confidence | SSI scenarios | Non-SSI scenarios |
|---|---|---|
| 0 | 12.5% (±24.3) | 33.3% (±41.3) |
| 1 | 33.3% (±18.1) | 40.0% (±22.0) |
| 2 | 22.4% (±10.8) | 41.2% (±16.8) |
| 3 | 28.1% (±8.0) | 40.9% (±12.0) |
| 4 | 18.0% (±6.4) | 36.6% (±10.5) |
| 5 | 29.1% (±5.6) | 40.1% (±7.2) |
| 6 | 20.0% (±4.1) | 50.0% (±6.9) |
| 7 | 18.0% (±4.2) | 63.6% (±6.4) |
| 8 | 13.3% (±4.1) | 68.8% (±6.5) |
| 9 | 9.0% (±3.7) | 82.3% (±5.7) |
| 10 | 8.3% (±3.9) | 80.6% (±6.6) |

**Table 4c: Accuracy by confidence cut-offs**

| Reported Confidence | SSI scenarios | | Non-SSI scenarios | |
|---|---|---|---|---|
| Greater than 6 | 14.8% | | 67.9% | |
| Less than 6 | 25.7% | P<0.001 | 39.5% | P<0.001 |
| | | | | |
| Greater than 7 | 12.9% | | 72.7% | |
| Less than 7 | 23.6% | P<0.001 | 43.1% | P<0.001 |
| | | | | |
| Greater than 8 | 10.4% | | 76.6% | |
| Less than 8 | 22.2% | P<0.001 | 48.7% | P<0.001 |
| | | | | |
| Greater than 9 | 8.7% | | 81.5% | |
| Less than 9 | 20.7% | P<0.001 | 52.7% | P<0.001 |

**Figure 3a: SSI Cases**: Accuracy (%) vs Confidence level (0-10)

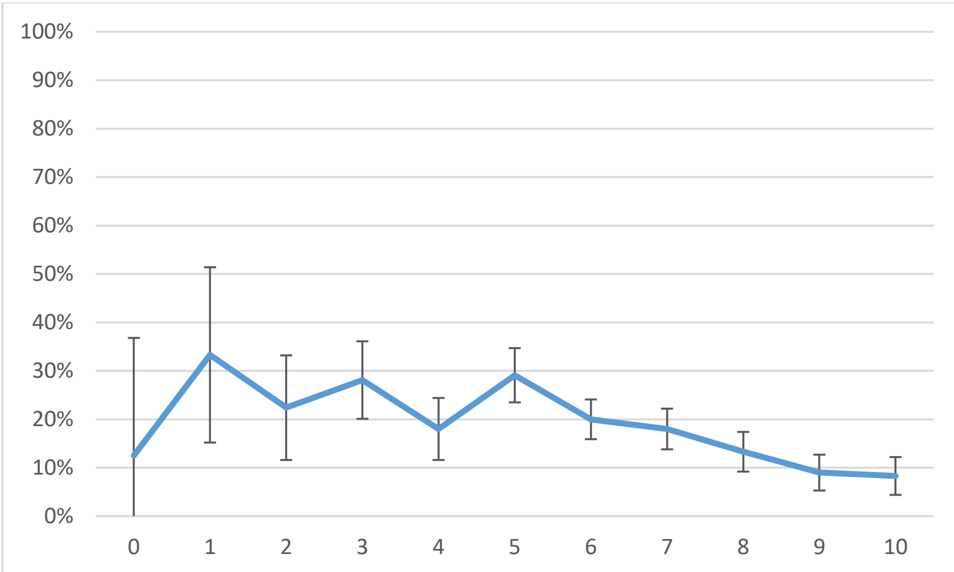
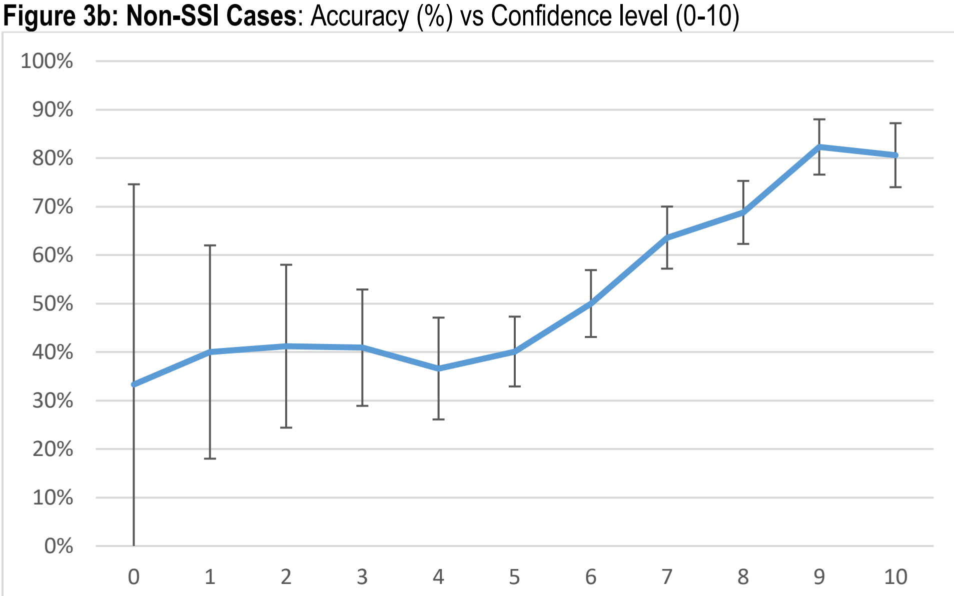



