

Prevalence of Ethical Oversight Among Studies Using Google Trends Data: a Cross-sectional Analysis

Mitchell Love, B.S., Office of Medical Student Research., Oklahoma State University Center for Health Sciences

Taimoor Khan, B.S., Office of Medical Student Research., Oklahoma State University Center for Health Sciences

Nicholas B. Sajjadi, B.S., Office of Medical Student Research., Oklahoma State University Center for Health Sciences

Ryan Ottwell, B.S., Office of Medical Student Research., Oklahoma State University Center for Health Sciences

Matt Vassar, Ph.D., Office of Medical Student Research & Department of Psychiatry and Behavioral Sciences., Oklahoma State University Center for Health Sciences.,

Jason Beaman, D.O., MPH, MS, Office of Medical Student Research & Department of Psychiatry and Behavioral Sciences., Oklahoma State University Center for Health Sciences.,

Micah Hartwell, Ph.D. Office of Medical Student Research & Department of Psychiatry and Behavioral Sciences., Oklahoma State University Center for Health Sciences.,

Corresponding Author:

Mitchell Love, Oklahoma State University Center for Health Sciences

Conflicts of Interest: No financial sources of support were provided during the conduct of this study. Matt Vassar reports grant funding from the National Institutes of Health, the U.S. Office of Research Integrity, and Oklahoma Center for the Advancement of Science and Technology, all outside the present work. Micah Hartwell receives research support through the US Department of Justice. All other authors have nothing to report related to the current study.

Abstract:

Background

The extent of ethical oversight among internet studies utilizing Google Trends (GT) data is largely unknown. GT data in prior medical research has provided a deeper understanding of public interest in and awareness of key medical issues, emphasizing the importance of ensuring ethical conduct in this field of research. Thus, we investigated the prevalence of Institutional Review Board (IRB) submission rates among GT studies.

Design

A systematic search of PubMed was conducted for observational studies using GT data published after 2012. We randomized and screened 563 articles in a masked, duplicate fashion. Title, PMID, publishing journal, publishing date, primary author credentials and country, potential correspondence for outside data, IRB statement, IRB sponsor, and funding statement were extracted. Reporting frequencies were calculated, and chi-square tests were conducted.

Results

Of the 76 included studies, 3 (3.95%) declared submission to an IRB, 11 (14.47%) declared no submission, and 62 (81.58%) made no declaration. Additionally, 30 (39.47%) reported a funding source, 11 (14.47%) reported no funding source, and 35 (46.05%) did not mention funding. Study funding correlated with likelihood of reporting ethical oversight ($\chi^2=9.9$, $P=0.043$).

Conclusions

Our findings suggest low reporting of ethical oversight among GT studies, possibly resulting from poor methodological documentation or widespread unfamiliarity of IRBs with internet research. Diminished submission likelihood among unfunded studies is possibly due to financial and temporal constraints. We recommend using stratified ethical review boards permitting some research to undergo truncated review, or otherwise clearly reporting the nature of review, encouraging production and limiting waste.

Introduction:

Conducting ethical research is critical for defining research goals, ensuring sustainable environments that foster scientific progress, and providing barriers against unethical practices leading to misinformation, erroneous conclusions, and harmful outcomes¹. Heinous breaches of human rights during research, such as those seen in the Tuskegee Syphilis Experiment² and others^{3,4}, emphasized the need to enhance and enforce ethical standards in research practice. The National Research Act of 1974 addressed this issue in the US by establishing Institutional Review Boards (IRBs) to review and oversee research projects' adherence to ethical standards.⁵ Title 45 Code of Federal Regulations Part 46 (45 CFR 46) is a federal law requiring institutions conducting research on human subjects (HSR) to receive ethical approval from an IRB before projects can begin⁶. IRB approval is now required in over 80 countries, thus reinforcing the established ethical precedent for all HSR.⁷

While IRB oversight has fortunately become ubiquitous, its role in the oversight of non-HSR — studies in which investigators do not obtain, study, analyze, or generate information through intervention or interaction with a living human being, including identifiable private information — remains unclear and even varies among IRBs⁸, specifically regarding studies using publicly available internet datasets. Google Trends (GT) is a source of such data and is becoming increasingly useful as a medical research tool.⁹ For example, analysis of GT data by Hartwell et al. revealed an association of environmental activist Greta Thunberg’s public appearances with spikes in Google searches for “Asperger Syndrome,” thus demonstrating the influence she, an individual living with Asperger Syndrome, may have had on the public’s awareness of psychiatric disorders.¹⁰ In another study involving public interest in a *Grey’s Anatomy* episode portraying domestic violence, the authors collaborated with the Rape, Abuse, & Incest National Network to collect and analyze data revealing significant increases in website traffic and call volume corresponding to increased Google searches for related terms following the episode’s airing¹¹. In another case, Allem et al. presented evidence of increased sales of at-home Human Immunodeficiency Virus tests following Charlie Sheen’s disclosure of his serostatus.¹² These examples highlight the importance of GT data even as its promising utility in medical research is continually discovered.

