

Spring 2014

An Exploration of Mechanical Turk as a Feasible Recruitment Platform for Cancer Survivors

Alaina Carr

University of Colorado Boulder

Follow this and additional works at: http://scholar.colorado.edu/honr_theses

Recommended Citation

Carr, Alaina, "An Exploration of Mechanical Turk as a Feasible Recruitment Platform for Cancer Survivors" (2014). *Undergraduate Honors Theses*. Paper 59.

This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.

An Exploration of Mechanical Turk as a Feasible Recruitment Platform for Cancer Survivors

Alaina Carr
Department of Psychology and Neuroscience
University of Colorado at Boulder

Thesis Advisor: Dr. Joanna Arch
Department of Psychology and Neuroscience

Dr. Rolf Norgaard, Program for Writing and Rhetoric
Dr. Richard Olson, Department of Psychology and Neuroscience (Honors Representative)

April 4, 2014

ABSTRACT

Objective: The recruitment of cancer survivors for participation in psychosocial studies can be challenging. Recruitment can be slow and costly, sample sizes are often small, and retention rates are low. These challenges are particularly pronounced in recruiting young adult cancer survivors (e.g. ≤ 40 years of age). One possible solution to this problem is to assess the feasibility and reliability of using Amazon's Mechanical Turk (MTurk), an online marketplace, as a survey research recruitment tool for cancer survivors. No known research to date has assessed the utility of MTurk as a recruitment tool for cancer populations. The present study seeks to address this gap in research by assessing the feasibility and validity of MTurk as a recruitment platform for cancer survivors in general and for young adult cancer survivors in particular. An additional goal is to assess cancer survivors general psychiatric symptoms and use of formal psychosocial support and the degree to which the need for such support is linked to their experience of having cancer.

Methods: During a 3-week period, a U.S. sample of cancer survivors ($n = 166$ total, $n = 146$ who fully completed the survey), defined as persons with cancer or a history of cancer (of any type), were recruited on MTurk to complete a series of questionnaires relating to cancer and general psychosocial functioning. The first survey assessed the presence of U.S. cancer survivors on MTurk and to determine the extent to which cancer survivors could be recruited quickly and at low cost compared to traditional recruitment methods and whether their responses were valid and reliable. One week after completing the first survey, participants were re-contacted through email to complete a second survey.

Results: 166 participants consented to our first survey and 146 of those 166 participants fully completed the survey. Two participants provided insufficient cancer type (i.e. No cancer)

and 1 participant reported residency outside of the U.S. We additionally assessed the reliability and validity of participant reporting on sociodemographic and cancer type items across both surveys and found 9 participants provided inconsistent information. We examined whether participants may have fabricated responses by using the MMPI-2 $F(p)$ scale and found 23 participants scored high on this measure of malingering. These participants were excluded from analyses resulting in a final Survey 1 sample of 111 participants. When assessing reliability and validity of participant reporting, we found for the majority of participants (88.75%), geographic and cancer type reporting was honest and consistent across both surveys. Participants on MTurk had low non-response error in fully completing the first and second surveys (87.96%; 89.89%) and total scores of questionnaires (AAQ-II, PHQ-9, & BEAQ) administered across both surveys were highly correlated ($r = .83$; $r = .78$; $r = .85$) suggesting adequate test-retest reliability. Our findings indicated there is a particularly strong presence of younger cancer survivors (Median age = 38 years). Breast, cervical, melanoma, ovarian, skin cancer, and lung cancer were the most commonly represented cancers within our sample. Additionally, we assessed feasibility of collecting data longitudinally and found 61.00% of participants ($n = 89$) responded to a second survey sent out one week after the first survey.

Conclusion: This study demonstrated that MTurk can be used with relative success as a survey research recruitment tool for cancer survivors, particularly for one-time surveys. Data was collected quickly (< 1month), at a relatively low-cost (< \$2.00/participant), and across a broad geographic range within the U.S. Cancer survivor respondents were younger on average than the national norm, suggesting that MTurk might represent a particularly viable

recruitment strategy for young adult cancer populations (≤ 40 years old). *Keywords:* Cancer Survivors, Amazon's Mechanical Turk, Cancer Recruitment, Internet

INTRODUCTION

In 2012, Americans living with a history of cancer - that is, cancer survivors - numbered nearly 14 million and by 2022, this number will increase to nearly 18 million (Siegal et al., 2012). The cancer survivor community will continue to grow due to increased early detection, improvement in medical treatments, and the aging U.S. population (Pollack et al., 2005; Stanton, 2012). Medical and psychosocial research and treatment centers will expand to address the challenges cancer survivors face during life beyond cancer diagnosis and treatment, which include problems with interpersonal relationships, feelings of alienation or isolation, fear of reoccurrence or death, anxiety and depression, and long-term treatment side effects (Hewitt, Rowland, & Yancik, 2003; Stanton, 2012).

Despite these efforts researchers still face many challenges in recruiting cancer survivors for psychosocial research. Challenges include difficulty locating cancer survivors, lack of institutional commitment (lack of available staffing or time committed to the IRB process and subsequent study), lack of patient interest, and poor retention rates (Ganz et al., 2009). Young adults (< 40 years of age) have been particularly under-represented in cancer survivorship research due to difficulty recruiting this mobile population (Rabin, Horowitz, & Marcus, 2012; Siegal et al., 2012; Stanton, 2012). Yet young age is the steadiest demographic predictor of poor quality of life, poor emotional functioning, and unmet needs among cancer survivors (Rabin et al., 2012; Siegal et al., 2012; Stanton, 2012). As the population living with a history of cancer continues to grow, the formation and implementation of evidence-based methods for promoting the health and well-being of

cancer survivors are critical (Stanton, 2012). In sum, finding effective recruitment methods for cancer survivors in general and young adult cancer survivors in particular, represents a critical challenge to psycho-oncology research.