**Figure 3b: Non-SSI Cases**: Accuracy (%) vs Confidence level (0-10)




Overall, appropriate triage was recommended 27.4% of all cases; 32.9% correct in SSI cases and 19.0% correct in non-SSI cases. Over-triage occurred in 45.7% of non-SSI cases and under-triage in 57.6% of SSI cases regardless of correct diagnosis. (Table 5a) Healthcare professionals had on average higher proportion of over-triaging and under-triaging as compared to non-healthcare professionals (58.3% vs 44.8%, p=0.016 over-triage, 64.7% vs 57.1%, p=0.11 under-triage). A personal history of surgery was associated with lower occurrence of over-triaging (43.1% vs 48.6%, p=0.045). Early escalation of care specific to the SSI cases, such as presenting to clinic or the Emergency Room more than 2 days before the correct time of SSI diagnosis, was recommended by 42.6% of participants, being least common among non-healthcare professionals (41.8% vs 56.0%, p=0.003).

Management recommendation concordance with diagnosis (i.e. the correct management recommendation for the participant indicated diagnosis) was correct 50.7% of the time. Over all cases regardless of true diagnosis, the appropriate level of management was recommended in 50.3% of cases given a SSI diagnosis, and 51.1% of cases not given an SSI diagnosis. Management recommendation concordance to the given diagnosis was no different among those with a personal history of surgery, experience as a caretaker, or medical professionals. (Table 5b)

**Table 5a: Management Recommendations**

| | Undertreatment of SSI | | Overtreatment of Non-SSI | | Overtreatment of SSI | |
|---|---|---|---|---|---|---|
| Overall | 57.6% | | 45.7% | | 42.6% | |
| | | | | | | |
| Surgery experience | | | | | | |
| Personal history of surgery | 57.8% | | 43.1% | | 40.8% | |
| No history of surgery | 57.3% | p=0.844 | 48.6% | p=0.045 | 44.8% | p=0.072 |
| | | | | | | |
| Has cared for family member/friend recovering from surgery | 57.2% | | 44.6% | | 41.9% | |
| No history as a caregiver | 58.0% | p=0.694 | 47.1% | p=0.356 | 43.6% | p=0.440 |
| | | | | | | |
| Healthcare professional | 64.7% | | 58.3% | | 56.0% | |
| Non-healthcare professional | 57.1% | p=0.110 | 44.8% | p=0.016 | 41.8% | p=0.003 |

**Table 5b: Management recommendation concordance:** participant treatment recommendation appropriateness for their reported diagnosis

| | Any diagnosis | | SSI diagnosed | | SSI not diagnosed | |
|---|---|---|---|---|---|---|
| Overall | 50.7% | | 50.3% | | 51.1% | |
| | | | | | | |
| Surgery experience | | | | | | |
| Personal history of surgery | 51.0% | | 51.6% | | 50.5% | |
| No history of surgery | 50.4% | p=0.708 | 49.0% | p=0.292 | 51.8% | p=0.595 |
| | | | | | | |
| Has cared for family member/friend recovering from surgery | 51.0% | | 51.5% | | 50.5% | |
| No history as a caregiver | 50.3% | p=0.687 | 48.8% | p=0.275 | 51.8% | p=0.602 |
| | | | | | | |
| Healthcare professional | 49.0% | | 50.0% | | 41.9% | |
| Non-healthcare professional | 50.8% | p=0.618 | 54.4% | p=0.370 | 51.6% | p=0.080 |