The recent and rapid emergence of GT studies in medical research raises questions regarding ethical dilemmas that may require IRB oversight despite the lack of human subjects. GT data is also frequently used in conjunction with secondary analyses performed on other publicly available datasets, such as the Behavioral Risk Factor Surveillance System, the Youth Risk Behavior Surveillance System, and the National Health and Nutrition Examination Survey — all of which have been deemed as not requiring IRB approval for use in research^{6,13–15}. All of the aforementioned GT studies were published declaring that their study design did not constitute HSR, and thus were not subjected to IRB oversight. However, the extent of submission to an IRB for the designation of non-HSR among GT studies is currently unknown. Thus, the primary aim of this study is to determine the IRB submission rates of cross-sectional analyses utilizing GT data. Findings from this study may help to identify barriers to efficient research production as well as to contribute evidence for use in establishing guidelines for ethically conducting internet research.

Methods:

Systematic Search and Eligibility: A systematic search of the PubMed database was conducted on 12/23/2020 for the terms "google trends", "*google.trends*", "*trends.google*", and "'google' AND 'trends'" published after 2012. All search results were downloaded as a CSV file and imported into Stata 16.1 (College Station, TX). To include journals that frequently publish Google Trends studies, we included journals with at least 5 search returns. Further we included only cross-sectional studies of GT data, excluding editorials and comments.

Data Extraction: Two of us (ML and TK) were trained on screening cross-sectional/observational study designs on 12/23/2020. We then conducted data extraction in a masked, duplicate fashion, in accordance with best practices as noted in the Cochrane Handbook (CITE) using a pilot-tested Google Form. Extracted data included article title, name of the journal, Year of publication, credentials of the first author, article type, external contact for additional data procurement, the

dataset used, country of the primary author, IRB statement, sponsor of IRB, and funding information. To extract IRB statements, we systematically searched each article for the terms “IRB”, “Review board” “Human Subjects” and “ethic*” (For Ethic (s, al) Approval).

Statistical Analysis: Frequencies and percentages were calculated for all extracted characteristics. Further, we determined rates of IRB submission and non-submission (PI determination of non-HSR), IRB determinations and decisions. We then used chi-square or fisher’s exact to measure associations between IRB submission and institution and county of first authors (and IRB, if noted) and extracted journal and study characteristics.

Reproducibility and Ethics Review: This protocol was uploaded to the Open Science Framework (OSF, <https://osf.io/zhb2g>) to ensure reproducibility and allow all record keeping for data inquiries. This study was submitted to Oklahoma State University Center for Health Sciences Institutional Review Board as a university requirement and was declared to be non-human subjects research according to HHS 45 CFR 46.²

Results:

The search of PubMed returned 1,703 results from 859 different journals. After excluding journals with fewer than 5 entries, 563 articles from 56 journals remained for screening. Following the screening process, 76 studies were included in our analysis. The majority of studies generated from the United States (25, 32.9%), followed by the United Kingdom with 9, South Korea, Italy, and India with 4 each, and 19 other countries with 3 or fewer. Of the 76 studies, 35 made no mention of funding support for the research, 30 reported having funding for study, and 11 reported that the study was not funded.

Of the 76 studies, 3 (3.95%) reported submitting for IRB approval, 11 (14.47%) reported not submitting for IRB approval, and 62 (81.58%) made no declaration. These results and funding data for included studies are summarized in Table 1 and Figure 1. Each of the 3 studies reporting submission for IRB approval were published in journals requiring a formal ethics determination made by IRB (Table 2). We identified 3 studies that curated additional data through correspondence with individuals or direct surveys of individuals— two of which made no mention of IRB approval, and 1 declaring it was not subject to oversight according to the Ethics Committee of the Capital Region of Denmark (Section 14 (1) of the Committee Act. 2)¹⁶

Table 1. Journals in which submission to the IRB statement was included and the institutions that provided it.