Current outreach mediums for recruitment utilized by behavioral cancer studies include: cancer registries, mass media advertising (newspapers, newsletters, flyers, radio, brochures, and television), online advertising (e-mail, search engines, affiliate websites, and online communities), telephone-based recruitment, clinical-based recruitment, and other outreach efforts through the community including partnerships with churches and cancer support groups, word of mouth, use of direct mail (Rabin, et al., 2012; Stanton et al., 2013). Research using these recruitment methods tends to yield samples that are not representative of the larger population of adults diagnosed with cancer, which can result in inadequate recruitment as a major rate-limiting step in behavioral research on cancer survivors (Rabin et al., 2012; Stanton et al., 2013). For example, Stanton and colleagues (2013), in recruiting for three nationwide psychoeducational trials for cancer patients, demonstrated that cancer information programs (i.e. Cancer Information Service and American Cancer Society) and cancer registries of individuals open to participating in research (i.e. Avon Army of Women program) served as worthwhile resources for prostate and breast cancer patient recruitment with restrictive eligibility criteria. Yet they indicated these recruitment resources yielded a participant profile characterized by high education (college graduate and above), high family income, and a particularly high percentage of non-Hispanic whites. They found that these demographic characteristics are representative of the Cancer Information Service, American Cancer Society, and Avon Army of Women pools but are not representative of the larger prostate and breast cancer survivor populations. Stanton and colleagues (2013) noted that

other resources for recruitment are needed if attaining greater sociodemographic diversity in cancer patient accrual is a research goal.

As noted, recruiting young adult cancer survivors has proven particularly challenging. Approximately 70,000 individuals in their twenties and thirties are diagnosed with cancer each year, yet are dramatically underrepresented in cancer survivorship research, and face significant medical and psychosocial challenges. In a recent study for a web-based physical activity intervention, Rabin, Horowitz, and Marcus (2012) found that in-person recruitment for young adult cancer survivors at an oncology clinic yielded the greatest percentage of participants enrolled for an exercise-based intervention (per cancer survivor approached): 38% ($n = 13$) of those approached agreed to be screened ($n = 5$) resulting in 1 enrolled participant (8% of those approached). However, when considering the relative yield of each strategy (absolute number of young adult cancer survivors enrolled), they found that mailings appeared to be the most productive strategy with a yield of 8 enrollees out of 770 mailings. Both strategies (in person and mass mailing), however took significant time and resources and recruitment was restricted to a local sample. Noted recommendations for recruitment and retention improvement included attending to convenience issues to make participation easier and utilizing resources of low-cost to reach a large number of eligible participants (Rabin et al., 2012; Stanton et al., 2013).

One potentially viable alternative for addressing several of these challenges in the recruitment of cancer survivors is the utilization of novel internet-based recruitment sources. Psychosocial research has begun to be conducted specifically on Amazon's Mechanical Turk (MTurk), a relatively new crowdsourcing site (founded in 2005) with access to one of the largest, stable, and diverse subject pool for low cost experiments (Mason & Suri, 2012). To

our knowledge, MTurk has not yet been explored as a potential research recruitment site for cancer survivors.

MTurk has already been used in several online studies involving behavioral and social research (Mason & Suri, 2012). There have been at least two studies demonstrating that the behavior of subjects on MTurk is comparable to the behavior of laboratory subjects (Mason & Suri, 2012; Paolacci, Chandler, & Ipeirotis, 2010). Paolacci, Chandler, and Ipeirotis (2010), for example, conducted a behavioral study comparing the quality of data provided by subjects recruited in online labor markets (MTurk & Internet Discussion Boards) to offline methods of recruiting subjects (a Midwestern University subject pool). Their demographic data suggests that MTurk workers are at least as representative of the U.S. population as traditional subject pools, with gender, race, age, and education of Internet samples corresponding to the population more closely than college undergraduate samples and other internet based samples in general. One additional concern to researchers conducting web-based experiments is that unsupervised subjects tend to be less attentive than subjects in a lab with an experimenter. To address this concern, they used a catch trial to identify which subjects failed to pay close attention provided to the survey. They found subjects in the three subject pools (MTurk, the Midwestern university sample, & Internet Discussion Boards) did not differ in terms of attention provided to the survey. The MTurk sample had the lowest catch trial failing rate (4.17%) compared to the Midwestern University sample (6.47%) and Internet Discussion Board sample (5.26%) (Paolacci, et al., 2010). Additionally, MTurk subjects were more likely to have a low non-response error (i.e. subjects that accessed and fully completed the study) compared to subjects from Internet Discussion Boards (91.6% and 66.7% respectively) suggesting MTurk strongly diminished

the potential for non-response error in online research (Paolacci et al, 2010) Paolacci and colleagues (2010) findings confirmed that MTurk participants yield reliable data.

Compared to social psychological studies on MTurk, studies on MTurk for clinical and health psychology research are not as prevalent. Shapiro, Chandler, and Mueller (2013) provide clinical psychology researchers with the first study to date examining the utility of MTurk for conducting research on psychopathology. Their longitudinal research design also assessed the reliability of participant reporting by comparing responses to demographic items across two surveys spaced one week apart. They found that the vast majority (97%) provided consistent demographic data across the two surveys. Additionally, MTurk workers were shown to match or exceed the prevalence of depression, social anxiety disorder, and trauma exposure in the general population (Shapiro et al., 2013). Their findings suggest that MTurk has potential to serve as a useful tool in recruiting clinical and subclinical psychiatric populations.

MTurk has structural and technical advantages that could make it a useful recruitment tool for cancer survivors, that address some of the limitations and recommendations for recruitment posed by previous investigators (e.g. Rabin et al., 2012; Stanton et al., 2013). First, studies and data can be conducted quickly at a low cost (rewards can be as low as \$0.01 and rarely exceed \$1.00; the median reservation wage of \$1.38 per hour) (Horton & Chilton, 2010; Paolacci et al., 2010; Shapiro et al., 2013). MTurk has the added benefit of offering a built-in mechanism to pay workers, which reduces the difficulties of compensating large numbers of individuals for their participation in studies (Mason & Suri, 2012). Second, workers tend to come from diverse backgrounds, have a wide range of age, ethnicity, socio-economic status, and geographic region (Mason & Suri, 2012). Specifically in the U.S., the

mean age of MTurk workers are 33 years old, workers have lower reported annual household income (66.7% of workers have a reported annual household income <60k), and there is a prevalence of female workers (64.85%) (Paolacci et al., 2010). MTurk's relatively young population could be useful particularly in reaching and recruiting young cancer survivors (Paolacci et al., 2010). Third, a worker's reputation defined by approval rate (percentage of the worker's submitted HITs have been approved by the Requester of the work) has a direct effect on the future HITs that a worker can complete (Ross et al., 2010). As a result, workers on MTurk tend to complete HITs with honesty and accuracy as to avoid having their work rejected by a requester (Paolacci et al., 2010; Shapiro et al., 2013).

