Comparing Amazon's Mechanical Turk Platform to Conventional Data Collection Methods in the Health and Medical Research Literature

Karoline Mortensen, Ph.D.[1] *and Taylor L. Hughes, B.S.*[2]

[1]Department of Health Sector Management and Policy School of Business Administration , University of Miami, Coral Gables, FL, USA; [2]Duke University, Durham, NC, USA.

BACKGROUND: The goal of this article is to conduct an assessment of the peer-reviewed primary literature with study objectives to analyze Amazon.com's Mechanical Turk (MTurk) as a research tool in a health services research and medical context.
METHODS: Searches of Google Scholar and PubMed databases were conducted in February 2017. We screened article titles and abstracts to identify relevant articles that compare data from MTurk samples in a health and medical context to another sample, expert opinion, or other gold standard. Full-text manuscript reviews were conducted for the 35 articles that met the study criteria.
RESULTS: The vast majority of the studies supported the use of MTurk for a variety of academic purposes.
DISCUSSION: The literature overwhelmingly concludes that MTurk is an efficient, reliable, cost-effective tool for generating sample responses that are largely comparable to those collected via more conventional means. Caveats include survey responses may not be generalizable to the US population.

KEY WORDS: Amazon Mechanical Turk; MTurk; Alternate data sources; Health and medical research.

J Gen Intern Med 33(4):533–8
DOI: 10.1007/s11606-017-4246-0

Amazon.com's Mechanical Turk (MTurk) is an online, web-based platform that started in 2005 as a service to allow researchers to "crowdsource" labor-intensive tasks for workers registered on the site to complete for compensation.[1, 2] MTurk has rapidly become a source of subjects for experimental research and survey data for academic work, as its representativeness, speed, and low cost appeal to researchers.[2, 3] Researchers post links to surveys and experiments and use MTurk to crowdsource the survey, collect the data, and compensate workers.[4] A Google Scholar search of "Amazon Mechanical Turk" revealed 15,000 results published between 2006 and 2014[3] and 17,400 results by mid-2017. MTurk is the largest online crowdsourcing platform,[4] with about one-third of the tasks related to academic tasks.[5] The growing popularity of MTurk has led to questions about its soundness as a subject pool; MTurk is the most studied nonprobability sample available to researchers.[3]

The MTurk pool of potential workers is vast, diverse, and inexpensive. MTurk has 500,000 registered users[3] with 15,000 individual US workers at any given time.[6] MTurkers have been paid as little as $0.05 to complete 10- to 15-min tasks.[4] Researchers can collect data from large enough samples to generate significant statistical power at one-tenth of the cost of traditional methods.[4] The MTurk population is more representative of the population at large than other online surveys and produces reliable results.[2, 3, 6–11]

There is a rapidly growing literature exploring the generalizability of MTurk responses to other data collection methods. Data obtained via MTurk surveys and experiments are at least as reliable as those obtained via traditional methods, are attractive for conducting internally and externally valid experiments, and the advantages outweigh the disadvantages.[3, 8–15] However, the benefits and drawbacks to using MTurk in the health and medical literature are largely unexplored beyond a taxonomy of how MTurk has been used in health and medical research.[16] This article is the first synthesis to assess the peer-reviewed literature that has a study objective to analyze MTurk as research tool in a health services research and medical context and uses MTurk for part or all the results. The results from this synthesis can guide academic researchers as they explore the strengths and weaknesses of employing MTurk as an academic research platform.