# Discussion

Surgical site infections are a leading cause of post-operative morbidity and mortality in the United States, affecting up to 300,000 people annually and incurring up to $10 billion in healthcare expenditures. Post-operative wound monitoring is increasingly difficult to accomplish given the trend towards shorter hospital stays and no established post-discharge wound surveillance system. One solution currently being explored is the use of wound photographs to help link objective data from a patient's outpatient recovery to provider evaluation. Large scale implementation of such a photographic surveillance program would generate a new patient-generated health datastream, creating a new demand for surgical outpatient care monitoring and potentially overwhelming an established healthcare system designed for in-person evaluations. Even with increasing use of telemedicine, no standards have been developed for the remote evaluation of surgical wounds for evidence of infection. Regardless of the method of post-discharge evaluation by surgical care providers, the standard for escalating post-discharge surgical care still relies on the patient's (or caregiver's) ability to recognize an evolving problem and initiate contact, typically via telephone triage or direct in-person evaluation through an unscheduled clinic or emergency room visit.

As such, the feasibility of SSI detection by lay people with various levels of prior experience with SSI is a relevant problem.

This study is the first to investigate photographic wound evaluation utilizing crowdsourcing. These findings demonstrated a poor ability of untrained crowds to diagnose SSI using photos, with a sensitivity and specificity of 18.1% and 59.6%, respectively. Aside from those with a prior personal history of SSI, none of the subpopulations analyzed were observed to have improved diagnostic accuracy and no subpopulation had a diagnostic accuracy greater than 50%. This suggests experience with surgical wounds did not confer an improved ability of non-healthcare professionals to diagnose SSI using photographic findings. This study also observed no difference in the diagnostic accuracy between self-reported healthcare professions and non-healthcare professions. Sanger et al. demonstrated a diagnostic accuracy of 76% among SSI experts, compared to the 34.6% demonstrated here.[30] Future studies may help to evaluate how non-expert surgeons will perform on such tasks as little is known how other front-line practitioners, such as ER physicians, Family Physicians, and others will be able to utilize wound photography for triage or diagnosis. Further large scale studies of surgeons and other medical professionals are required.

These results must be interpreted with the structure of our survey in mind. Typically, sensitivities and specificities are calculated for a two-option question (i.e. disease or no disease) and compared against 50% odds of being correct. As participants were asked to evaluate multiple days of a patient's wound, the reported sensitivity and specificity must be compared against the complex probability of arriving at the correct answer by chance (i.e. random guessing). If these surveys were answered in a random fashion we would expect an overall diagnostic accuracy (i.e. the chance of arriving at the correct answer after randomly guessing) of 6.24% with a sensitivity and specificity of 8.55% and 2.74%, respectively. Comparing this to our diagnostic accuracy result of 34.6% and sensitivity and specificity of 18.1% and 59.6%, reveals that participants performed much better than random guesses. Regardless, a sensitivity of only 18.1% means that over 80% of SSI cases failed to be detected; an unacceptable level for use in clinical practice. While our analysis classified "early" diagnosis of SSI as incorrect, one could consider including these as correctly identifying patients with SSI as doing so would ensure additional evaluation. By expanding our analysis to include the "early" diagnoses as correct, the sensitivity increases to 57%. While this is greatly improved, it still does not provide the level of accuracy needed for such a screening activity.

The use of wound photography to monitor for SSI has been explored with several inpatient studies of vascular[21,23,31] and abdominal surgery[16] patients. All studies compared physician wound evaluation at bedside to physician-generated photographs, finding moderate agreement (65-76%) in the diagnosis of SSI between the two methods, and sensitivities and specificities ranging 42-71% and 65-97%, respectively. While these figures are significantly higher than those observed in this study, the subgroup of medical professional in this study did equally poorly. The authors of prior wound photography studies mention the importance of training physicians on using photographs to diagnosis SSIs. The participants in this study did not receive any such training past the simple definitions of a SSI, likely contributing to the observed poor performance. Other crowdsourcing studies have demonstrated the use of non-medical professionals to review medical data and images after simple training didactic. Lendvay et al.[32] demonstrated that non-medical professionals could be trained to view pathology slides to diagnose bladder cancer with an accuracy of 92%. Likewise, crowdsourcing has been used for assessment of surgical technique,[33,34] electrocardiogram interpretation,[35] fundus exam,[36] and assessing for tumor markers.[37] Further investigation is needed to determine what elements are required in a training curriculum for non-medical professionals to achieve adequate diagnostic performance pertaining to SSI.