Fourth, MTurk makes workers anonymous to requestors (only identifiable by a unique worker ID) and this can increase response rate and honesty as well (Shapiro et al., 2013). Workers' unique ID can be used by requesters to identify workers who have already completed a HIT and researchers can exclude these workers accordingly (Paolacci et al., 2010). To further ensure honesty among workers, online labor market creators have their own strong financial incentives to prevent users from having multiple accounts and use a terms-of-use agreement and technical approaches to prevent multiple accounts (Horton, Rand, & Zeckhauser, 2011). Lastly, each worker has a unique worker ID that can restrict what types of HITs a workers can see and complete (Paolacci et al., 2010; Shapiro et al., 2013).

Thus requesters can require workers to have a particular "qualification" (i.e. national origin, age, worker reputation, etc.) which could be a useful tool in targeting specific subgroups of cancer survivors (Shapiro et al., 2013).

This study aims to address the utility of Mechanical Turk's online Internet community as a valid and feasible platform in recruiting cancer survivors to behavioral survey research. No known work to date has assessed the presence of cancer survivors among the MTurk community or the feasibility of recruiting such participants to complete survey research. To address this gap, we aimed to recruit 250 cancer survivors within a relatively brief time span (< 1 month), for low cost (< \$2.00/ participant), and characterize them with regard to cancer type and history, socio-demographics and psychosocial treatment preferences. Finally we aimed to assess the feasibility of conducting longitudinal research with cancer survivors on Mechanical Turk by aiming to recruit our sample for 2 surveys spaced one week apart. A related goal was to assess the validity of the data by including a widely-used test of malingering, Minnesota Multiphasic Personality Inventory-2 Infrequency-Psychopathology Scale ($F(p)$ scale, Arbisi & Ben-Porath, 1995), and including other longitudinal checks (i.e. consistency of cancer type and demographic data reported across both surveys) on the accuracy and reliability of the reported data.

METHODS

Participants

Participants were recruited from Amazon's Mechanical Turk according to the following eligibility criteria: 1) They were U.S. residents, defined by self-report, ownership of a U.S. bank account through Amazon, required social security number (SSN) or individual tax identification number (ITIN) for Amazon's worker account registration as well as Amazon's U.S. Resident Tax Information, and GeoIP estimate with longitudinal and latitudinal coordinates of the computer accessing the survey located in the U.S.; 2) Fluent in

English; 3) At least 18 years of age or older; 4) Had at least a 90% HIT approval rate (meaning 90% of the worker's submitted HITs have been approved by the Requester of the work); and 5) Identified as a cancer survivor, defined here as a person with cancer or a history of cancer (of any type). Non-U.S. residents were excluded because most of our measures have not been validated in non-Western samples and some questions were not relevant to oncology care settings outside of the U.S. Of the original 166 participants that consented to the study, 146 participants completed the first survey, 1 responded from a non-American Internet protocol address, and 2 provided insufficient cancer type information at Survey 1 (i.e., "no cancer"). Of the remaining 143 participants, 23 scored highly on the MMPI-2 measure of malingering discussed below, and 9 provided inconsistent demographic and cancer information at Survey 2 resulting in a final Survey 1 sample of 111 participants that represent the focus of the analyses that follow (See Figure 1). Participants provided their informed consent online. The study was approved by the University of Colorado Boulder Institutional Review Board.

Procedures

In Survey 1, participants completed a well-being survey of approximately 22 minutes of administered questionnaires relating to cancer and general psychosocial functioning. The focus of this paper will be on two of the study-specific cancer-related measures, sociodemographic and support preferences, which includes the measures listed below. Participants who were paid \$0.50 for approximately 22 minutes and recruitment took approximately three weeks. This rate of pay (\$1.42 per hour) is about average for MTurk HITs as the median reservation wage for tasks performed on MTurk is \$1.38 per hour

(Horton & Chilton, 2010; Shapiro et al., 2013). After a worker's response to the first HIT was accepted on MTurk, the worker was emailed indicating they were eligible to complete a second HIT, were assigned a custom qualification (numeric specific criteria assigned to each worker in order to only have those qualified workers access the second survey), and would receive further instruction in one week.

One week after completing the first part of the study, participants were re-contacted through a second email to complete a second survey in exchange for \$0.70. See measures below for the second survey content. Up to five email invitations were sent out to remind participants to complete the second survey if participants still had yet to complete it following the previous email. The five email invitations resulted in eighty-nine (61.0%) participants responding to this request over the course of ~4 weeks, with eighty fully completing the second survey. Of these eighty Survey 2 participants, eight exceeded the established cutoff on a test of malingering at Survey 1 (see malingering measure below) and an additional 9 participants provided inconsistent demographic and cancer history information between surveys 1 and 2. These 17 participants were excluded from further analyses, resulting in a final Survey 2 sample of 63 participants (See Figure 2). Survey 1 participants who participated in Survey 2 did not differ from those on demographic and cancer type. In Survey 2, participants completed the AAQ-II, PHQ-9, BEAQ, and Support Preferences once again. Participants were also asked to provide demographic information and Brief Cancer History again as a mechanism to identify potential data validity issues (i.e. reporting different demographic information and cancer type would indicate lack of validity).

Survey 1 Measures

To decrease time burden for Mechanical Turk workers, we evaluated constructs as efficiently as possible using sample-appropriate, psychometrically sound measures.

The 9-item *Patient Health Questionnaire* (PHQ-9, Kroenke, Spitzer, & Williams, 2001) was designed to briefly assess depressive symptoms in medical settings. The PHQ-9 is roughly half the length of other common depression measures, which decreases the time burden for participants (Kroenke et al., 2001). Major depression is suspected if 5 or more of the 9 depressive symptom criteria have been existent at least “more than half of the days” in the past 2 weeks, and 1 of the symptoms is depressed mood or anhedonia (Kroenke et al., 2001). The PHQ-9 demonstrates good sensitivity (.84) and fair specificity (.72) in medical settings (Kroenke et al., 2001).

The *Minnesota Multiphasic Personality Inventory-2, Infrequency Psychopathology Scale F(p)* (MMPI-2, Arbisi & Ben-Porath, 1995) includes 27 MMPI-2 items answered infrequently (<10% of the time) by both the MMPI-2 normative sample and psychiatric inpatients (Arbisi & Ben-Porath, 1995). The MMPI-II *F(p)* scale demonstrates good construct validity and incremental validity (Arbisi & Ben-Porath, 1995). The higher the *F(p)* score, the more likely it is that the participant has faked bad or malingered in their self-report responses. We used a gender-specific T score corresponding to five standard deviations above the normed mean to designate a malingered response, as recommended by Arbisi & Ben-Porath (1995). A little over fifteen percent ($n= 23$; 15.75%) of our sample score above this cut off and were excluded from analyses.