METHODS

A literature search was performed for articles published between 2005 and mid-February 2017 using Google Scholar and PubMed databases. Searches for variations of the terms Mechanical Turk, MTurk, health, healthcare, clinic*, and medic yielded an initial total of 331 non-duplicative articles. Two reviewers (TH and KM) screened the articles first by title review, eliminating those that did not pertain to health as defined by the World Health Organization,[17] leaving 181

Received August 3, 2017
Revised September 29, 2017
Accepted November 21, 2017
Published online January 4, 2018

Table 1 Summary of Research Findings

Research study	Objective	Description
Aghdasi, Bly, White, Hannaford, Moe, and Lendvay (2015)	To assess the ability of a large group of laypersons using a crowdsourcing tool (MTurk) to grade a surgical procedure (cricothyrotomy) performed on a simulator	When compared to expert head and neck surgeons, the crowdsourced group took significantly less time to complete the analyses (10 h vs. 60 days), and the assessment of complex surgical performance by laypersons matched those of the experts (correlation coefficient 0.833). CSATS using MTurk workers provides an efficient, accurate, and inexpensive method of evaluating surgical performance, even when applied to complex procedures
Arch and Carr (2016)	To assess the presence of cancer survivors on MTurk and the feasibility of using it as an efficient, cost-effective, and reliable platform for psycho-oncology research	MTurk was shown to be a successful tool for recruiting young adult cancer survivors in particular, and the participants recruited were more geographically, medically, and socio-demographically diverse groups of cancer survivors than with many other available psycho-oncology recruitment sources. It was cost-effective, time-efficient, and the data obtained were reliable
Arditte, Çek, Shaw, and Timpano (2015)	To examine the similarities and differences of clinical phenomena reported by MTurk workers and the general US population	Results suggested that MTurk workers were significantly more likely to report clinical symptoms associated with social anxiety and depression than traditional community or epidemiological samples. Psychopathology researchers are encouraged to exercise caution when generalizing MTurk findings to the larger US population
Bardos, Friedenthal, Spiegelman, and Williams (2016)	To evaluate, using MTurk, a cloud-based workforce to assess patients' perspectives of health care	The authors were able to effectively and efficiently collect survey data from a large national pool, including quantitative and qualitative data about patients' knowledge and experience with miscarriage. These quality data were collected from over 1000 respondents over a 3-day window for under $300 USD (4 to 69 times less than other survey methods)
Boynton and Richman (2014)	To illustrate the utility of MTurk to recruit a diverse sample of adults for participation in an online daily diary study of alcohol use	Multilevel models of daily alcohol data derived from the MTurk sample replicated findings commonly reported in daily diary studies of alcohol use
Brady, Villanti, Pearson, Kirchner, Gupta, and Shah (2014)	To develop and validate the use of MTurk as a method of fundus photograph grading	MTurk workers were able to quickly and correctly categorize retinal images of diabetic patients into normal and abnormal, though there remains a need to improve MTurkers' ability to correctly rate the degree and severity of retinopathy. MTurk offers a novel and inexpensive means to reduce skilled grader burdens and increase screening for diabetic retinopathy
Briones and Benham (2016)	To compare data from MTurk to data on psychological stress and sleep quality obtained from a college student sample	The data obtained from MTurk were statistically equivalent to data from the conventional undergraduate college sample
Brown and Allison (2014)	To compare MTurk workers' evaluations of abstracts on nutrition and obesity to authors' expert expectations	MTurkers reached consensus in 96% of reviewed abstracts, and over 99% of their ratings of obesogenicity of foods were complete and usable. Crowdsourcing on MTurk can be an economical and timesaving approach to evaluate large bodies of published literature
Chen, White, Kowalewski, Aggarwal, Lintott, Comstock, Kuksenok, Aragon, Holst, and Lendvay (2013)	To test a web-based grading tool using crowdsourcing on MTurk and Facebook	Surgery-naive MTurk workers were able to assess robotic suturing performance equivalent to experienced faculty surgeons in a shorter timeframe
Deal, Lendvay, Haque, Brand, Comstock, Warren and Alseidi (2016)	To assess the feasibility and reliability of CSATS for evaluation of technical skills in general surgery	Qualitative feedback from both crowd workers and faculty experts was "remarkably similar." Use of MTurk is a reliable basic tool for the standardization of surgical technique evaluation
Gardner, Brown, and Boice (2012)	To investigate MTurk as a research tool for measuring body size estimation and dissatisfaction	MTurker's assessment of their body size based on anthropometric data were similar to the results of three previous studies using traditional samples. MTurk can serve as a viable method for collecting data on perception and attitudes of body image quickly and inexpensively
Good, Nanis, Wu, and Su (2015)	To compare MTurk workers' disease mention annotations to classifications by the "expert-crafted gold standard"	Overall, the MTurk workers generally performed at a high level for the task relative to the gold standard annotation. The MTurk worker population was able to effectively process biomedical text
Harber and Leroy (2015)	To illustrate the utility of crowdsourcing on MTurk for occupational health surveillance	MTurk was used to recruit and obtain information from employed individuals with asthma, who answered questions about the interaction of their asthma and work. Crowdsourcing methods are extremely effective and have potential for occupational health surveillance tools because of their efficiency, effectiveness, and financial viability