<i>Article Title</i>	<i>Journal</i>	<i>Journal Requirement?</i>	<i>Institution Providing Requirement</i>
Respiratory Syncytial Virus Tracking Using Internet Search Engine Data	BMC Public Health	Yes	University of Arizona
Analysis of Dermatologic Conditions in Turkey and Italy by using Google Trends Analysis in the Era of the COVID-19 Pandemic	Dermatologic Therapy	Yes	Usak University, Ministry of Health Scientific Research
Tracking Internet Interest in Anabolic-Androgenic Steroids using Google Trends	Int. J Drug Policy	Yes	University of Glasgow

Figure 1: Rates of submission to IRB among studies using GoogleTrends data.

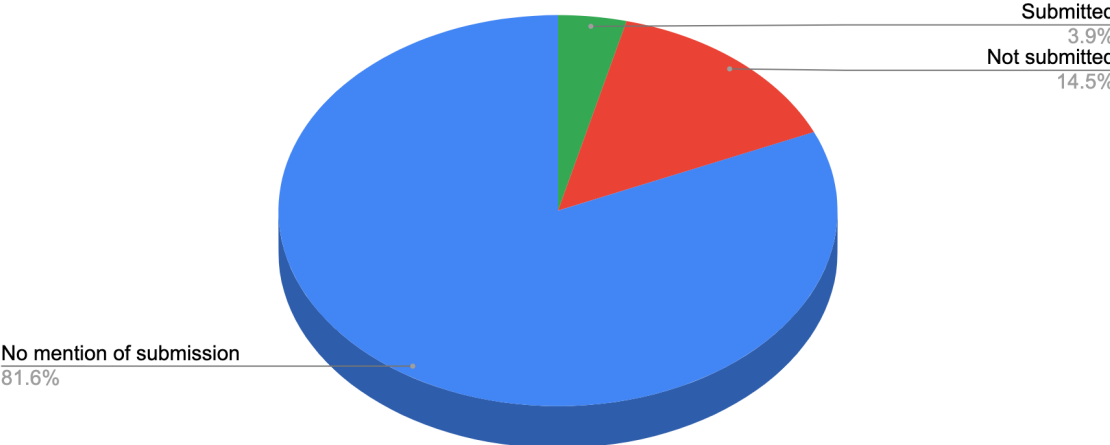
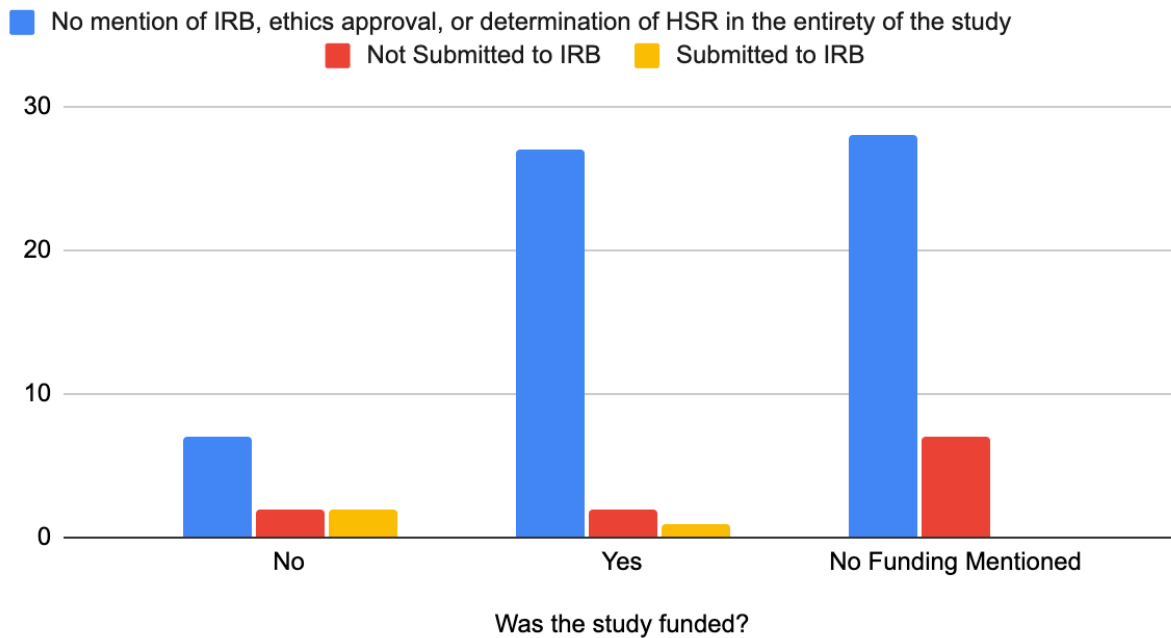


Table 2. Sources of supporting information and methods of data collection.