The *Acceptance and Action Questionnaire-II* (AAQ-II) assesses the construct of acceptance, experiential avoidance, and psychological inflexibility (Bond et al., 2011). The AAQ-II indicates satisfactory structure, reliability, and validity with a mean alpha coefficient

of .84 (.78-.88) and a 3 and 12-month test-retest reliability of .81 and .79, respectively (Bond et al., 2011).

The *Brief Experiential Avoidance Questionnaire* (BEAQ, Gámez et al., 2013) is a 15-item self-report measure that demonstrates expected associations with measures of avoidance, psychopathology, and quality of life. The BEAQ demonstrates good internal consistency and strong convergence with respect to each of the Multidimensional Experiential Avoidance Questionnaires (MEAQ). The MEAQ was developed to assess a broad range of experiential avoidance and the BEAQ includes content from each of the MEAQ's six subscales and tends to be more strongly associated with measures of avoidance across populations (Gámez et al., 2013).

Study-specific *Demographics* and *Brief Cancer History* questionnaires asked about basic socio-demographics, cancer type, cancer stage, cancer related experiences (high anxiety, distress, or depression prior and since diagnosis), psychosocial support previously and/or currently received, and cancer treatment type(s).

(Follow Up Survey) Survey 2 Measures

In the follow up survey - the 2nd HIT MTurk workers were given- we administered the AAQ-II, PHQ-9, BEAQ, & support preferences for a second time. Additionally, we administered the same socio-demographic questionnaire and slightly altered the *Brief Cancer History* to a 10-item questionnaire.

Worker Targeting and Protection Strategies

1) Assigning Worker Qualifications and Requirements:

Survey 1 participants could only complete the first survey if they met the eligibility criteria outlined by the research team. In MTurk, only workers qualified to complete the HIT

can preview the HIT. Workers were required to have a 90% HIT approval rate for all Requesters' HITs, be 18 years or older, and located in the U.S. The HIT approval rate (%) for all requesters is a statistic associated with a worker who does work over time on MTurk and is based on how well the worker has accurately and satisfactorily completed all previous HITs. The Amazon Mechanical Turk Participation Agreement requires that workers and requesters that register and use their site certify that they are at least 18 years or older. Furthermore, Amazon's Mechanical Turk Participation Agreement requires workers to acknowledge and agree to all services used in the site. Any services performed for a Requester will be performed outside of the United States if a worker is not a resident or citizen of the United States. The participation agreement and the worker criteria outlined in the HIT dually ensure that workers that completed our Wave 1 survey were 18 years or older and located in the U.S.

For Survey 2, we created a custom qualification type in MTurk that assigned a particular qualification ID to the participants that successfully completed Survey 1. This custom qualification type, titled "Completion of Survey on Well-Being among Cancer Survivors", only allowed those workers who completed and received payment for the first survey to view and access the second HIT. Additional eligibility criteria noted earlier were not necessary to include in the 2nd HIT requirements because workers already indicated they met those requirements upon completion of the 1st HIT.

2) Tracking Subjects to Ensure Independent Responses

We used two different methods to ensure that participants were not accessing our study HIT multiple times and providing multiple responses to our surveys. The first method was through the use of TurkGate, a tool for researchers that recruits through MTurk but run

their studies on other sites (Darlow & Goldin, 2013). TurkGate provides better control and verification of MTurk workers' access to an external site (Darlow & Goldin, 2013). TurkGate allowed us to group our HITs together so that workers completing the Survey 1 and 2 surveys could only access one survey each. Once a worker had accessed our surveys they are denied future access to the same survey. Additionally TurkGate prevents workers from returning to a survey even if they accidentally closed out of it and from previewing our survey.

The second method used to ensure workers were not accessing the HITs multiple times and providing multiple responses was through, Qualtrics, an external survey software site used for online data collection. We used Qualtrics to program and present the surveys. Qualtrics has survey settings that prevent people from taking the same survey more than once through assignment of unique Response IDs. Qualtrics assigns a response ID unique to that participant and independent from MTurk and TurkGate. These two methods vastly reduced or eliminated the risk of having our subject pool contaminated by multiple responding.

3) Non-Response Error and Verification of Survey Completion

We used additional features in TurkGate and Qualtrics to ensure that participants were answering all questions and fully completing the surveys. Qualtrics has a forced response option that can be used on every survey item ensuring every question is responded to before the participant can move on to the next survey question. A participant cannot continue with the survey if he or she has not filled out every question item. Additionally, to verify that only those who agreed to the informed consent form could proceed with the survey, we implemented an end of survey feature in Qualtrics that jumped to an "end of survey" page and redirection back to MTurk if participants did not give their consent. In

addition, Turkgate uses secure codes to verify that workers are completing their surveys. The secure codes indicate whether or not the information in the first half of a code (i.e. Worker ID, Group Name, and any custom key-value pairs) when hashed with a private encryption key matches the second half of the code (Darlow & Goldin, 2013). This helps prevent users from fabricating their own codes and checks if any codes have been duplicated. In summary, we employed multiple technological features to ensure the highest data quality among study participants.

RESULTS

Hypothesis 1: Measure the Reliability and Validity of Participant Reporting

We assessed participant' honesty and consistency in reporting using multiple approaches. First, we compared their responses to geographic (i.e. "In which state and country do you currently reside?") and cancer type items (i.e. "What type of cancer were you diagnosed with?") across the two surveys. We also used the GeoIP addresses (longitudinal and lateral coordinates of their computer location) of the participants that completed survey 1 and 2 to dually ensure participants were completing the surveys in the location they self-reported. Geographic and cancer type information remained consistent across time periods for 88.75% (71 /80) of the 80 participants who completed both surveys. Nine MTurk participants provided demographic and cancer information at Survey 2 that differed from what they provided at Survey 1 – these participants were excluded from the analyses.

Malingering. We also examined whether participants may have fabricated responses by using the MMPI-2 $F(p)$ scale. Slightly over fifteen percent of our sample ($n = 23$, 15.75%) scored above the 5 SD cutoff that the scale authors suggest indicates probable

malingering (Arbisi & Ben-Porath, 1995). Participants who scored above this cutoff were excluded from further analyses.

These participants excluded from analysis did not differ on their demographic characteristics except the proportion of women malingerers to non-malingerers was smaller than of the proportion of male malingerers to non-malingerers $\chi^2(2, 134) = 6.25, p \leq .05$. This indicates that male participants were more likely to malingering than female participants in our Survey 1 sample. Participants that malingered were also younger on average than those who did not malingering ($M=33.52, SD=12.50$, and $M=41.03, SD= 14.71$), $t(35.857) = -2.540, p \leq .05$.