(*continued on next page*)

Table 1. (continued)

Research study	Objective	Description
Harris, Mart, Moreland-Russell, and Caburnay (2015)	To evaluate the ability of crowdsourcing on MTurk to classify Twitter postings containing diabetes information into 9 topic categories, relative to classifications by experts	Classification by MTurkers relative to experts was reliable at the good or excellent level. MTurk may be a reliable, quick, and economical way for researchers to code large amounts of data gathered from social media
Hipp, Manteiga, Burgess, Stylianou, and Pless (2016)	To validate the use of MTurkers to interpret images from webcams to explore the effects of built environment changes on active transportation	Classification by MTurkers relative to classifications developed by trained research assistants resulted in an objective, cost-effective alternative to traditional methods
Holst, Kowalewski, White, Brand, Harper, Sorensen, Truong, Simpson, Tanaka, Smith, and Lendvay (2015)	To evaluate the ability of crowd workers to provide valid performance scores of surgical skills in live tissue when compared to the gold standard of assessment	Crowdsourcing basic surgical skills compares favorably to assessments by expert surgeons, and it provides larger volume feedback in a shorter period of time than expert assessors. However, further research is needed to link CSATS to clinical outcomes to confidently presume that non-medically trained workers can accurately assess surgical skills
Khare, Burger, Aberdeen, Tresner-Kirsch, Corrales, Hirchman, and Lu (2015)	To compare drug indication cataloging of FDA drug labels by MTurk workers to the annotation by domain experts	The MTurk workforce's judgments on cataloging drug indications from FDA drug labels achieved an aggregated accuracy of 96%. Employing MTurkers results in significant cost and time saving while reaching accuracy comparable to that of domain experts
Kim and Hodgins (2016)	To evaluate the validity and internal and retest reliability of data obtained from addiction populations on MTurk	Self-reported data for alcohol and gambling populations are of high quality, though caution is warranted because of significant differences in the cannabis sample
Kuang, Agro, Stoddard, Bray, and Zeng-Treitler (2015)	To explore the application of online crowdsourcing for health informatics research, specifically the testing of medical pictographs	The MTurk group scored significantly higher on depicting pictographs with discharge instructions than the traditional in-person sample. Crowdsourcing is a viable complement to traditional in-person surveys, but it cannot replace them
Lee, Lee, Keane, and Tufail (2016)	To evaluate the feasibility of using MTurk as a platform for performing manual segmentation of macular optical coherence tomography (OCT) images	Using MTurk workers to manually perform OCT segmentation generated thousands of data points in a timely and cost-effective manner, and it showed a high degree of interrater reliability. Though this study shows promise of the novelty of using crowdsourced workers for OCT segmentation, it remains unknown whether the accuracy of the data is comparable to evaluations by trained ophthalmologic experts
Lloyd, Yen, Pietrobon, Wiener, Ross, Kokorowski, Nelson, and Routh (2014)	To assess the feasibility of using MTurk online platform to elicit utility values for performing cost-utility analyses of multiple treatment choices	Mturk participants were not representative of the US population, but were more representative than other traditional convenience samples, such as undergraduate campuses. Response rates using MTurk were within acceptable range for survey research and utility values were similar to previous studies. However, these data were not validated against a standard sample using face-to-face interview techniques, and the authors cannot recommend widespread adoption until validated
Maclean and Heer (2013)	To compare MTurkers' ability to identify medically relevant terms in patient-authored text to annotations by registered nurses	The inter-rater reliability scores for the MTurk and a group of 30 registered nurses on medical word identification tasks were nearly identical, MTurkers performance is comparable in quality to those given by medical experts, and they are an acceptable approximation for expert judgment
Mitry, Peto, Hayat, Blows, Morgan, Khaw, and Foster (2015)	To evaluate the performance and repeatability of crowdsourcing the classification of normal and glaucomatous discs from optic disc images	MTurk workers were able to categorize optic disc images with high sensitivity (83–88%), but poor specificity (35–43%). Crowdsourcing represents a cost-effective image analysis method with good repeatability
Mitry, Zutis, Peto, Hayat, Khaw, Morgan, Moncur, Trucco, and Foster (2016)	To develop a novel online tool to facilitate large-scale annotation of digital retinal images and assess the accuracy of MTurk worker grading using the tool compared to expert classification	Annotation of abnormalities retinal fundus photograph images by ophthalmologically naive MTurk workers is comparable to expert annotation, and the highest agreement with expert annotation was achieved in workers that underwent compulsory training. MTurk has the potential to deliver timely, accurate, and cost-effective image retinal analysis
Mortensen, Musen, and Noy (2013)	To compare MTurkers' ability to verify correct and incorrect biomedical relationships to the correct answers determined by experts	A method developed to verify ontology relations applied to crowdsourced MTurk workers verified 86% of the relations. High-performance, cost effective strategies can be deployed via an MTurk workforce

