PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Statistical tools used for analyses of frequent users of emergency department: A scoping review
AUTHORS	Chiu, Yohann; Racine-Hemmings, François; Dufour, Isabelle; Vanasse, Alain; Chouinard, Maud-Christine; Bisson, Mathieu; Hudon, Catherine

VERSION 1 - REVIEW

REVIEWER	Ambikaipakan Senthilselvan University of Alberta Canada
REVIEW RETURNED	07-Jan-2019

GENERAL COMMENTS	The reviewer does not understand the need for this review. Did the authors expect the the statistical tools used in identifying variables associated with frequent ED use (or predicting the risk of becoming a frequent user) be different from those used for other types of outcome. Without looking at conclusions from this review, the reviewer could predict that chi-square test and logistic regression would be the most frequently used statistical tools.The review does not add any new knowledge to the current literature.

REVIEWER	Eric Christensen University of Minnesota, USA
REVIEW RETURNED	12-Jan-2019

GENERAL COMMENTS	The objective of this study was to provide an overview of the statistical tools used in analyses of frequent ED users. I commend the authors for a thorough, well-researched, well-executed, well-written study. While self-serving, I was pleased that my own study of frequent ED users made the cut for inclusion. Of course, I have some comments that I hope will be helpful to the authors. Line 44. I find the word choice “consensual” as poor. I think consensus would be better. Lines 88-89. Can the authors add a sentence to describe the method YC used to exclude the 3087 studies based on the title and abstract alone? I assume that it relates to what the authors already laid out in stages 2 and 3, but a little description here would be helpful.

	Line 95. The authors write “A reference search yielded five relevant articles.” I assume they are referring to the references of the 80 articles yield the additional five articles. Please clarify. Table 1 is an excellent summary. It seems to me that the authors need to do more with the discussion section. The review of studies, while very well done, needs more in the discussion so that we can understand why it matters. The authors make the statement in lines 173-178 about mixing statistical tools for better learning. While I don’t necessarily agree or disagree, I think it should be backed up or explained more. Is this a hunch or is there more behind it? The authors discuss that machine learning is often a “black box” approach—it works but no one knows why (lines 187-190). Being curious, I want to know why, so I am skeptical of these. Can the authors address this issue? Also, I would like to see the authors address in the discussion section the application of models for frequent ED use. What tools can or would be best to incorporate in some automated fashion into health systems or insurers IT tools? I expected before reading the article that the authors would indicate what were the most effective from all of the tools used to predicted frequent ED use. They seem to favor machine learning and quantile regression. Can you be more clear?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Comment: The reviewer does not understand the need for this review. Did the authors expect the statistical tools used in identifying variables associated with frequent ED use (or predicting the risk of becoming a frequent user) be different from those used for other types of outcome. Without looking at conclusions from this review, the reviewer could predict that chi-square test and logistic regression would be the most frequently used statistical tools. The review does not add any new knowledge to the current literature.

Response: Chi square test and logistic regression are statistical tools widely used in the literature, not only for frequent emergency department users. Though this statement would be a good guess, it was still relevant to document it with rigorous evidence. A scoping review is a thorough study defined by a rigorous methodology. Furthermore, this type of review allows synthetizing research evidence in order to identify the extent, range, and gaps in the research. Though we agree that our primary result may be intuitive, we are also convinced of the relevance of the study.

Reviewer 2

Comment: The objective of this study was to provide an overview of the statistical tools used in analyses of frequent ED users. I commend the authors for a thorough, well-researched, well-executed, well-written study. While self-serving, I was pleased that my own study of frequent ED users made the cut for inclusion. Of course, I have some comments that I hope will be helpful to the authors.

Response: We thank you for this positive and constructive feedback.

Comment: Line 44. I find the word choice “consensual” as poor. I think consensus would be better.

Response: Thank you for this suggestion. “Consensus” is indeed more appropriate; it has been changed.

Comment: Lines 88-89. Can the authors add a sentence to describe the method YC used to exclude the 3087 studies based on the title and abstract alone? I assume that it relates to what the authors already laid out in stages 2 and 3, but a little description here would be helpful.

Response: YC read the title and abstract for the 4,534 studies. Based on the title and abstract, any study that were explicitly not appropriate for the scoping review was discarded (e.g. studies about frequent use of inpatient services, descriptive studies, etc.). In case of uncertainty, the study was included to be read at the next stage. We completed the selection process (lines 90 92).

Comment: Line 95. The authors write “A reference search yielded five relevant articles.” I assume they are referring to the references of the 80 articles yield the additional five articles. Please clarify.

Response: We added some details (lines 98-100).

Comment: Table 1 is an excellent summary.

Response: We thank you for this feedback.

Comment: It seems to me that the authors need to do more with the discussion section. The review of studies, while very well done, needs more in the discussion so that we can understand why it matters. The authors make the statement in lines 173-178 about mixing statistical tools for better learning. While I don’t necessarily agree or disagree, I think it should be backed up or explained more. Is this a hunch or is there more behind it?

Response: This is a relevant comment. We believe that each statistical tool can help in discovering different but nonetheless significant results. For instance, while logistic regression is widely (and appropriately) used to assess the relationship between a binary dependent and some independent variables, it is rarely used for automatic variable selection or for evaluating the importance of each independent variable. Those steps can be performed with other tools (lasso regression and random forest, for instance), hence our statement that mixing them can help achieving different goals. We added those points to the discussion (lines 196-205).

Comment: The authors discuss that machine learning is often a “black box” approach—it works but no one knows why (lines 187-190). Being curious, I want to know why, so I am skeptical of these. Can the authors address this issue?

Response: We thank you for this suggestion. We added some details as to why machine learning is often qualified as a “black box” (lines 215-226), though it still displays good performances. In this context, this emphasizes the need for researcher and clinicians to work together.

Comment: Also, I would like to see the authors address in the discussion section the application of models for frequent ED use. What tools can or would be best to incorporate in some automated fashion into health systems or insurers IT tools? I expected before reading the article that the authors would indicate what were the most effective from all of the tools used to predicted frequent ED use. They seem to favor machine learning and quantile regression. Can you be more clear?

Response: We do not favor one approach or the other; in fact, we believe that they are complementary and should be used together (lines 193-205). Our philosophy in the discussion was to identify relevant statistical tools not yet or scarcely used in studies about frequent ED users. For

instance, quantile regression is more adapted to studying a continuous dependent variable that exhibits heterogeneous relationships with the independent variables. Regarding machine learning tools, they require large datasets but can predict accurately any type of outcome. Health data are becoming more diverse and available, thus we believe that machine learning are adequate for those types of data. Our key message in the discussion is that some statistical tools can be more effective than others depending on the objective, but mixing them may be more appropriate than relying on only one.