The confidence in diagnosis reported by participants in this study was higher for correct response in non-SSI cases. Furthermore, increased reported confidence was associated with improved diagnostic accuracy in correctly identifying non-SSI cases, indicating a potential method to subdivide response data to improve the accuracy of ruling out a potential SSI. Unfortunately, level of confidence did not appear to be as useful in the SSI scenarios, with higher confidence levels associated with incorrect responses. Wirthlin et al.[21] evaluated physician's diagnostic confidence in remote assessments of wounds, finding high reported certainty across all cases regardless of accuracy. This is consistent with our findings in that a participant's confidence in their answer had no relation to the accuracy of diagnosis, and in the subgroup of healthcare professionals, was actually associated with poorer accuracy in SSI cases.

Management recommendations were correct in only 27% of cases, with almost half of all non-SSI cases directed to seek treatment and 57% of SSI not directed to seek treatment. This is in contrast to prior research that demonstrated physicians making appropriate treatment recommendations in 66-92% of cases.[21] However, our findings indicated a large discordance between presumed diagnosis and management recommendation, with only 50% of participant's selected diagnoses receiving appropriate care, likely hindering the appropriateness of management recommendation. This likely represents a lack of knowledge of appropriate management for SSIs, that could possibly be improved with training didactics.

Prior to drawing conclusions from this study, several limitations must be discussed. First, the quality of collected data is potentially biased due to our recruitment via Amazon Mechanical Turk. While quality control questions aimed to screen out unthoughtful responses, it is unclear how well these questions ensure participants give their full attention to the survey, as the average time to survey completion was less than 2minutes. The broad demographics of the mTurk population provides a good baseline generalizability of the study, however alternative recruitment populations varying in expertise may be better suited for our tasks. While our participants were younger than the general population, this group is a good representation of potential persons available to participate in crowdsourcing tasks. Second, most SSI scenarios had the wounds opened on the day the SSI diagnosis was made. As participants would recognize such a change in the wound photos, assessment and quantification of late diagnoses was not possible. Third, our photo database source included a majority of light-skinned patients, limiting the generalizability of our findings to decision making using photos of wounds of darker skin tones. Fourth, as most of the SSI scenarios had wounds opened the day after SSI diagnosis (thus indicating the diagnosis), we were only able to assess a missed diagnosis and could not quantify a delay in diagnosis. While this is sufficient in quantifying the occurrence of missed diagnosis, it did not allow assessment of the degree of delay in diagnosis. As the major concern of any surveillance system is missed diagnoses, information on the degree of delay would have added meaningful detail to our findings. Finally, the photos used were not patient generated and provide a less than pragmatic view of how an outpatient wound surveillance program would function. The differences between patient generated and physician generated photos are not known, however the quality of photo has been shown to not affect diagnostic accuracy.[23]

Despite these limitations, this study highlights the possibility of using wound photography for outpatient SSI surveillance. Currently, the Center for Disease Control does not have defined criteria or guidelines for the use of wound photography to diagnose SSI. As the use of telemedicine and mHealth for post-operative care increases, in-person clinical encounters may be augmented or replaced, and it is imperative that SSI evaluative standards are established. Expert consensus on specific wound photo features relevant to SSI is a starting place for future prospective evaluation of the role of serial wound photography in SSI monitoring and surveillance. Future crowdsource participants will benefit from photo based SSI criteria, allowing for directed training and educational programs on how to identify, diagnose and triage SSI.

## Conclusions

This study suggests poor diagnostic accuracy for surgical site infection in the assessment of wound photographs by untrained people. While crowdsourcing is a promising method for analysis of large amounts of medical data, its use in monitoring photographs of post-operative wounds for signs of developing infection has yet to be validated. Additional work is needed to better understand how best to train and apply crowdsourcing for outpatient surveillance and detection of SSI.

## Acknowledgments/ Funding

Institute of Translational Health Science (ITHS) grant support (UL1TR000423 from NCRR/NIH).
I would like to acknowledge my mentors and co-investigators on this project: Kristy Kummerow Broman, MD, MPH, Ross J. Lordon, BS, Patrick C. Sanger, PhD, Julie B. Cooper, MPA, William B. Lober, MD, Benjamin K. Poulose, MD MPH, and Heather L. Evans, MD MS.
I would also like to acknowledge the generous feedback from the mPOWEr research group and Dr. Gabrielle Van Ramshorst who provided access to the dataset that formed the basis of the case scenarios for the surveys. Dr. Gaskill had full access to all study data and takes full responsibility for the integrity of the data and the accuracy of the data analysis.