(continued on next page)

Table 1. (continued)

Research study	Objective	Description
Powers, Boonjindasup, Pinsky, Dorsey, Maddox, Su, Gettman, Sundaram, Castle, Lee, and Lee (2016)	To compare the assessment of surgeons' technical performance of renal artery and vein dissection during robotic partial nephrectomy done by crowdsourced workers and expert surgeon graders	Crowdsourced ratings on MTurk were highly correlated with surgical content experts' assessments. Crowdsourcing provides a rapid, cost-effective, scalable alternative or adjunct to surgical expert ratings
Santiago-Rivas, Schnur, and Jandorf (2016)	To explore the use of MTurk crowdsourcing for cluster analysis of the assessment of sun protection beliefs relative to the clustering by experts	The authors conclude that the results of their study provide a potential alternative approach to developing future sun protection initiatives in the population
Schleider and Weisz (2015)	To test the feasibility of MTurk as a platform to obtain reports from parents on their family function and youth mental health	Parents on MTurk provided high-quality data, and the authors conclude that MTurk was successful in achieving enrollment goals and comparable to other studies using different samples in attrition, race/ethnicity, and enrollment of single parents
Shao, Guan, Clark, Liu, Santelices, Cortes, and Merchant (2015)	To explore the yields, speed, and costs of recruitment and participant diversity in a worldwide, internet-based study of HIV/AIDS and HIV testing knowledge	MTurk yielded the most participants, recruited participants at the fastest rate, had the highest completion-to-enrollment ratios, and lowest cost per completion for English-speaking platforms relative to other platforms like Google and Facebook. However, international MTurk respondents tend to be well-educated participants from South and Southeast Asia, so the results may not be demographically reflective of the global population
Shapiro, Chandler, and Mueller (2013)	To assess the utility of using Amazon's MTurk for conducting research on psychopathology	The prevalence of depression, general anxiety, and trauma exposure among MTurk workers matched or exceeded prevalence in the general population, allowing researchers to access participants with the full range of symptoms as they would in the general population
Turner, Kirchhoff, and Capurro (2012)	To examine crowdsourcing as a method to gather feedback on the design of health promotion messages for oral health	Crowdsourcing has the potential to reach more diverse populations than convenience sampling, while substantially reducing the time and cost of gathering participant feedback
White, Kowalewski, Dockter, Comstock, Hannaford, Lendvay (2015)	To test whether crowdsourced layworkers could discriminate robotic surgical skill levels of two dry-laboratory surgical tasks in agreement with expert faculty surgeons	Evidence shows that crowdsourced workers' assessments are largely in agreement with expert evaluators and are less costly. Limitations of generalizability beyond dry-laboratory setting (potential issues with generalizability to actual human surgery performance) and robotic surgery (i.e., to laparoscopic surgery, open surgeries, etc.)
Wu, Hultman, Diegidio, Hermiz, Garimella, Crutchfield, and Lee (2017)	To test a conjoint analysis of attributes favored by patients seeking an esthetic surgeon, comparing data gathered from MTurk and a pilot study of a university-wide community	Results from a conjoint analysis of desired attributes of an esthetic surgeon were similar among an anonymous university survey and an MTurk survey. The authors conclude that MTurk benefits include broad, diverse, anonymous participant pools, low-cost, rapid data collection, and high completion rate
Wymbs and Dawson (2015)	To determine the utility of MTurk for studying adults with ADHD by screening workers for ADHD diagnostic histories, symptoms, and diagnostic comorbidities	Upon comparison with DSM-V's and CDC's current estimates for the national prevalence of ADHD in childhood and adulthood, diagnostic prevalence, demographic correlates, symptom profiles, and internalizing comorbidities are consistent with studies of "offline" populations. MTurk offers an efficient and inexpensive way to gather large quantities of clinically relevant data from adults with ADHD
Yu, Willis, Sun, and Wang (2013)	To compare Mturkers' interpretations of medical pictograms for pharmaceuticals to those of two judges, and of a panel in another study that used an open-ended, in-person panel	Crowdsourcing via MTurk can be used as an effective and inexpensive approach for participatory evaluation of medical pictograms. The authors noted that misinterpretations of pictograms made by MTurkers were a result of concepts that were difficult to depict graphically and thus the workers exposed design problems, rather than lack of skill