Test-Retest Reliability. The AAQ-II, PHQ-9, and the BEAQ were given at Survey 1 and Survey 2. AAQ-II scores across the two surveys were highly correlated $r = .83$, as were scores on the PHQ-9 $r = .78$ and the BEAQ $r = .85$, suggesting adequate test-retest reliability in participant responses.

Non-response error. We looked at the number of people that accessed and consented to both surveys but did not fully complete it. 87.96% of participants at Survey 1 fully completed the survey and 89.89% of participants at Survey 2 fully completed the second survey.

Hypothesis 2: Assess the Presence of Cancer Survivors on Mechanical Turk.

Table 1 illustrates the demographic characteristics of the cancer survivor sample that responded to Survey 1 and Table 2 presents cancer types reported from the Survey 1 sample. Participants identified as Cancer survivors were 41.03 years of age ($SD = 14.71$) on average (range = 19-75 years old, median = 38 years) and younger than the general U.S. cancer registry population ($M= 64$) (Stanton et al., 2013). Compared to the general U.S. cancer

registry population, our participants were less educated (2 year college vs. college graduate or above), and have a lower income (31-40k vs. $\geq 60k$) (Stanton et al., 2013). Additionally, participants predominately self-identify as white (81.10%), female (68.50%), and with regard to family and relationships, 36.90% identified themselves as married and 49.50% reporting having one or more children. The Survey 1 sample represented a broad geographic range of the U.S. with 34 states represented among responders, with the highest percentage of participants residing in Florida (9.90%), California (9.00%), Illinois (8.10%), and Georgia (6.30%).

As presented in Table 2, breast (24.03%), cervical (9.90%), melanoma (9.00%), ovarian (7.20%), basal cell/ squamous cell skin carcinoma (7.20%), lymphoma/ Hodgkin lymphoma (5.40%), and lung/ carcinoma of the lung/ small cell lung cancer (5.40%) were the most common cancer types reported among participants.

The majority of participants reported being in stage 1 (localized) or 2 (locally advanced) when diagnosed with cancer (36.00% and 30.60%); see Table 3. Regarding cancer treatment, 42.60% had surgery, 25.40% had chemotherapy, 20.80% had radiation, 7.10% had hormonal treatment, and 4.10% had other forms of treatment.

Table 4 presents the *Brief Cancer History* support preferences reported from Survey 1. Slightly less than half of participants from Survey 1 reported previously struggling with high anxiety, depression, or distress prior to cancer diagnosis (47.70%). However when asked about struggling with high anxiety, depression, or distress since cancer diagnosis, there was nearly a 20.00% increase (67.60%). Despite this sharp rise in reported struggle with high anxiety, depression, or distress since cancer diagnosis, only 34.20% of participants are

currently seeking any kind of treatment or help for anxiety, depression, distress, or other emotional difficulties and 41.40% of participant have previously sought treatment or help.

Hypothesis 3: Evaluate Response Rates from a Follow up Survey Administered 1 One Week Following the Original Survey

Eighty-nine participants of the original 146 participants (61.00%) responded to the second survey, which was administered one week, following the original survey and responses were collected in a roughly four-week period. Response rates to the second survey were consistent with other obtained response rates in follow-up studies for general populations, with one study finding a 60% response rate to a second survey administered within the first few months of collecting data (Chandler, Mueller, & Paolacci, 2013). However, the Survey 2 response rate (61.00%) was lower compared to consistent response rates in similar follow-up clinical studies conducted on MTurk (80.00%) (Shapiro et al., 2013). When controlling for malingering at Survey 1 and inconsistent demographic and cancer information (refer to the procedures section) from Survey 1 to Survey 2, 23 participants scored high on a test for malingering (8 of the 23 participants completed the second survey) and 9 participants provided inconsistent demographic and cancer information across both surveys. This resulted in final sample size of 63 participants for Survey 2.

DISCUSSION

The current study represents the first known exploration of MTurk as a potential recruitment platform for cancer survivors in behavioral survey research. We investigated study-specific cancer-related measures, sociodemographics, and support preferences, at

Survey 1 to descriptively characterize the cancer survivors recruited on MTurk, assess the validity and reliability of the data they provided, and evaluate the possibility of following up with them longitudinally one week after Survey 1 was completed. Feasibility of recruiting cancer survivors was assessed in Survey 1 by tracking indices of common barriers to the recruitment of this population including cost, time, and a lack of sociodemographic diversity. Reliability and validity of reported data was assessed via participant reporting on sociodemographics and cancer type across both surveys as well as by examining whether participants may have fabricated responses by using the MMPI-2 $F(p)$ scale. The possibility of follow-up one week after Survey 1 was completed was evaluated through participant response rate to the second survey.

Reliability and Validity of Participant Reporting

Consistent with earlier research on participant data quality on MTurk (Paolacci et al., 2010; Shapiro et al., 2013), participants in our study demonstrated good test-retest reliability on AAQ-II, PHQ-9, & BEAQ scores across Survey 1 and Survey 2. Both surveys demonstrated reasonably low non-response error with 87.96% of participants fully completing the first survey and 89.89% fully completing the second survey. These numbers may have been higher except that we programmed the survey to lock participants out if they left and came back to it at a later time. Our non-response rate was slightly higher compared to the non-response rates from a previous MTurk study on the general population with 91.60% of their participants ($n= 120$ of 131) fully completing their survey (Paolacci et al., 2010). Additionally, participants' honesty in reporting geographic and cancer type information remained consistent across time periods for 88.75% (71 / 80) of the 80 participants who completed both surveys. This suggests imperfect but reasonable data quality

provided by participants. Specific qualifications used at Survey 1 (i.e. 90% HIT approval rate, located in the U.S., fluent in English, ≥ 18 years) functioned as a prescreen tool to ensure visibility of the survey to those participants that meet the study's criteria. This also offered more methodologically rigorous and sophisticated research design than those typically conducted using online convenience samples (Chandler et al., 2013). A prior study using MTurk to study clinical populations recommended incorporating screening methods for workers' Internet protocol address to ensure better data quality due to a substantial portion of their participants ($n = 33$ of 530) completing their survey from an Internet protocol address located outside of the U.S. (Shapiro et al., 2013). In comparison, we used specific qualifications and a GeoIP estimate (longitudinal and latitudinal coordinates of the computer accessing the survey located in the U.S.) at Survey 1 and found 1 worker's reported residency located outside of the U.S. This ensures participants are from a nation where questionnaires are validated on the normed U.S. population.

