PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Identifying the predictors of avoidable emergency department attendance after contact with the NHS 111 phone service: Analysis of 16.6 million calls to 111 in England in 2015-17
AUTHORS	Egan, Mark; Murar, Filip; Lawrence, James; Burd, Hannah

VERSION 1 - REVIEW

REVIEWER	Hamde Nazar Newcastle University, UK
REVIEW RETURNED	12-Jul-2019

GENERAL COMMENTS	I thoroughly enjoyed reading this manuscript. The need for this work is well justified and the methodology well articulated to explore and achieve this aim. The analyses is extensive with some sophisticated tests which I recommend need specialist review. The findings reported are sufficiently important to inform practice and service optimisation. Despite being complicated in places due to the nature of the data and analyses, the work is well written and presented. I do not have any suggestions for revisions for this work.

REVIEWER	Michael Schull ICES, Toronto University of Toronto
REVIEW RETURNED	26-Jul-2019

GENERAL COMMENTS	Thanks for the opportunity to review this paper. The following comments are offered to potentially strengthen it. 1) In the abstract, it is confusing as to whether your denominator is all NHS 111 calls, or only those in which patients were not advised to go to the ED. This could be clarified. 1a) It would be helpful to understand whether you prespecified any hypothesis(es). 2) how was it determined that 50% of ED visits received "no treatment, or advice and guidance only"? Does this also mean no tests (eg blood or imaging) to rule-out serious illness? The fact that no treatment was given does not necessarily equate with avoidable ED visits (eg chest pain with a rule-out AMI set of investigations may require no treatment).

	3)Can you explain why you chose a time period from March 2015 to October 2017? 4) You state the NHS 111 dataset was initially about 18M out of 38M calls; based on what criteria did you exclude the 20M calls? 5) You linked NHS 111 call records to SUS data to determine whether an ED visit was made, but does the SUS data give the type of clinical data to determine the avoidability of the ER visit? 6) As part of your outcome definition, you defined avoidable based on HRG VBO7Z, VBO8Z, VBO9Z and it is not clear why these are considered necessarily avoidable 7) Please define "clinical input" during a NHS 111 call 8) It is quite striking that for every 1000 NHS 111 calls, 129 had an ambulance dispatched and yet 32 were still considered "avoidable". Similarly, of 92 patients told to attend ED, 39 were deemed avoidable. This suggests the definition of avoidable based on post-hoc disposition (ie after the clinical assessment) may be an over-estimate of the number of visits that did not need to attend an ED. 9) Please provide a rationale for the GBM analysis. What is the advantage or difference compared with OLS? 10) Your conclusion in the discussion that your "finding does at least suggest that 111 is not causing a large fraction of callers to inappropriately seek treatment at the ED" is not supported by your data based on your definition of avoidable ED visit. You found that, of patients NHS 111 advised not to go to ED, only 5% did so. However, of the patients NHS 111 did advise to go to ED, a substantial proportion were avoidable by your definition. Therefore, your data does indeed suggest that NHS 111 is causing a substantial number of patients to attend the ED for avoidable visits. 12) Of course, this calls into question your definition of avoidable, and indeed the post-hoc determination that many patients had "avoidable" visits comes only after a clinical evaluation, presumably face to face by an MD.

REVIEWER	Janette Turner ScHARR, University of Sheffield, UK
REVIEW RETURNED	02-Aug-2019

GENERAL COMMENTS	I think the paper could be improved with some additional clarifications of 1. How the sample of 18,127,605 was derived from the 38million calls made to NHS111 (i.e why weren't all 38 million used - what factors reduced this - was it site participation (this is hinted at later), exclusion of certain types of calls etc 2. How was the data linkage carried out - the declaration on data suggests this was done externally. Were NHS Number and other identifiers used, what was the match rate for linked records? 3. The paper makes some assumptions on the part of the reader that they understand how NHS111 works. A sentence or two earlier in the introduction, e.g explaining what a DX code is, would be helpful for a general reader