articles. After abstract and full-text review, 35 articles were included in the final analysis.

RESULTS

The 35 articles that met the criteria of primary peer-reviewed article, MTurk used in part or all of the results, and an objective of the study was to analyze MTurk as a research tool in a health services research and medical context are described briefly in Table 1.[12, 18–51] A number of strengths of using MTurk in an academic health services setting were identified in the literature. The studies were overwhelmingly supportive of the economical, cost-effective nature of MTurk.[18, 19, 23, 25, 26, 29–34, 37, 38, 40, 41, 45–49, 51] Additional strengths include the time-saving component of using MTurk, reliability, and high quality. Accurate,[34] effective,[29, 30, 51] performance comparable to quality of medical experts,[18, 26, 33, 34, 39, 41, 43, 48] high

verification,[42] reliable,[27, 31] objective,[32] statistically equivalent to data from other samples,[12, 22, 24, 38, 49, 50] diverse,[19, 21, 47, 49] and viable,[28, 36] high quality,[35, 42, 45] among other strengths, were consistent conclusions in the literature.

The weaknesses are dominated by the identified strengths, but important to note. Four studies[20, 36–38] noted three caveats: (1) researchers should exercise caution when generalizing MTurk findings to the US population;[20, 36] (2) despite a high degree of inter-rater reliability in the MTurk sample, it is unknown whether the accuracy of the data is comparable to evaluations by trained ophthalmologic experts;[37] (3) the data were not validated against a sample using face-to-face interview techniques.[38] The literature overwhelming concludes that MTurk is an efficient, reliable, cost-effective tool for a variety of tasks with results comparable to those collected via more conventional means. However, results from surveys on MTurk should not be generalized to the US population.

