Using PubMed to Generate Email Lists of Participants for Healthcare Survey Research: A Simple and Practical Approach

Mohamed KHALIFA[1]

Centre for Health Informatics, Australian Institute of Health Innovation, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia

Abstract. Survey research is one of the most essential domains of evaluation and measurement in healthcare and social sciences. Online surveys are considered the most economical of the three main data collection methods, followed by telephone interviewing, while face-to-face interviewing is the most expensive. Even though they have many advantages, online surveys have very low response rates. The objective of this paper is to demonstrate a practical and simply replicable approach for using PubMed to generate large email lists of potential participants for healthcare survey research. In addition to personalizing each email, researchers can use a range of strategies to improve the response rate, including sending reminders, adding the updated response rate to the reminders, and stating the average time it would take participants to complete the survey. Moreover, acknowledging participants, using financial and non-financial incentives and contacting participants through their affiliated organization, can significantly improve participants response rate.

Keywords. Healthcare, Survey Research, Participants Emails, Questionnaires.

1. Introduction

Survey research is one of the most essential domains of evaluation and measurement in healthcare and social sciences. This wide domain includes evaluation interventions through asking questions of participants. A survey can take a wide range of types, starting from the simplest paper-based feedback forms to the most intensive one-on-one in-depth interviews [1]. Online questionnaires have been considered the most economical of the three main data collection methods, followed by telephone interviewing, while face-to-face interviewing is the most expensive [2]. Questionnaires can support both quantitative and qualitative research methods. In addition to answering quantitative questions, respondents could provide rich, free text feedback [3]. Online questionnaires have many advantages. They are inexpensive, practical, provide a quick way to get results, scalable, provide comparability with easy analysis and visualization, offer actionable data while keeping respondent anonymity, have no time constraints, and can cover every aspect of a topic. Disadvantages include providing dishonest answers or leaving unanswered questions. Questionnaires cannot convey feelings and emotions, there is lack of personalization, unconscientious responses, in

---

[1] Corresponding Author, Mohamed KHALIFA; Email: mohamed.khalifa@mq.edu.au.

addition to the questionnaire or survey fatigue problem [4]. The lower response rate of online surveys has been a major concern for survey researchers. There is a wide variety of factors that can affect response rates, such as the content and presentation, sampling methods, contact delivery modes, invitation design, using reminders and incentives [5]. One way of overcoming the low response rate of online survey is to contact a larger sample of participants [6]. Compared to other healthcare databases, PubMed includes the largest number of publications and offers an optimal update frequency. PubMed remains the best tool in biomedical electronic research [7]. The objective of this paper is to demonstrate a practical and simply replicable approach for using PubMed to generate large email lists of potential participants for healthcare survey research.