<i>Article Title</i>	<i>Type of Additional Data Collected</i>	<i>Type of Correspondence</i>	<i>Who was contacted?</i>
Computational Models of Consumer Confidence from Large-scale Online Attention Data: Crowd-sourcing Econometrics	Consumer Confidence Data	CCI Questionnaires, Economist's Confidence Survey	National Bureau of Statistics of China
Internet Searches Offer Insight into Early-season Pollen Patterns in Observation-free Zones	National Allergy Bureau Pollen Concentration	Requested by Mail	American Academy of Allergy, Asthma, & Immunology
Public Perception of the Vegetative State/Unresponsive Wakefulness Syndrome: A Crowdsourced Study	Perception of Wakefulness Syndrome	Online Survey	Individuals via Online Survey Platform

We found a significant association between a study being funded and their reporting of ethical approval ($\chi^2=9.9$, $P= 0.043$), with 7 of the 11 (63.6%) studies without funding declaring their studies did not require IRB oversight or meet the requirement of human subjects research (Figure 2). No other significant associations were found.

Figure 2. Rate of submission to IRB vs study funding source



Discussion:

Our results suggest that the majority of institutions conducting GT studies and journals frequently publishing GT studies do not require IRB oversight or approval, which may be contributing to the low reporting of IRB involvement among GT studies. These findings may be due to the poor methodological documentation known to be problematic among GT studies as noted in a 2014 systematic review by Nuti et al¹⁷. Researchers may be submitting GT studies for IRB approval yet failing to report submission, either because it is not required by the publisher, or because IRBs are unaware of how to properly handle overseeing online research, often ignoring its complexities¹⁸. In 2006, Buchanan et al contacted over 700 IRBs in the US regarding ethical oversight of internet research finding that only 9% of respondents recommended or required ethical training for internet research. The lack of guidance is likely due to the exponential increase in internet research beginning at the onset of the millenium¹⁹, preventing IRBs from having enough time to develop ethical guidelines regarding its conduct. The novelty of internet research further complicates its insurgence, as the nuance of its ethical dilemmas aren't yet fully recognized²⁰, making guideline development even more difficult. In this context, our results may support developing methodological reporting standards for GT studies that would enhance research quality by

requiring authors to declare the nature of a study's ethical review, whether by IRBs or by investigators.

Our results suggest that unfunded studies were less likely to report IRB oversight and to report having not submitted to IRB for designation as non-HSR. This may be partly due to the limited resources available to unfunded researchers. In the case of HSR, associated costs of ethical review are miniscule in comparison to the value of ensuring the ethical conduct of research and for protecting the subjects, the institution, and the investigators, though the same cannot be said of low-risk non-HSR. Temporal constraints of the review process may also hinder progress. Inefficient and inconsistent behaviors of participating IRBs, such as misplaced paperwork, reiterations to settle issues, and reviews taking longer than expected, have previously been reported²¹. Kano et al. characterized the aforementioned financial and temporal barriers among 89 US IRBs overseeing the conduct of low-risk medical student research, pointing out inconsistencies in rulings and often insurmountable costs in the absence of funding²². Institutions have sought to ameliorate this issue by using stratified ethical review boards which permit some research to undergo limited review, or even to totally forgo review, depending on the nature of the study or of the data being used²³. This may involve principal investigators making legitimate judgement calls based more on threats to subject anonymity or the potential for harm as opposed to a human subjects designation. Of course, allowing investigators to make non-HSR is not without risks, including factors such as physical, psychological, social, and private harm, as well as legal ramifications. However, research waste could be minimized by allowing studies to be either safely subjected to limited review or exempt from it entirely.