	4. For the GP practice characteristics description I couldn't understand if this is for the individual patients actual GP practice or a more general estimate for a local area 5. As a non-statistician I'd appreciate a couple of sentences about what a gradient boosting model is I have no idea! 6. The potential costs described in the results are a simple estimate of ED attendances only but it cannot be assumed this would be an absolute saving as a reduction may need to be substituted with a contact elsewhere in the health system (e.g. a primary care contact). It needs to be clear this is only potentially an ED not a system saving and that this calculation is assuming the avoidable ED attendance is an additional contact rather than an alternative contact. The objectives didn't set out to measure costs and these have been extrapolated. I think this part may actually be more suitable for the discussion as an illustration of the potential consequences of the findings rather than reported in the results. One of the interesting points is the clasification of high, medium and low risk for avoidable attendance and how characteristics of each risk level might be used in the future to refine call assessment. I'd like to see a bit more discussion of the characterstics the authors think are important within each risk level. Overall this is a well written paper with a clear and straighforward question and message that addresses a subject area which is rife with speculation about inappropriate triage but is sadly lacking in actual evidence. This analysis goes some way towards fulling the evidence gap.

REVIEWER	Xinhua Yu University of Memphis, US
REVIEW RETURNED	24-Oct-2019

GENERAL COMMENTS	This study explored the predictors of avoidable emergency department visits using national 111 triage call data. The main strengths are large dataset linked with clinical use information, and appropriate candidate predictors. However, there are a few issues the authors should address: 1) The authors excluded a large number of calls. Are those excluded calls through various criteria similar to those analyzed calls in terms age, gender, and other clinical and call characteristics? Even in large data, selection bias may still exist and worth checking. 2) About 6 million calls were excluded in the regression analysis. Are those excluded similar to those included in the regression analysis? 3) Among control variables, there were 96 different disposition codes. How were these codes used in the analysis? Did the authors regroup them or use in individual code? 4) The description of data cleaning and control variables are quite thorough. However, there is little mention of how statistical analysis was done in the method section. For example, what statistical tests were used to compare various characteristics?

	How were logistic regression and OLS regression formulated and implemented? In what software the authors used to analyze them? 5) The outcome is avoidable ED use (yes/no). However, it seems the outcome should be three levels: avoidable ED, non-avoidable ED, and no ED. It is not appropriate to combine non-avoidable ED with no ED. The multinomial logistic regression may be better suited to model the probability of avoidable ED among those not advised to attend ED. 6) The event of avoidable emergency department use rate was 5.4% for calls not advised to attend ED. Statistically, the rate is low, and OLS is unlikely appropriate. Some researchers accept OLS for binary outcomes, if the event rate is about 20-80%. Clearly, this study has a much lower rate. I would suggest using probit model to get predicted probability if the authors intended to do so. Or the authors can obtain marginal predicted probability based on logistic regression as well. 7) The authors used gradient boosting model to explore the explainable variations by each predictor (roughly speaking, relative importance of each predictor), and also use GBM to re-classify the population into three groups with high, medium, and low risk of seeking avoidable ED visit. This is appropriate, but the authors should describe in details how GBM was set up, and how the relative importance of predictors was assessed, and how these three groups were classified. Without sufficient details, it is difficult for readers understand the method behind the paper. 8) Although this is not a methodological paper, the authors compared the area under the curve (AUC) between logistic regression and GBM, and found there was only a small improvement in predictability. It is not unexpected. Therefore, is it possible present the logistic regression results only? Overall, this paper addressed an interesting and important questions about the effectiveness of 111 triage call, and potential health care cost saving and waste in the nation.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 (Hamde Nazar)

I thoroughly enjoyed reading this manuscript. The need for this work is well justified and the methodology well articulated to explore and achieve this aim. The analyses is extensive with some sophisticated tests which I recommend need specialist review. The findings reported are sufficiently important to inform practice and service optimisation. Despite being complicated in places due to the nature of the data and analyses, the work is well written and presented. I do not have any suggestions for revisions for this work.