2. Methods

In simple and replicable five steps, and as an example, we will generate an email list of emergency department clinicians and researchers who published at least one PubMed indexed paper, over the last five years. Step one includes using the advanced search function of the PubMed (https://www.ncbi.nlm.nih.gov/pubmed/advanced) to search for papers. We will use the search key words: Emergency Medicine OR Emergency Department OR Emergency Departments OR Emergency Service OR Emergency Services in the [Title/Abstract] AND the same search key words in the [Affiliation], to make sure retrieved papers not only are about emergency medicine, but also have been conducted by clinicians and researchers who belong to the emergency medicine or emergency department at their organizations. Step two includes retrieving the results from the PubMed, in the MEDLINE format, to a text file to be saved on the computer. Step three includes extracting the emails listed in the MEDLINE results text file, using a free online email extraction website, specialized in extracting emails from text files (http://convertcsv.com/email-extractor.htm). Step four includes preparing a MS Excel sheet with the extracted emails. Step five includes using the mail merge function of the MS Office, through MS Excel, Word, and Outlook, to send emails to the identified participants of the survey research. Figure 1 shows the five steps approach of identifying publications, retrieving and extracting emails of their authors.
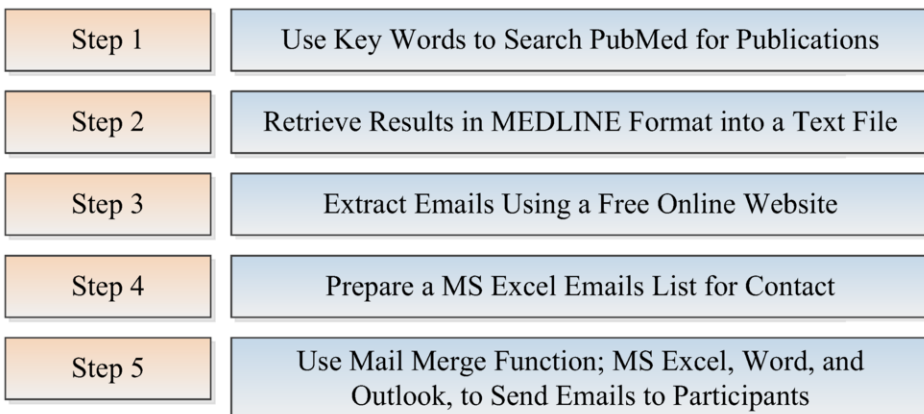
Step 1	Use Key Words to Search PubMed for Publications
Step 2	Retrieve Results in MEDLINE Format into a Text File
Step 3	Extract Emails Using a Free Online Website
Step 4	Prepare a MS Excel Emails List for Contact
Step 5	Use Mail Merge Function; MS Excel, Word, and Outlook, to Send Emails to Participants

Figure 1. Five steps approach of identifying publications, retrieving and extracting emails of their authors

3.   Results

The results of the search on the PubMed retrieved 13,608 identified papers published in the area of emergency medicine or emergency department by emergency medicine clinicians and researchers over the last five years. Retrieving the MEDLINE format results into a text file generated a 55 Mega Byte data text file containing the titles, abstracts, author information, and other relevant information of the retrieved publications. The extraction of authors' emails generated 5,817 unique emails.

4.   Discussion and Conclusion

We have described in detail a simple method to generate a large emails list of emergency medicine clinicians and researchers through using the advanced search function of the PubMed. PubMed is one of the best resources for biomedical published research. The keywords and the search fields used ensure that the identified papers are relevant to the emergency medicine. Using the last five years as a timeframe, ensures that a large percentage of the emails are still valid and attended by their owners. This method generates emails without names, which might lower the response rate. It is reported that decreasing costs associated with designing and administering online surveys would make survey fatigue more prevalent [8]. On the other hand, it is discussed by many researchers that personalizing the email invitations significantly improves the response rate of participants as it gives the impression of focused message and targeted audience [9]. However, using the mail merge function enables researchers to send a separate email directly to each participant, which is better than sending bulk emails using the blind carbon copy function. Bulk emails are sometimes identified by email agents, such as MS Outlook and Gmail, as spam and are automatically sent to the trash, or simply are overlooked by the recipients considering them as commercial, low priority or less relevant communications [10]. Sending survey invitation emails anonymously could generate a response rate between 1 to 5%, while sending personalized salutation emails could raise this rate up to 10 to 15% [11]. Alternately, the names of the researchers still can be extracted, from the MEDLINE format results file that we downloaded earlier, through checking the Affiliation (AD) field which comes after the Author (AU) field in the MEDLINE format text file. However, if this process is done manually it will take quite a long time, 30 seconds on average for each email, which makes the total time needed to extract the names for the 5,817 emails identified earlier about 48 hours of work. This time can be much reduced if a programming code was written for this task, which is out of the cope of this paper.