The utility and versatility of GT data warrants measures to ensure its place in scientific research while being subjected to the same ethical standards that are expected universally, as is the case for other forms of internet research²⁴. An example of GT utility was demonstrated with its use as a public health surveillance tool related to the Ebola virus outbreak in 2014²⁵. Alicino et al evaluated the relative search volume of the term "Ebola" during this epidemic, discovering a strong correlation between the number of registered Ebola cases and the relative search volumes in three African countries. Further, monitoring GT data spoke to the influence of media coverage of the virus, as peaks in search queries were observed as the disease was first reported outside of Africa during the latter half of 2014²⁵. The importance of GT studies is growing, as widespread patient access to health information via the internet presents new benefits²⁶ and challenges^{27,28}, the understanding of which could equip clinicians and researchers with beneficial knowledge leading to higher quality care and more effective public health initiatives. Any hindrance to the ethical and efficient production of GT studies should be a serious concern for those looking to use this tool to add to the literature base.

Ethical conduct is paramount in research, regardless of whether human subjects are involved. While the human subjects designation guides ethical social, clinical, and biomedical research, its role in internet research is often debated as being unfit for non-biomedical studies or studies that do not involve human interaction at all²⁹, as is generally the case with secondary data analysis and methodological research, particularly in the case of GT. We recommend the development of institutionally standardized protocols allowing junior investigators and PI's to make the non-HSR determination of GT studies as a feasible and effective way to reduce waste and encourage research production. Any rare or questionable cases could be submitted to an IRB, thus limiting this expense

to fewer studies. Additionally, if novel protocols arise for which there are no precedents and investigators intend to conduct multiple studies using the same protocol, only the initial protocol should undergo review, preventing redundant review of an unchanging protocol. A similar approach was successfully carried out in 2015 by the University of Iowa College of Nursing and the Human Subjects Office, which developed a decision algorithm for use by Doctor of Nursing Practice students to make the non-HSR designation while requiring attestation of adherence to the protocol. At the end of 2 years, 96.3% of their projects were deemed as non-HSR, and the institution concluded that the process required less time by students, faculty, and the IRB in preparing and processing review requests, thus leading to the timely review of research projects³⁰. If developing such a determination algorithm is not feasible, we recommend that all studies be submitted to an IRB for review to avoid discrepancies, and that submission be stated in the methodology section of GT study protocols so that the nature of review is certain and documented.

Strengths and Limitations

Limitations of this study include the relatively small sample size. This may be due in part to the novelty of GT studies, which have only recently become more common. Another limitation is the general lack of knowledge regarding the ethical dilemmas of internet research which may leave IRB evaluation, or any proposed ethical review protocol, ineffective in preventing ethical infraction. However, we believe the benefits of these studies far outweigh the risks. A strength of this study includes being the first investigation to characterize the state of ethical oversight of GT studies, which may serve to raise awareness and establish sound ethical practices, securing GT studies' place in medical research. We also conducted the screening and extracting in a masked, duplicate fashion by two investigators. Finally, we adhered to a pre-established protocol to improve research transparency and integrity.

Conclusion

Internet research is becoming a prominent tool in the medical field, allowing researchers to understand societal interactions with medical information on large scales. Thus, ensuring ethical practices will establish the place of internet studies in research methodology, preserving its integrity and validity and allowing its maximum potential to be discovered. Currently, GT studies reside in an ethical grey area. Our research suggests that these studies are less likely to have undergone ethical review and to have mentioned their exemption from review based on local ethical protocols. This could be due to widespread IRB unfamiliarity with GT studies or due to the resource barriers researchers potentially encounter in obtaining ethical approval of non-HSR. Regardless, ethical evaluation of GT studies is critical. We recommend the use of stratified ethical review boards to preserve the advantages of the IRB approval process while mitigating its limitations. If developing such a protocol is unfeasible, third party ethical oversight of GT studies should be conducted.