Despite the overall honesty and consistency in participant reporting, some concerns about data quality surfaced in this study particularly with the sizable minority of participants (15.75% of the Survey 1 sample) endorsing items consistent with malingering. This suggests that a minority of MTurk participants were motivated to fake psychological distress. A possibility is that participants perceived psychological distress to be an interest to the researcher given the title of the survey, "Survey on Well-Being among Cancer Survivors" and provided high responses of psychological distress as a means to meet the perceived interests of the researcher or to have access to more surveys (increase % HIT approval rate, higher paying surveys, etc.).

Presence of Cancer Survivors on MTurk

In line with our hypothesis, findings revealed a presence of cancer survivors among the MTurk community. Further confirming our hypothesis, there was a prevalence of young adult cancer survivors ($M= 41.03$, $SD=14.71$) in particular. Our study yielded a sample of adult cancer survivors who were significantly younger than large cancer registry populations ($M = 64$ years) (Stanton et al., 2013). Given the widespread use of the Internet by young adults, MTurk could also be a promising nationwide recruitment strategy for mobile young adult cancer survivors given the young average age of our sample. Additionally, our sample had a higher percentage of females (68.50%). The reported sociodemographic characteristics of our sample indicated participants are less educated (2 year college) and have lower annual family income (31-40k) than major cancer registry populations (college graduate or above; $\geq 60k$) (Stanton et al., 2013). However, our sample of cancer survivors MTurk and cancer registry samples similarly reported a higher percentage of non-Hispanic white participants (81.10% vs. 82.64%) (Stanton et al., 2013). These findings suggest that MTurk could be a viable recruitment method for cancer survivors if relative economic and educational diversity are research goals. However, researchers may wish to consider alternative recruitment strategies if greater racial/ ethnic diversity is a major goal.

Cancer survivors appeared more willing to disclose relatively personal information about their experience of cancer, including previous and current type(s) of psychosocial support received and experiences with high anxiety, distress, and depression prior to diagnosis and since diagnosis.

Follow Up Longitudinally One Week Following the Original Survey

To assess the feasibility of follow-up data collection in sophisticated research design on MTurk, we contacted participants one week after completing Survey 1 to take a follow-up

survey. 89 participants (61.00%) responded to the second survey, which is consistent in follow-up studies for general populations, with one study finding a 60% response rate to a second survey administered within the first few months of collecting data (Chandler et al., 2013). However, our study's response rate (61.00%) was lower compared to response rates in similar follow-up clinical studies on MTurk (80.00%) (Shapiro et al., 2013). More adequate approaches are needed in future follow-up studies.

Study Limitations

This study had several limitations. First, the response rate of our second survey (61.00%) was low compared to the obtained follow-up response rates of an MTurk clinical psychology study (80.00%) (Shapiro et al., 2013). A possible explanation for a lower response rate is that pay rate was relatively low for both of our surveys (\$1.41 per hour for Survey 1; \$1.99 per hour for Survey 2) compared to the Shapiro et al. longitudinal MTurk study (\$2.25 per hour for Survey 1; \$2.40 per hour for Survey 2) (Shapiro et al., 2013). Prior studies found that MTurk task response rates increased with higher wages (Mason & Suri, 2012). An additional possible explanation for a lower response rate is the study was not advertised as a longitudinal survey on both of the HITs. Second, our Survey 1 response rate was lower than we hypothesized, with only 166 cancer survivors responding within < 1-month period as opposed to the 250 we were aiming for. Thus, recruiting a very large sample of cancer survivors (for example, 1000+) might be challenging on Mechanical Turk. Third, our study demonstrated the feasibility of using MTurk for survey-oriented research studies with cancer populations but we still have not tested the possibility of using MTurk to recruit for more involved research beyond survey research (e.g. online interventions). Fourth, a sizable minority (15.75%) of our participants scored above a suggested cutoff on a

malingering scale, suggesting that a minority of respondents were not honest and requires future studies to include such approaches to help ensure honest responding. Fifth, we locked out participants who began the survey, took a break, and attempted to finish it later. In that cancer survivors often report difficulties concentrating and high levels of fatigue; this feature may have accidentally excluded legitimate participants from completing the survey. Future studies of cancer survivors on MTurk should omit this feature.

Future Directions

Future studies could use a more targeted survey with a HIT title that included a particular targeted cancer type(s). In using a HIT title that advertises for specific cancer types, we would learn whether recruitment of particular cancer types in larger numbers is feasible. Given the high prevalence of female cancer survivors in our sample and on Mechanical Turk in general, future studies could additionally target Mechanical Turk cancer survivors with breast or cervical cancer. Additionally, an investigation of the feasibility of implementing web-based interventions could be useful in targeting particular cancer survivors groups (e.g., depressed or anxious cancer survivors) that report lower quality of life. Difficulty regaining quality of life is most commonly seen in women and in people diagnosed with cancer at a young age (Siegal et al., 2012). This could be a compelling future direction given the prevalence of younger and female cancer survivors in our sample.

Summary and Conclusion

Despite these limitations, the findings from this study suggest that researchers studying cancer survivors and young adult cancer survivors in particular should consider Mechanical Turk as a potentially useful recruitment strategy. Our findings suggest that use of MTurk requires use of data safeguards such as use of malingering tests because a minority

of participants appeared to be falsely responding. Nonetheless, the majority of participants provided seemingly honest and consistent responses across the two surveys. Relatively high data quality provided by participants as well as the prevalence of survivors of a wide variety of cancers across very broad US geographic regions suggest MTurk could be a viable alternative to recruitment strategies previously considered by researchers (i.e. mailings, in person oncology clinic recruitment, cancer registries, etc.). As the population living with a history of cancer continues to grow, the formation and implementation of evidence-based methods for understanding and promoting the health and well-being of cancer survivors are critical. Using MTurk as a recruitment platform could address several of the recruitment challenges psycho-oncology researchers currently face in pursuing this work.