-> We thank Dr. Nazar for taking the time to review this work and for her positive assessment of the article.

Reviewer: 2 (Michael Schull)
Thanks for the opportunity to review this paper. The following comments are offered to potentially strengthen it.
We thank Dr. Schull for his review and thoughtful feedback.

1) In the abstract, it is confusing as to whether your denominator is all NHS 111 calls, or only those in which patients were not advised to go to the ED. This could be clarified.

-> Thank you for this comment. Upon re-reading the abstract we agree this was unnecessarily confusing. We have now amended the text to read:
“Of 16,563,946 calls to 111, 12,894,561 (77.8%) were not advised to go to ED (i.e. they were advised to either attend primary care, attend another non-ED healthcare service, or to self care). Of the calls where the patient was not advised to go to the ED, 691,783 (5.4%) resulted in the patient making an avoidable ED attendance within 24 hours.”

1a) It would be helpful to understand whether you prespecified any hypothesis(es).
-> We have now added text clarifying that “The analysis was considered exploratory and consequently did not examine pre-specified hypotheses.”

2) how was it determined that 50% of ED visits received "no
treatment, or advice and guidance only"? Does this also mean no tests (eg blood or imaging) to rule-out serious illness? The fact that no treatment was given does not necessarily equate with avoidable ED visits (eg chest pain with a rule-out AMI set of investigations may require no treatment).
-> We thank the reviewer for this perceptive point.
Supplement A now provides the source for the ‘50%’ statistic - it draws on Hospital Episode Statistics 2016-17 data, and is created by examining the proportion of attendances where the “First A&E Treatment '2 character description field'” recorded that the treatment administered during the attendance was “Guidance/advice only” or “None (consider guidance/advice option).”

However, the reviewer is correct that this coding may not always capture cases where a patient received a medical test prior to receiving one of these codes. We have therefore included an additional statistic which examines the first ED investigation code (rather than treatment code) assigned to each attendance, and which clarifies that when using this metric, a large proportion (41%) of attendances have initial ED investigations recorded as ‘None’ (note these do not include cases where the investigation was recorded as ‘Null/Blank’ - these are recorded in a separate row). We consider this new statistic to be a useful complement to the existing 50% statistic stated in the introduction, in that both suggest that many ED attendances do appear to be due to low-intensity health issues, albeit with the caveat that the high-level Hospital Episode Statistics data does not allow us to cleanly parse out the attendances among this group which may still have been best suited for the ED.

The supplement reads:
“The 9.7 million figure includes only attendances with valid ED treatment records and includes attendances at both major consultant-led departments (Types 1 and 2) Minor Injury Units and Walk-in Centres (Types 3 and 4). The 50% figure is calculated by summing the rows “Guidance/advice only” and “None (consider guidance/advice option)” in Table 19 in https://files.digital.nhs.uk/publication/e/7/acci-emer-atte-eng-2016-17-data.xlsx. Similarly, Table 17 in the same dataset records that, among the 19.6 million attendances in 2016-17 with valid investigation records, the first investigation recorded for 8 million of them (41%) of them was 'None'.”

3) Can you explain why you chose a time period from March 2015 to October 2017?
4) You state the NHS 111 dataset was initially about 18M out of 38M calls; based on what criteria did you exclude the 20M calls?
-> We have now added text clarifying that the time period and sample size examined in the report involved the use of the data available at the time the research was conducted:
"The data initially contained 18,127,605 observations, where each row was a call made to 111 between 31 March 2015 and 31 October 2017. This represents almost half (47%) of the 38,585,200

million calls made to 111 between March 2015 and October 2017 - we were not able to access the full universe of calls as data from some 111 sites had not yet undergone the data linkage procedures necessary for inclusion in the analysis. For this same reason we were not able to access calls outside the stated time period."