Corresponding Author: *Karoline Mortensen, Ph.D.; Department of Health Sector Management and Policy School of Business Administration University of Miami, Coral Gables, FL, USA (e-mail: Kmortensen@bus.miami.edu).*

Compliance with Ethical Standards:

Conflict of Interest: *The authors declare no conflicts of interest.*

REFERENCES

1. **Redmiles EM, Kross S, Pradhan A, Mazurek ML**. How well do my results generalize? Comparing security and privacy survey results from MTurk and web panels to the US; 2017. Technical Report of the Computer Science Department at the University of Maryland. http://drum.lib.umd.edu/handle/1903/19164.
2. **Paolacci G, Chandler J, Ipeirotis P.** Running experiments on Amazon Mechanical Turk. Judgment and decision making. 2010;5(5):411–419. https://doi.org/10.2139/ssrn.1626226.
3. **Chandler J, Shapiro DN.** Conducting clinical research using crowdsourced convenience samples. Annual review of clinical psychology. 2016;12:53–81. https://doi.org/10.1146/annurev-clinpsy-021815-093623.
4. **Pittman M, Sheehan K.** Amazon's Mechanical Turk a digital sweatshop? Transparency and accountability in crowdsourced online research. Journal of media ethics. 2016;31(4):260–262. https://doi.org/10.1080/23736992.2016.1228811.
5. **Hitlin P.** Research in the crowdsourcing Age, a case study.; 2016. http://www.pewinternet.org/2016/07/11/research-in-the-crowdsourcing-age-a-case-study/.
6. **Stewart N, Harris AJL, Bartels DM, Newell BR, Paolacci G, Chandler J.** The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. Judgment and decision making. 2015;10(5):479–491. https://doi.org/10.1017/CBO9781107415324.004.
7. **Behrend TS, Sharek DJ, Meade AW, Wiebe EN.** The viability of crowdsourcing for survey research. Behavioral research methods. 2011;43(3):800–813. https://doi.org/10.3758/s13428-011-0081-0.
8. **Berinsky AJ, Huber GA, Lenz GS.** Evaluating online labor markets for experimental tesearch: Amazon.com's Mechanical Turk. Political analysis. 2012;20(3):351–368. https://doi.org/10.1093/pan/mpr057.
9. **Buhrmester M, Kwang T, Gosling SD.** Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspectives in psychological science. 2011;6(1):3–5. https://doi.org/10.1177/1745691610393980.
10. **Woods AT, Velasco C, Levitan CA, Wan X, Spence C.** Conducting perception research over the internet: a tutorial review. PeerJ. 2015;3:e1058. https://doi.org/10.7717/peerj.1058.
11. **Sheehan KB**. Crowdsourcing research: Data collection with Amazon's Mechanical Turk. Commun Monogr. 2017;0(0):1–17. https://doi.org/10.1080/03637751.2017.1342043.
12. **Shapiro DN, Chandler J, Mueller PA.** Using Mechanical Turk to study clinical populations. Clinical psychological science. 2013;1(2):213–220. https://doi.org/10.1177/2167702612469015.
13. **Casler K, Bickel L, Hackett E.** Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. Computers in human behavior. 2013;29(6):2156–2160. https://doi.org/10.1016/j.chb.2013.05.009.
14. **Horton JJ, Rand DG, Zeckhauser RJ.** The online laboratory: Conducting experiments in a real labor market. Experimental economics. 2011;14(3):399–425. https://doi.org/10.1007/s10683-011-9273-9.
15. **Mason W, Suri S.** Conducting behavioral research on Amazon's Mechanical Turk. Behavioral research methods. 2012;44(1):1–23. https://doi.org/10.3758/s13428-011-0124-6.
16. **Ranard BL, Ha YP, Meisel ZF, et al.** Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. Journal of general internal medicine. 2014;29(1):187–203. https://doi.org/10.1007/s11606-013-2536-8.