REFERENCES

- Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology Scale. *Psychological Assessment*, 7(4), 424.
- Avis, N. E., & Deimling, G. T. (2008). Cancer survivorship and aging. *Cancer*, 113(S12), 3519-3529.
- Ben-Porath, Y. S., & Butcher, J. N. (1989). The comparability of MMPI and MMPI-2 scales and profiles. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1(4), 345.
- Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, 42(4), 676-688.
- Chandler, J., Mueller, P., & Paolacci, G. (2013). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior research methods*, 1-19.
- Darlow, A., & Goldin, G (2013). Turkgate. *Turkgate*. Retrieved March 27, 2014, from <http://gideongoldin.github.io/TurkGate/>
- Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., Suzuki, N., & Watson, D. (2013, September 23). The Brief Experiential Avoidance Questionnaire: Development and Initial Validation. *Psychological Assessment*. Advance online publication. doi: 10.1037/a0034473
- Ganz, P. A., Land, S. R., Antonio, C., Zheng, P., Yothers, G., Petersen, L., ... & Ko, C. Y. (2009). Cancer survivorship research: the challenge of recruiting adult long term cancer survivors from a cooperative clinical trials group. *Journal of Cancer Survivorship*, 3(3), 137-147.
- Hewitt, M., Rowland, J. H., & Yancik, R. (2003). Cancer survivors in the United States: age, health, and disability. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 58(1), M82-M91.
- Horton, J. J., & Chilton, L. B. (2010, June). The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce* (pp. 209-218). ACM.
- Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399-425.

- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. *Journal of general internal medicine, 16*(9), 606-613.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods, 44*(1), 1-23.
- Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making, 5*(5), 411-419.
- Pollack, L. A., Greer, G. E., Rowland, J. H., Miller, A., Doneski, D., Coughlin, S. S., ... & Ulman, D. (2005). Cancer survivorship: a new challenge in comprehensive cancer control. *Cancer Causes & Control, 16*(1), 51-59.
- Rabin, C., Horowitz, S., & Marcus, B. (2012). Recruiting Young Adult Cancer Survivors for Behavioral Research. *Journal of clinical psychology in medical settings, 1*-4.
- Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010, April). Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 2863-2872). ACM.
- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science, 1*(2), 213-220.
- Siegel, R., DeSantis, C., Virgo, K., Stein, K., Mariotto, A., Smith, T., ... & Ward, E. (2012). Cancer treatment and survivorship statistics, 2012. *CA: a cancer journal for clinicians, 62*(4), 220-241.
- Stanton, A. L. (2012). What happens now? Psychosocial care for cancer survivors after medical treatment completion. *Journal of Clinical Oncology, 30*(11), 1215-1220.
- Stanton, A. L., Morra, M. E., Diefenbach, M. A., Miller, S. M., Perocchia, R. S., Raich, P. C., ... & Marcus, A. C. (2013). Responding to a significant recruitment challenge within three nationwide psychoeducational trials for cancer patients. *Journal of Cancer Survivorship, 1*-12.

Figure 1. Participant Flow Survey 1.

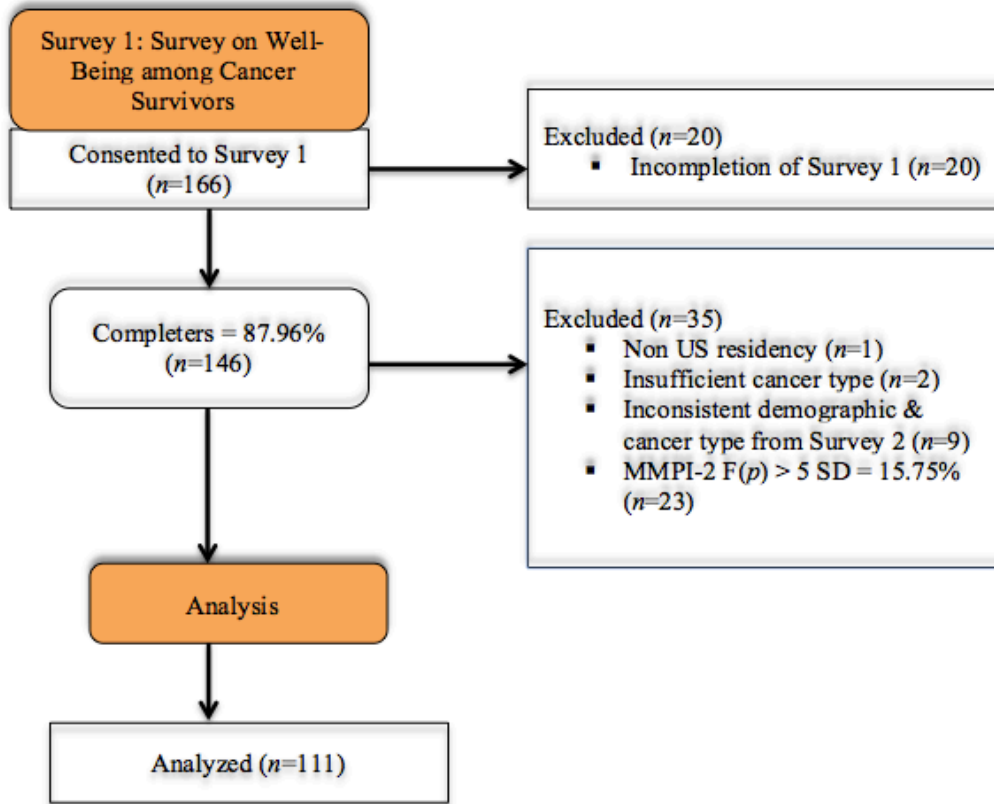
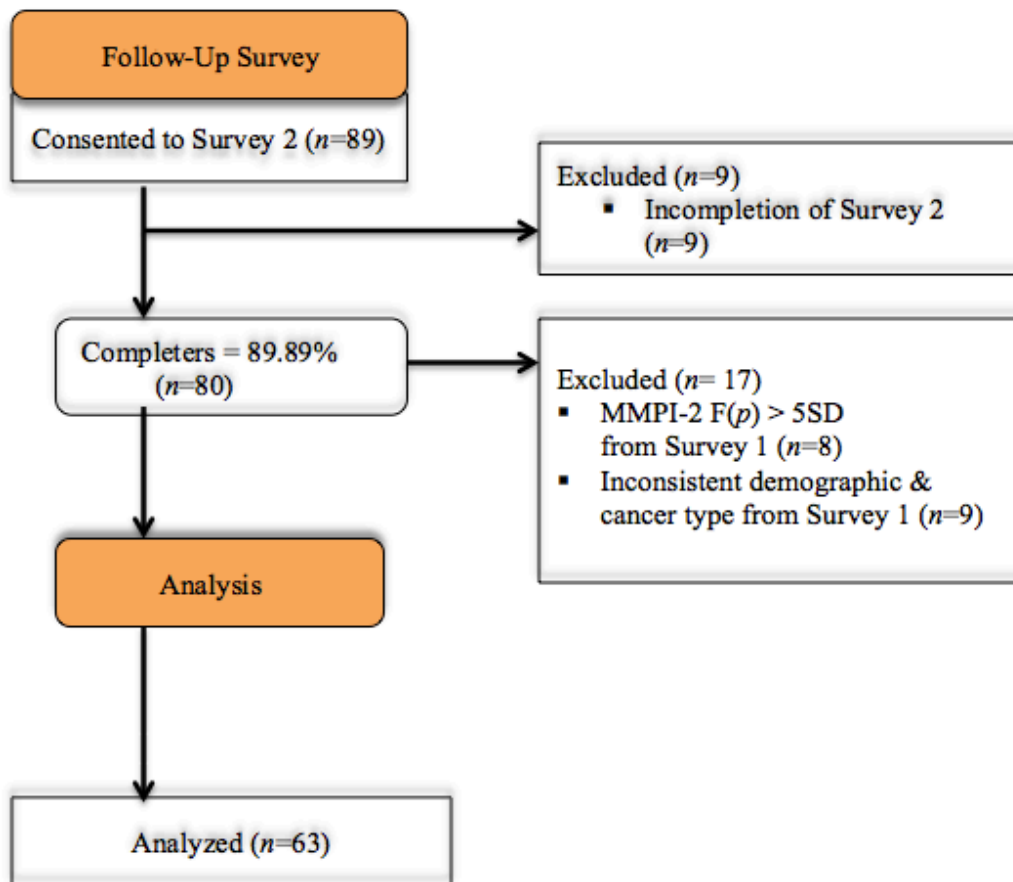