5) You linked NHS 111 call records to SUS data to determine whether an ED visit was made, but does the SUS data give the type of clinical data to determine the avoidability of the ER visit?

The data we have does not allow us to conclusively identify certain ED attendances as being certainly 'avoidable'. However, we do follow the conventional approach used in other recent NHS research of assigning 'avoidable' status to attendees with low-intensity ED treatment/investigation codes, and which "[were] not admitted, not referred to another healthcare specialist by the ED, and did not die in the ED department. (p5)" All of these indicators (eg treatment code, whether the patient was admitted or referred) are available in the data we hold; however, we do not have access to clinical notes made about the attendances which could, in some circumstances, clarify whether certain attendances which appeared avoidable may in fact have been entirely appropriate (eg instances where the patient had been explicitly told by a healthcare professional to attend the ED).

6) As part of your outcome definition, you defined avoidable based on HRG VBO7Z, VBO8Z, VBO9Z and it is not clear why these are considered necessarily avoidable
-> Thank you for highlighting this oversight. We have now expanded text on pages 4 and 5 (tracked-changed version) to include the following:
"These HRG codes represent relatively low-intensity health assessments and were therefore considered more likely to capture attendances which could have been safely treated elsewhere - with the caveat that it is not certain that all these attendances should certainly not have attended the ED (e.g. some patients may have attended the ED at the explicit instruction of a healthcare professional, even though they ended up receiving low intensity treatment). This particular list of HRG codes was adapted from those used in a 2017 study by the North of England Commissioning Support Unit which also examined avoidable admissions."

7) Please define "clinical input" during a NHS 111 call
-> We have now expanded on this:
"a binary variable indicating whether the call had clinical input from a doctor, nurse, or other clinician (21% of calls did involve the patient speaking to a healthcare professional like this)"

8) It is quite striking that for every 1000 NHS 111 calls, 129 had an ambulance dispatched and yet 32 were still considered "avoidable". Similarly, of 92 patients told to attend ED, 39 were deemed avoidable. This suggests the definition of avoidable based on post-hoc disposition (ie after the clinical assessment) may be an over-estimate of the number of visits that did not need to attend an ED.
-> We agree with the reviewer that these figures are surprising. Regarding our use of the term 'avoidable' - we have now expanded text on pages 4 and 5 (tracked-changed version) to clarify that this term will in some cases misclassify some attendances:
"These HRG codes [those used to define attendances as 'avoidable'] represent relatively low-intensity health assessments and were therefore considered more likely to capture attendances which could have been safely treated elsewhere - with the caveat that it is not certain that all these attendances should certainly not have attended the ED (e.g. some patients may have attended the ED at the explicit instruction of a healthcare professional, even though they ended up receiving low intensity

treatment). This particular list of HRG codes was adapted from those used in a 2017 study by the North of England Commissioning Support Unit which also examined avoidable admissions."

We have also added a sentence to highlight this issue as a study limitation in the Discussion
-> "Key limitations included the relatively crude criteria we used to define ED attendances as 'avoidable' (i.e. this relied principally on post-hoc ED disposition codes and did not incorporate any clinical notes which could have provided more nuanced information about the patient's health issue), our lack of controls for other characteristics likely predictive of health behaviour (e.g. patients' education, risk aversion, and health history)"

9) Please provide a rationale for the GBM analysis. What is the advantage or difference compared with OLS?
-> We apologise for the lack of detail originally provided regarding the rationale for the gradient boosted model. We have now added the following information regarding this:
"A GBM models the outcome measure as the result of a series of decision trees. Each tree attempts to identify areas where the others make poor predictions and correct for that, resulting in generally strong predictive performance even in the presence of complex nonlinear relationships or interactions between the predictors and the outcome (a situation in which other techniques such as OLS may not perform so well). It is one of the best-performing predictive algorithms for tabular data. (o10)"