## References


1. Zimlichman E, Henderson D, Tamir O, Franz C, Song P, Yamin CK, et al. Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system. JAMA Intern Med. 2013;173: 2039–2046. doi:10.1001/jamainternmed.2013.9763
2. Anderson DJ, Podgorny K, Berríos-Torres SI, Bratzler DW, Dellinger EP, Greene L, et al. Strategies to Prevent Surgical Site Infections in Acute Care Hospitals: 2014 Update. Infect Control Hosp Epidemiol. 2014;35: 605–627. doi:10.1086/676022
3. Mahmoud NN, Turpin RS, Yang G, Saunders WB. Impact of surgical site infections on length of stay and costs in selected colorectal procedures. Surg Infect. 2009;10: 539–544.
4. Cruse PJ, Foord R. The epidemiology of wound infection. A 10-year prospective study of 62,939 wounds. Surg Clin North Am. 1980;60: 27–40.
5. Kirkland KB, Briggs JP, Trivette SL, Wilkinson WE, Sexton DJ. The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. Infect Control Hosp Epidemiol. 1999;20: 725–730. doi:10.1086/501572
6. Perencevich EN, Sands KE, Cosgrove SE, Guadagnoli E, Meara E, Platt R. Health and Economic Impact of Surgical Site Infections Diagnosed after Hospital Discharge. Emerg Infect Dis. 2003;9: 196–203. doi:10.3201/eid0902.020232
7. Kazaure HS, Roman SA, Sosa JA. Association of postdischarge complications with reoperation and mortality in general surgery. Arch Surg. 2012;147: 1000–1007.
8. Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR, The Hospital Infection Control Practices Advisory Committee. Guideline for Prevention of Surgical Site Infection, 1999. Infect Control Hosp Epidemiol. 1999;20: 247–280. doi:10.1086/501620
9. Sands K, Vineyard G, Platt R. Surgical site infections occurring after hospital discharge. J Infect Dis. 1996;173: 963–970.
10. Kent P, McDonald M, Harris O, Mason T, Spelman D. Post-discharge surgical wound infection surveillance in a provincial hospital: follow-up rates, validity of data and review of the literature. ANZ J Surg. 2001;71: 583–589.
11. Martone WJ, Nichols RL. Recognition, prevention, surveillance, and management of surgical site infections: introduction to the problem and symposium overview. Clin Infect Dis. 2001;33: S67–S68.
12. Petherick ES, Dalton JE, Moore PJ, Cullum N. Methods for identifying surgical wound infection after discharge from hospital: a systematic review. BMC Infect Dis. 2006;6: 170. doi:10.1186/1471-2334-6-170
13. Haley RW. The scientific basis for using surveillance and risk factor data to reduce nosocomial infection rates. J Hosp Infect. 1995;30 Suppl: 3–14.
14. Condon RE, Schulte WJ, Malangoni MA, Anderson-Teschendorf MJ. Effectiveness of a surgical wound surveillance program. Arch Surg Chic Ill 1960. 1983;118: 303–307.
15. Whitby M, McLaws M-L, Collopy B, Looke DFL, Doidge S, Henderson B, et al. Post-discharge surveillance: can patients reliably diagnose surgical wound infections? J Hosp Infect. 2002;52: 155–160. doi:10.1053/jhin.2002.1275
16. van Ramshorst GH, Vrijland W, van der Harst E, Hop WCJ, den Hartog D, Lange JF. Validity of Diagnosis of Superficial Infection of Laparotomy Wounds Using Digital Photography: Inter- and Intra-observer Agreement Among Surgeons. Wounds Compend Clin Res Pract. 2010;22: 38–43.
17. Mead PB, Pories SE, Hall P, Vacek PM, Davis JH, Gamelli RL. Decreasing the incidence of surgical wound infections. Validation of a surveillance-notification program. Arch Surg Chic Ill 1960. 1986;121: 458–461.