Figure 2. Participant Flow Survey 2.



**Table 1 Final sample
demographic characteristics**

	Survey 1 (<i>n</i> = 111)
Age, in years, <i>Mean (SD)</i>	41.03 (14.71)
Education, <i>Mean (SD)</i>	4.16 (1.32)
	4= 2 yr. college degree
Household Income, <i>Mean (SD)</i>	4.01 (2.16)
	4= 31-40k
Sex (Female), %(<i>n</i>)	68.50% (76/111)
Reported race/ethnicity, %(<i>n</i>)	
White/ Caucasian	81.10% (90/111)
Black/ African American	7.20% (8/111)
Hispanic/ Latino/a	6.30% (7/111)
Asian Am/ Pacific Islander	2.70% (3/111)
Biracial	1.80% (2/111)
Native Am/ Alaskan Native	.90% (1/111)
Marital status, %(<i>n</i>)	
Married	36.90% (41/111)
Partnered but not married	26.10% (29/111)
Single	19.80% (22/111)
Divorced/ Separated	8.10% (9/111)
Other	9.00% (10/111)
Children (1+), %(<i>n</i>)	49.50% (55/111)

Table 2 Reported Cancer Type.

Cancer Type, %(n)	Survey 1 (<i>n</i> =111)
Breast	24.03% (27/111)
Cervical	9.90% (11/111)
Melanoma	9.00% (10/111)
Ovarian	7.20% (8/111)
Skin/Basal Cell/ Squamous Cell Carcinoma	7.20% (8/111)
Lymphoma/ Hodgkin Lymphoma	5.40% (6/111)
Lung/ Carcinoma of the Lung/ Small Cell Lung Cancer	5.40% (6/111)
Thyroid/ Papillary Thyroid	4.50% (5/111)
Uterine/Endometrial	4.50% (5/111)
Prostate	3.60% (4/111)
Colon Cancer	2.70% (3/111)
Bone	2.70% (3/111)
Chronic Myelogenous Leukemia/ Acute Lymphoblastic Leukemia	1.80% (2/111)
Cholangiocarcinoma	1.80% (2/111)
Testicular	1.80% (2/111)
Lingual/ Oral/ Tongue	0.90% (1/111)
Kidney	0.90% (1/111)
Cerebral Pilocytic Astrocytoma	0.90% (1/111)
Appendix	0.90% (1/111)
Pancreatic	0.90% (1/111)
Neurofibrosarcoma	0.90% (1/111)
Brain	0.90% (1/111)
Throat (Laryngeal and Pharyngeal)	0.90% (1/111)
Intestinal/ Celiac Diseases/ Crohn's Disease	0.90% (1/111)

Table 3. Stage and Treatment Characteristics.

Characteristics.	Survey 1 (n=111)
What stage of cancer were you diagnosed with?	
Stage 0	9.00% (10/111)
Stage 1	36.00% (40/111)
Stage 2	30.60% (34/111)
Stage 3	8.10% (9/111)
Stage 4	6.30% (7/111)
Other	9.90% (11/111)
What type(s) of treatment have you had?	
Chemotherapy	25.40%
Radiation	20.80%
Surgery	42.60%
Hormonal Treatment	7.10%
Other	4.10%

Table 4. Brief Cancer History	Survey 1 (n=111)
PRIOR to your cancer diagnosis, did you ever struggle with high anxiety, depression, or distress? (Yes)	47.70% (53/111)
SINCE your cancer diagnosis, have you struggled with high anxiety, depression, or distress? (Yes)	67.60% (75/111)
Are you CURRENTLY seeking any kind of help or treatment for anxiety, depression, distress, or other emotional difficulties? (Yes)	34.20% (38/111)
If YES, what type(s) of treatment are you currently doing?	
Counseling and Psychotherapy	24.10%
Medication	36.20%
Both counseling/ psychotherapy and medication	12.10%
Alternative Medicine (yoga, acupuncture, medication, etc.)	19.00%
Support, therapy, or skills group	6.90%
Other	1.70%
To what degree do you think these difficulties were related to your having had cancer?	
Not at all Related	15.80%
Somewhat Related	42.10%
Moderately Related	18.40%
Mostly Related	18.40%
Completely Related	5.30%
Have you ever PREVIOUSLY sought treatment or help for anxiety, depression, distress, or other emotional difficulties? (Yes)	41.40% (46/111)
If YES, what type(s) of treatment have you done in the past?	
Counseling and Psychotherapy	26.50%
Medication	24.10%
Both Counseling/ psychotherapy, and medication	27.70%
Alternative Medicine (yoga, acupuncture, medication, etc.)	13.30%
Support, therapy, or skills group	8.40%
Other	0.00%
To what degree do you think these difficulties were related to your having had cancer?	
Not at all Related	60.90%
Somewhat Related	15.20%
Moderately Related	13.00%
Mostly Related	4.30%
Completely Related	6.50%