References

1. What is ethics in research & why is it important? nih.gov. Accessed January 9, 2021. <https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm>
2. Brandt AM. Racism and research: the case of the Tuskegee Syphilis Study. *Hastings Cent Rep.* 1978;8(6):21-29.
3. Zimbardo PG. On the ethics of intervention in human psychological research: with special reference to the Stanford Prison Experiment. *Cognition.* 1973;2(2):243-256.
4. McLeod SA. The milgram experiment. *Simply Psychology.* Published online 2007. http://www.edmotivate.com/uploads/4/7/6/4/47648491/milgram_experiment.pdf
5. Moon MR. The History and Role of Institutional Review Boards: A Useful Tension. *AMA Journal of Ethics.* 2009;11(4):311-316.
6. Electronic Code of Federal Regulations. Accessed December 23, 2020. <https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML>
7. Grady C. Institutional Review Boards: Purpose and Challenges. *Chest.* 2015;148(5):1148-1155.
8. Vitak J, Proferes N, Shilton K, Ashktorab Z. Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards. *J Empir Res Hum Res Ethics.* 2017;12(5):372-382.
9. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. *JMIR Public Health and Surveillance.* 2019;5(2): e13439. doi:10.2196/13439
10. Hartwell M, Keener A, Coffey S, Chesher T, Torgerson T, Vassar M. Brief Report: Public Awareness of Asperger Syndrome Following Greta Thunberg Appearances. *J Autism Dev Disord.* Published online August 18, 2020. doi:10.1007/s10803-020-04651-9
11. Torgerson T, Khojasteh J, Vassar M. Public Awareness for a Sexual Assault Hotline Following a Grey's Anatomy Episode. *JAMA Intern Med.* 2020;180(3):456-458.
12. Allem J-P, Leas EC, Caputi TL, et al. The Charlie Sheen Effect on Rapid In-home Human Immunodeficiency Virus Test Sales. *Prev Sci.* 2017;18(5):541-544.
13. Approved Sources of Public Use Data. Accessed December 23, 2020. <https://louisville.edu/research/humansubjects/lifecycle/approved-sources-of-public-use-data>
14. Public use datasets - research. Published June 18, 2019. Accessed December 23, 2020. <http://research.fiu.edu/irb/public-use-datasets/>
15. GUIDANCE: Data Sets Not Requiring IRB Review. Accessed December 23, 2020.

<https://www.nber.org/guidance-data-sets-not-requiring-irb-review>

16. National Committee on Health Research Ethics – nvk.dk. Accessed January 12, 2021. <http://www.nvk.dk/english>
17. Nuti SV, Wayda B, Ranasinghe I, et al. The use of google trends in health care research: a systematic review. *PLoS One*. 2014;9(10): e109583.
18. Buchanan EA, Ess CM. Internet research ethics and the institutional review board: current practices and issues. *SIGCAS Comput Soc*. 2009;39(3):43-49.
19. Denissen JJA, Neumann L, van Zalk M. How the internet is changing the implementation of traditional research methods, people's daily lives, and the way in which developmental scientists conduct research. *Int J Behav Dev*. 2010;34(6):564-575.
20. Watson M, Jones D, Burns L. Internet research and informed consent: An ethical model for using archived emails. *Int J Ther Rehabil*. 2007;14(9):396-403.
21. Silberman G, Kahn KL. Burdens on research imposed by institutional review boards: the state of the evidence and its implications for regulatory reform. *Milbank Q*. 2011;89(4):599-627.
22. Kano M, Getrich CM, Romney C, Sussman AL, Williams RL. Costs and inconsistencies in US IRB review of low-risk medical education research. *Med Educ*. 2015;49(6):634-637.
23. Annette Markham, Umea University, Sweden & Loyola University, Chicago Elizabeth Buchanan, University of Wisconsin-Stout, USA. Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0). aoir.org. Accessed January 14, 2021. <https://aoir.org/reports/ethics2.pdf>
24. Williams S. Williams, SG (June 2012). *The Ethics of Internet Research Online Journal of Nursing Informatics (OJNI)*, 16 (2). Published online 2012.
25. Alicino C, Bragazzi NL, Faccio V, et al. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. *Infect Dis Poverty*. 2015;4:54.
26. Jadad AR, Haynes RB, Hunt D, Browman GP. The Internet and evidence-based decision-making: a needed synergy for efficient knowledge management in health care. *CMAJ*. 2000;162(3):362-365.
27. Miller MP, Arefanian S, Blatnik JA. The impact of internet-based patient self-education of surgical mesh on patient attitudes and healthcare decisions prior to hernia surgery. *Surg Endosc*. 2020;34(11):5132-5141.
28. Zhou H, Zhang J, Su J. Internet access, usage and trust among medical professionals in China: A web-based survey. *Int J Nurs Sci*. 2020;7(Suppl 1): S38-S45.

29. Bassett EH, O’Riordan K. Ethics of Internet research: Contesting the human subjects research model. *Ethics Inf Technol*. 2002;4(3):233-247.
30. Foote JM, Conley V, Williams JK, McCarthy AM, Countryman M. Academic and Institutional Review Board Collaboration to Ensure Ethical Conduct of Doctor of Nursing Practice Projects. *J Nurs Educ*. 2015;54(7):372-377.