   In addition to personalizing each email by including the correct salutation and last name of the recipient, researchers can use a range of strategies to improve response rate of online surveys. These strategies include sending up to three reminders two weeks apart, adding the updated response rate to the reminder emails to encourage others to participate when they see the study is progressing, and stating the average time it would take a participant to complete the survey in the text of the invitation email [12]. Providing the participants with the option to request to be acknowledged in the publication of the study could also provide a good incentive to improve their response rate, especially when they are active researchers [13]. Moreover, financial and non-financial incentives of participants proved to have a significantly positive influence on improving their response rate [14]. Studies show that online survey response rates can

be increased by low to no cost incentive and text appeal strategies. There is a robust evidence that both low-cost cash prize lottery incentives and cost-free text appeal interventions targeting an individual's egotistic need for approval may increase the survey response rate [15]. It is also suggested that contacting participants through their affiliated organizations, such as professional bodies, associations, and colleges, can significantly improve their response rate to the online survey. Because this would enhance the credibility of the survey study, increase the trust of participants and their commitment towards completing the survey. However, this necessitates more efforts, resources and time to organize [16]. As a conclusion, the recently developing easiness of contacting a larger number of participants for survey research is associated with a lower response rate due to many reasons. However, we still can overcome this challenge by either generating larger participants' lists or using strategies to improve their response rates. It is true that the retrieved emails are publicly available through online publications, however, it is essential to consider different data protection and privacy regulations. It is advised that such lists should not be used for purposes other than research and that contacting each researcher should be accompanied by an explanation of how we got their emails and why we are contacting them.

References

[1] R. Kumar, *Research methodology: A step-by-step guide for beginners*, Sage Publications Limited (2019).
[2] W.E. Saris and I.N. Gallhofer, *Design, evaluation, and analysis of questionnaires for survey research*, John Wiley & Sons (2014).
[3] J. Brannen, Combining qualitative and quantitative approaches: an overview, *Mixing methods: Qualitative and quantitative research*, Routledge (2017). p. 3-37.
[4] S. Debois, Advantages and disadvantages of questionnaires, SurveyAnyPlace Blog (2016).
[5] W. Fan and Z. Yan, Factors affecting response rates of the web survey: A systematic review, *Computers in human behavior* **26** (2) (2010), 132-139.
[6] D.A. Dillman, J.D. Smyth and L.M. *Christian, Internet, phone, mail, and mixed-mode surveys: the tailored design method*, John Wiley & Sons (2014).
[7] M.E. Falagas, E.I. Pitsouni, G.A. Malietzis and G. Pappas, Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses, *The FASEB journal* **22** (2) (2008), 338-342.
[8] K. Fosnacht, S. Sarraf, E. Howe and L.K. Peck, How important are high response rates for college surveys? *The Review of Higher Education* **40** (2) (2017), 245-265.
[9] D. Heerwegh, T. Vanhove, K. Matthijs and G. Loosveldt, The effect of personalization on response rates and data quality in web surveys, *International Journal of Social Research Methodology* **8** (2) (2005), 85-99.
[10] A.N. Joinson and U.D. Reips, Personalized salutation, power of sender and response rates to Web-based surveys, *Computers in Human Behavior*, **23** (3) (2007), 1372-1383.
[11] A. Petrovčič, G. Petrič and K. L. Manfreda, The effect of email invitation elements on response rate in a web survey within an online community, *Computers in Human Behavior* **56** (2016), 320-329.
[12] J. McPeake, M. Bateson and A. O'Neill, Electronic surveys: how to maximise success, *Nurse researcher* **21** (3) (2014).
[13] C.T. Cunningham, H. Quan, B. Hemmelgarn, T. Noseworthy, C. A. Beck, E. Dixon and N. Jetté, Exploring physician specialist response rates to web-based surveys, *BMC medical research methodology* **15** (1) (2015), 32.
[14] Y.I. Cho, T.P. Johnson and J.B. VanGeest, Enhancing surveys of health care professionals: a meta-analysis of techniques to improve response, *Evaluation & the health professions* **36** (3) (2013), 382-407.
[15] M.J. Pedersen and C.V, Nielsen, Improving survey response rates in online panels: Effects of low-cost incentives and cost-free text appeal interventions, *Social Science Computer Review* **34** (2) (2016), 229-243.
[16] F. Anseel, F. Lievens, E. Schollaert and B. Choragwicka, Response rates in organizational science, 1995–2008: A meta-analytic review and guidelines for survey researchers, *Journal of Business and Psychology* **25** (3) (2010), 335-349.