10) Your conclusion in the discussion that your "finding does at least suggest that 111 is not causing a large fraction of callers to inappropriately seek treatment at the ED" is not supported by your data based on your definition of avoidable ED visit. You found that, of patients NHS 111 advised not to go to ED, only 5% did so. However, of the patients NHS 111 did advise to go to ED, a substantial proportion were avoidable by your definition. Therefore, your data does indeed suggest that NHS 111 is causing a substantial number of patients to attend the ED for avoidable visits.
12) Of course, this calls into question your definition of avoidable, and indeed the post-hoc determination that many patients had "avoidable" visits comes only after a clinical evaluation, presumably face to face by an MD.
We thank the reviewer for these insightful points. We have now updated the text in the discussion to read:

"Although our analysis could not answer the counterfactual of "would overall avoidable ED attendances be higher or lower if the 111 service did not exist?", our findings do at least suggest that relatively few 111 patients end up making unadvised attendances at ED which could likely have been safely treated elsewhere. However, Figure 2 also found that a surprisingly large proportion of patients who were advised by 111 to attend the ED did nonetheless end up receiving low-intensity treatment (such that even these attendances were classified as 'avoidable' as defined in this study). Clarifying the precise nature of these 'advised and avoidable' attendances was outside the scope of this study but warrants further investigation".

Reviewer: 3 (Janette Turner)

1. How the sample of 18,127,605 was derived from the 38million calls made to NHS111 (i.e why weren't all 38 million used - what factors reduced this - was it site participation (this is hinted at later), exclusion of certain types of calls etc
-> We have now added text clarifying that the time period and sample size examined in the data was all the data available at the time the research was conducted:
"The data initially contained 18,127,605 observations, where each row was a call made to 111 between 31 March 2015 and 31 October 2017. This represents almost half (47%) of the 38,585,200

million calls made to 111 between March 2015 and October 2017. We were not able to access the full universe of calls as data from some 111 sites had not yet undergone the data linkage procedures necessary for inclusion in the analysis. For this same reason we were not able to access calls outside the stated time period."

2. How was the data linkage carried out - the declaration on data suggests this was done externally. Were NHS Number and other identifiers used, what was the match rate for linked records?
-> We have now added a sentence regarding how the data linkage was performed:
"Each call was linked with a Secondary Use Services (SUS) record of whether the patient attended ED within 24 hours of the 111 call. This linkage was conducted by researchers at NHS England and patients' NHS number was used as the matching variable."

3. The paper makes some assumptions on the part of the reader that they understand how NHS111 works. A sentence or two earlier in the introduction, e.g explaining what a DX code is, would be helpful for a general reader
-> Thank you for highlighting this oversight; we have now expanded text on page 4 (tracked-changes document) to explicitly note the meaning of Dx codes:
"Note that the following final dispositions, recorded at the end of 111 using so-called 'Dx' codes, were considered irrelevant to the research question [...]"

4. For the GP practice characteristics description I couldn't understand if this is for the individual patients actual GP practice or a more general estimate for a local area
-> We have now added detail to clarify that GP practices were identified at the level of individual patients.
"For each GP practice in the data, which was recorded at the individual patient level, we included…"

5. As a non-statistician I'd appreciate a couple of sentences about what a gradient boosting model is I have no idea!
-> We apologise for the lack of detail originally provided regarding the rationale for the gradient boosted model. We have now added the following information regarding this:
"A GBM models the outcome measure as the result of a series of decision trees. Each tree attempts to identify areas where the others make poor predictions and correct for that, resulting in generally strong predictive performance even in the presence of complex nonlinear relationships or interactions between the predictors and the outcome (a situation in which other techniques such as OLS may not perform so well). It is one of the best-performing predictive algorithms for tabular data."