17. **Constitution of the World Health Organization**. 1946. http://www.who.int/about/mission/en/.
18. **Aghdasi N, Bly R, White LW, Hannaford B, Moe K, Lendvay TS.** Crowdsourced assessment of surgical skills in cricothyrotomy procedure. Journal of surgical research. 2015;196(2):302–306. https://doi.org/10.1016/j.jss.2015.03.018.
19. **Arch JJ, Carr AL**. Using Mechanical Turk for research on cancer survivors. Psychooncology. 2016; https://doi.org/10.1002/pon.4173.
20. **Arditte KA, Cek D, Shaw AM, Timpano KR.** The importance of assessing clinical phenomena in Mechanical Turk research. Psychological assessment. 2016;28(6):684–691. https://doi.org/10.1037/pas0000217.
21. **Bardos J, Friedenthal J, Spiegelman J, Williams Z.** Cloud based surveys to assess patient perceptions of health care: 1000 respondents in 3 days for US $300. JMIR research protocols. 2016;5(3):e166. https://doi.org/10.2196/resprot.5772.
22. **Boynton MH, Richman LS.** An online daily diary study of alcohol use using Amazon's Mechanical Turk. Drug and alcohol review. 2014;33(4):456–461. https://doi.org/10.1111/dar.12163.
23. **Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP.** Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. Journal of medical internet research. 2014;16(10):e233. https://doi.org/10.2196/jmir.3807.
24. **Briones EM, Benham G.** An examination of the equivalency of self-report measures obtained from crowdsourced versus undergraduate student samples. Behavioral research methods. 2016. https://doi.org/10.3758/s13428-016-0710-8.
25. **Brown AW, Allison DB.** Using crowdsourcing to evaluate published scientific literature: Methods and example. PLoS One. 2014;9(7):e100647. https://doi.org/10.1371/journal.pone.0100647.
26. **Chen C, White L, Kowalewski T, et al.** Crowd-Sourced Assessment of Technical Skills: A novel method to evaluate surgical performance. Journal of surgical research. 2014;187(1):65–71. https://doi.org/10.1016/j.jss.2013.09.024.
27. **Deal SB, Lendvay TS, Haque MI, et al.** Crowd-sourced assessment of technical skills: an opportunity for improvement in the assessment of laparoscopic surgical skills. American journal of surgery. 2016;211(2):398–404. https://doi.org/10.1016/j.amjsurg.2015.09.005.
28. **Gardner RM, Brown DL, Boice R.** Using Amazon's Mechanical Turk website to measure accuracy of body size estimation and body dissatisfaction. Body image. 2012;9(4):532–534. https://doi.org/10.1016/j.bodyim.2012.06.006.
29. **Good BM, Nanis M, Wu C, Su AI.** Microtask crowdsourcing for disease mention annotation in PubMed abstracts. Pacific symposium on biocomputing. 2015:282–293. https://doi.org/10.1142/9789814644730_0028.
30. **Harber P, Leroy G.** Assessing work–asthma interaction with Amazon Mechanical Turk. Journal of occupational medicine. 2015;57(4):381–385. https://doi.org/10.1097/JOM.0000000000000360.
31. **Harris JK, Mart A, Moreland-Russell S, Caburnay CA.** Diabetes topics associated with engagement on Twitter. Preventing chronic disease. 2015;12:E62. https://doi.org/10.5888/pcd12.140402.
32. **Hipp JA, Manteiga A, Burgess A, Stylianou A, Pless R.** Webcams, crowdsourcing, and enhanced crosswalks: Developing a novel method to analyze active transportation. Frontiers in public health. 2016;4:1–9. http://journal.frontiersin.org/article/10.3389/fpubh.2016.00097.