18. Martínez-Ramos C, Cerdán MT, López RS. Mobile phone–based telemedicine system for the home follow-up of patients undergoing ambulatory surgery. Telemed E-Health. 2009;15: 531–537.
19. Klasnja P, Pratt W. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. J Biomed Inform. 2012;45: 184–198. doi:10.1016/j.jbi.2011.08.017
20. Semple JL, Sharpe S, Murnaghan ML, Theodoropoulos J, Metcalfe KA. Using a Mobile App for Monitoring Post-Operative Quality of Recovery of Patients at Home: A Feasibility Study. JMIR MHealth UHealth. 2015;3: e18. doi:10.2196/mhealth.3929
21. Wirthlin DJ, Buradagunta S, Edwards RA, Brewster DC, Cambria RP, Gertler JP, et al. Telemedicine in vascular surgery: feasibility of digital imaging for remote management of wounds. J Vasc Surg. 1998;27: 1089–1100.
22. Armstrong KA, Semple JL, Coyte PC. Replacing Ambulatory Surgical Follow-Up Visits With Mobile App Home Monitoring: Modeling Cost-Effective Scenarios. J Med Internet Res. 2014;16. doi:10.2196/jmir.3528
23. Wiseman JT, Fernandes-Taylor S, Gunter R, Barnes ML, Saunders RS, Rathouz PJ, et al. Inter-rater agreement and checklist validation for postoperative wound assessment using smartphone images in vascular surgery. J Vasc Surg Venous Lymphat Disord. 2016;4: 320–328.e2. doi:10.1016/j.jvsv.2016.02.001
24. Seaman M, Lammers R. Inability of patients to self-diagnose wound infections. J Emerg Med. 1991;9: 215–219.
25. Huff C, Tingley D. “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. Res Polit. 2015;2: 2053168015604648.
26. Demographics of Mechanical Turk: Now Live! (April 2015 edition) | A Computer Scientist in a Business School [Internet]. [cited 29 Sep 2016]. Available: http://www.behind-the-enemy-lines.com/2015/04/demographics-of-mechanical-turk-now.html
27. van Ramshorst GH, Vos MC, den Hartog D, Hop WCJ, Jeekel J, Hovius SER, et al. A comparative assessment of surgeons’ tracking methods for surgical site infections. Surg Infect. 2013;14: 181–187. doi:10.1089/sur.2012.045
28. Centers for Disease Control. Surgical Site Infection Event [Internet]. 2016 [cited 5 Aug 2016]. Available: http://www.cdc.gov/nhsn/PDFs/pscManual/9pscSSIcurrent.pdf
29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42: 377–381. doi:10.1016/j.jbi.2008.08.010
30. Sanger PC, Hartzler A, Lordon RJ, Armstrong CA, Lober WB, Evans HL, et al. A patient-centered system in a provider-centered world: challenges of incorporating post-discharge wound data into practice. J Am Med Inform Assoc JAMIA. 2016;23: 514–525. doi:10.1093/jamia/ocv183
31. Murphy RX, Bain MA, Wasser TE, Wilson E, Okunski WJ. The reliability of digital imaging in the remote assessment of wounds: defining a standard. Ann Plast Surg. 2006;56: 431–436. doi:10.1097/01.sap.0000202146.92893.6a
32. Lendvay TS. Optical Biopsy of Bladder Cancer Using Crowd-Sourced Assessment. JAMA Surg. 2016;151: 88. doi:10.1001/jamasurg.2015.3225
33. Kowalewski TM, Comstock B, Sweet R, Schaffhausen C, Menhadji A, Averch T, et al. Crowd-Sourced Assessment of Technical Skills for Validation of Basic Laparoscopic Urologic Skills Tasks. J Urol. 2016;195: 1859–1865. doi:10.1016/j.juro.2016.01.005
34. Powers MK, Boonjindasup A, Pinsky M, Dorsey P, Maddox M, Su L-M, et al. Crowdsourcing Assessment of Surgeon Dissection of Renal Artery and Vein During Robotic Partial Nephrectomy: A Novel Approach for Quantitative Assessment of Surgical Performance. J Endourol. 2016;30: 447–452. doi:10.1089/end.2015.0665

35. Nyotowidjojo I, Erickson RP, Lee KS. Crowd-Sourcing Syncope Diagnosis: Mobile Smartphone ECG Apps. Am J Med. 2016;129: e17–e18. doi:10.1016/j.amjmed.2015.11.022
36. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP. Rapid Grading of Fundus Photographs for Diabetic Retinopathy Using Crowdsourcing. J Med Internet Res. 2014;16: e233. doi:10.2196/jmir.3807
37. dos Reis FJC, Lynn S, Ali HR, Eccles D, Hanby A, Provenzano E, et al. Crowdsourcing the general public for large scale molecular pathology studies in cancer. EBioMedicine. 2015;2: 679–687.