

Supplemental File 2-Data Cleaning Procedure

After the survey was shared on social media via a third party (i.e., by an individual), we noticed the response rate had a large increase (>200 responses within minutes). Because of this, we paused the survey to review the responses; it raised suspicions of receiving potential responses from spambots and/or non-eligible participants seeking the incentive reward. The responses were reviewed, and it was determined that our survey was targeted by spambots. As a result, we re-opened the survey link after 24 hours, asked participants not to share the link (either personally or via social media) and created a protocol to clean the data before analysis. Specifically, we 1) removed respondents who indicate student as highest level of training; 2) removed respondents who indicate outside of Canada as main location of practice; 3) removed respondents that did not complete at least 60-100% of the survey; 4) removed any responses to the qualitative survey question “Please list 2-3 benefits of being or becoming a scholarly practitioner” and “Please list 2-3 of the most significant challenges you’ve encountered/anticipate in becoming a scholarly practitioner” that are exact duplicates or nonsensical. The remaining responses were checked for conflicting data. If any responses had two or more conflicting data, they were removed. These might include, (i) respondents with outlier response times, defined as under 12 and over 28 minutes. These time limits are based on the average time it took participants to complete the survey during the pilot testing of the survey. During this process, the average time for completion of the survey was 17.7 minutes; (ii) respondents who provided same response to every closed-ended item on a survey page (i.e., straight lining);¹ (iii) responses that are gibberish (i.e., unintelligible responses) or nonsensical responses (e.g., responses that did not make sense in the context of the items asked). For example, indicating their age is 150 years old or they’ve supervised 20,000 students in the last 5 years; (iv) respondents who provided a contact email with random letters or end in numbers exceeding four digits as these characteristics are an indication of a bot generated email address and had similar characteristics of examples from Gmail bulk account creators that can be built or bought online²; (v) any direct duplicate emails that respondents included for an incentive prize. Finally, to claim incentives, respondents had to provide their full name and province of practice. With that information, they were cross-checked in their respective regulatory member public registry as proof that they were RTs. If they could not be cross-checked and would not provide proof of licensure, their data were removed.

Reference

1. Kim Y, Dykema J, Stevenson J, Black P, Moberg D. Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys. *Social Science Computer Review*. 2019;37(2):214-233. doi:10.1177/0894439317752406.
2. Wang Z, Qin M, Chen M, Jia C. Hiding Fast Flux Botnet in Plain Email Sight. *ATCS/SePrIoT@SecureComm*. 2017.