33. **Holst D, Kowalewski TM, White LW, et al.** Crowd-Sourced Assessment of Technical Skills (C-SATS): Differentiating animate surgical skill through the wisdom of crowds. Journal of endourology. 2015;29(10):1183–8. https://doi.org/10.1089/end.2015.0104.
34. **Khare R, Burger JD, Aberdeen JS, et al.** Scaling drug indication curation through crowdsourcing. Database. 2015;2015:bav016. https://doi.org/10.1093/database/bav016.
35. **Kim HS, Hodgins DC.** Reliability and validity of data obtained from alcohol, cannabis, and gambling populations on Amazon's Mechanical Turk. Psychology of addictive behaviors. 2017;31(1):85–94. https://doi.org/10.1037/adb0000219.
36. **Kuang J, Argo L, Stoddard G, Bray BE, Zeng-Treitler Q.** Assessing pictograph recognition: A comparison of crowdsourcing and traditional survey approaches. Journal of medical internet research. 2015;17(12):e281. https://doi.org/10.2196/jmir.4582.
37. **Lee AY, Lee CS, Keane PA, Tufail A.** Use of Mechanical Turk as a MapReduce framework for macular OCT segmentation. Journal of ophthalmology. 2016. https://doi.org/10.1155/2016/6571547.
38. **Lloyd JC, Yen T, Pietrobon R, et al.** Estimating utility values for vesicoureteral reflux in the general public using an online tool. Journal of pediatric urology. 2014;10(6):1026–1031. https://doi.org/10.1016/j.jpurol.2014.02.014.
39. **MacLean DL, Heer J.** Identifying medical terms in patient-authored text: a crowdsourcing-based approach. Journal of the american medical informatics association. 2013;20(6):1120–1127. https://doi.org/10.1136/amiajnl-2012-001110.
40. **Mitry D, Peto T, Hayat S, et al.** Crowdsourcing as a screening tool to detect clinical features of glaucomatous optic neuropathy from digital photography. PLoS One. 2015;10(2):1–8. https://doi.org/10.1371/journal.pone.0117401.
41. **Mitry D, Zutis K, Dhillon B, et al.** The accuracy and reliability of crowdsource annotations of digital retinal images. Translational vision science & technology. 2016;5(5):6. https://doi.org/10.1167/tvst.5.5.6.
42. **Mortensen JM, Musen MA, Noy NF.** Crowdsourcing the verification of relationships in biomedical ontologies. AMIA Annual symposium proceedings. 2013;2013:1020–1029.
43. **Powers MK, Boonjindasup A, Pinsky 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. Journal of endourology. 2016;30(4):447–452. https://doi.org/10.1089/end.2015.0665.
44. **Santiago-Rivas M, Schnur JB, Jandorf L.** Sun protection belief clusters: Analysis of Amazon Mechanical Turk data. Journal of cancer education. 2016;31(4):673–678. https://doi.org/10.1007/s13187-015-0882-4.
45. **Schleider JL, Weisz JR.** Using Mechanical Turk to study family processes and youth mental health: A test of feasibility. Journal of child and family studies. 2015;24(11):3235–3246. https://doi.org/10.1007/s10826-015-0126-6.
46. **Shao W, Guan W, Clark MA, et al.** Variations in recruitment yield, costs, speed, and participant diversity across internet platforms in a global study examining the efficacy of an HIV/AIDS and HIV testing animated and live-action video. Digital culture & education. 2015;7(1):40–86.
47. **Turner AM, Kirchhoff K, Capurro D.** Using crowdsourcing technology for testing multilingual public health promotion materials. Journal of medical internet research. 2012;14(3):e79. http://www.jmir.org/2012/3/e79/.
48. **White LW, Kowalewski TM, Dockter RL, Comstock B, Hannaford B, Lendvay TS.** Crowd-Sourced Assessment of Technical Skill: A valid method for discriminating basic robotic surgery skills. Journal of endourology. 2015;29(11):1295–1301. https://doi.org/10.1089/end.2015.0191.
49. **Wu C, Scott Hultman C, Diegidio P, et al.** What do our patients truly want? Conjoint analysis of an aesthetic plastic surgery practice using internet crowdsourcing. Aesthet Surg J. 2017;37(1):105–118. https://doi.org/10.1093/asj/sjw143.
50. **Wymbs BT, Dawson AE.** Screening Amazon's Mechanical Turk for adults with ADHD. J Atten Disord. 2015:1–10. https://doi.org/10.1177/1087054715597471.
51. **Yu B, Willis M, Sun P, Wang J.** Crowdsourcing participatory evaluation of medical pictograms using Amazon Mechanical Turk. Journal of medical internet research. 2013;15(6):e108. http://www.jmir.org/2013/6/e108/.