6. The potential costs described in the results are a simple estimate of ED attendances only but it cannot be assumed this would be an absolute saving as a reduction may need to be substituted with a contact elsewhere in the health system (e.g. a primary care contact). It needs to be clear this is only potentially an ED not a system saving and that this calculation is assuming the avoidable ED attendance is an additional contact rather than an alternative contact. The objectives didn't set out to measure costs and these have been extrapolated. I think this part may actually be more suitable for the discussion as an illustration of the potential consequences of the findings rather than reported in the results.
_. We thank the reviewer for this insightful point. We have now moved the text on the costs of the ED attendances from the Results section to the Discussion section and added the reviewer's recommended caveat about the overall system cost:
"Our analysis of the largest yet published dataset of linked 111 calls and subsequent ED attendances found that, of patients not advised by 111 to go to ED, around 1 in 20 (5.4%) made an avoidable Type 1 ED attendance within 24 hours of the call. Using the NHS national tariff charges present in the data for each ED attendance, we estimate that these avoidable attendances incurred tariff costs of £65 million (£2.1 million per month) over the March 2015 to October 2017 period covered in our data. If we

extrapolate this 5.4% incidence rate of avoidable attendances to all 38,585,200 calls made to 111 between March 2015 and October 2017 (i.e. including calls not in our data), this implies £58.8m in tariff charges were incurred per year by avoidable ED attendances. The cost to the NHS as a whole, however, is likely smaller than this, since patients who do not visit a Type 1 ED may instead attend another (albeit potentially cheaper) part of the healthcare system."

One of the interesting points is the classification of high, medium and low risk for avoidable attendance and how characteristics of each risk level might be used in the future to refine call assessment.
-> We thank the reviewer for this suggestion; we have now added the following text to the discussion which suggests how this information could potentially be used to improve the 111 service:
"In terms of practical implications of this research, we suggest that analysis of newer editions of the dataset examined in this report could be used to (i) provide tailored feedback to individual 111 call handlers and local leaders of 111 services regarding the proportion of their calls which result in avoidable ED attendance soon afterwards, and (ii) communicating which calls are at high-risk of an avoidable ED attendance (e.g. potentially using a traffic-light warning system where red warnings are used to identify high-risk calls) to 111 call-handlers, who could then provide extra resource for these patients (e.g. by spending extra time providing self-care instructions or guidance on how best to secure a GP appointment)."

Overall this is a well written paper with a clear and straighforward question and message that addresses a subject area which is rife with speculation about inappropriate triage but is sadly lacking in actual evidence. This analysis goes some way towards fulling the evidence gap.
-> We thank Dr. Turner for her review and for this positive assessment of the paper.

Reviewer: 4 (Xinhua Yu)
This study explored the predictors of avoidable emergency department visits using national 111 triage call data. The main strengths are large dataset linked with clinical use information, and appropriate candidate predictors. However, there are a few issues the authors should address:

1) The authors excluded a large number of calls. Are those excluded calls through various criteria similar to those analyzed calls in terms age, gender, and other clinical and call characteristics? Even in large data, selection bias may still exist and worth checking.
-> We thank the reviewer for flagging this issue. Regarding the initial drop in the number of calls from 38.6 million total calls made to 18.1 million in our initial analysis data, we have now added the following clarification:
"The data initially contained 18,127,605 observations, where each row was a call made to 111 between 31 March 2015 and 31 October 2017. This represents almost half (47%) of the 38,585,200 million calls made to 111 between March 2015 and October 2017. We were not able to access the full universe of calls as data from some 111 sites had not yet undergone the data linkage procedures necessary for inclusion in the analysis."

Unfortunately, we do not have access to the 20.5 million calls which were not included in our analysis data, so we are unable to compare whether they were significantly different from the calls in the data. However, our understanding is that the latest version of this data-set, held by NHS England, will over time include an increasing proportion of the total number of 111 calls made, as more and more 111 sites update their data-linkage processes.

2) About 6 million calls were excluded in the regression analysis. Are those excluded similar to those included in the regression analysis?

-> Thank you for highlighting that we did not provide an examination of these excluded calls. We have now additional detail in the Methods section: “This reduced the sample to 16,563,946 triaged calls, from which we produced descriptive statistics. Our main regression model used a smaller sample of 10,954,783 calls, as this retained only rows with complete information on the outcome measure and all control variables. Comparing the former and the latter samples, we saw a slight increase in the mean value of the outcome variable (from 5.4% to 5.6%), the proportion of calls between midnight and 4 am (a 5.2% increase), the proportion of calls that received clinical input (a 6.1% increase), the proportion of calls that happened on a bank holiday and in the Christmas period (6.3% and 5.8%, respectively), and a decrease in the proportion of calls from patients based in London (a 12.5% decrease); all other relative changes were smaller than 3.0%. The regression sample is therefore very similar to the sample used for descriptive statistics, with a small number of notable deviations. Missing-data imputation was not performed due to computational infeasibility, given the size of the dataset.”

3) Among control variables, there were 96 different disposition codes. How were these codes used in the analysis? Did the authors regroup them or use in individual code?
-> We have now clarified that the disposition control variable used in the analysis did include 96 categories (additionally, text on page 6 notes that all dispositions had at least 30 observations):
“Our analysis used 18 control variables, shown in past research to be important predictors of ED attendance in England, which fell into five broad categories:
...
3 call characteristics: which of the 96 NHS Pathways disposition codes assigned to the patient was assigned at the end of the call.”

4) The description of data cleaning and control variables are quite thorough. However, there is little mention of how statistical analysis was done in the method section. For example, what statistical tests were used to compare various characteristics? How were logistic regression and OLS regression formulated and implemented? In what software the authors used to analyze them?
-> We have now added the following sentence:
“Our regression specification consisted of a linear combination of all the control variables and did not include any interactions or transformations." (p7)
Furthermore, we have clarified that “All analyses were performed in Stata 14 and R 3.5.0." (p4)

5) The outcome is avoidable ED use (yes/no). However, it seems the outcome should be three levels: avoidable ED, non-avoidable ED, and no ED. It is not appropriate to combine non-avoidable ED with no ED. The multinomial logistic regression may be better suited to model the probability of avoidable ED among those not advised to attend ED.
-> We thank the reviewer for this suggestion. We followed the binary coding approach for avoidable ED attendances given that this was the approach typically used in past research, and we wished to ensure comparability of our results.

6) The event of avoidable emergency department use rate was 5.4% for calls not advised to attend ED. Statistically, the rate is low, and OLS is unlikely appropriate. Some researchers accept OLS for binary outcomes, if the event rate is about 20-80%. Clearly, this study has a much lower rate. I would suggest using probit model to get predicted probability if the authors intended to do so. Or the authors can obtain marginal predicted probability based on logistic regression as well.

-> We thank the reviewer for this suggestion. We appreciate the potential issues with using OLS when the base rate of the outcome measure is relatively low, as is the case in our study. For this reason we included logistic regression (odds ratios) output in column 1 of Table 1.

We attempted to provide marginal predicted probabilities from the logistic regression. This turned out to be computationally infeasible, given the size of our dataset. Our attempt to perform this calculation on a 1% subset of the full dataset was prohibitively time-consuming on the sole computer we are permitted to use for analysis of this data (the computer in question is in a password-protected Secure Data Room, and its use is rationed among our research team). Furthermore, due to the sensitivity of the dataset, we do not have the option to perform the analysis on a cloud server. However, if the reviewer and/or editor feel that the addition of marginal predicted probabilities to our report are essential (in addition to our existing presentation of logistic odds ratios) in order to meet the publication threshold, please let us know and we will aim to reserve sufficient time on the appropriate computer in Spring 2020.

7) The authors used gradient boosting model to explore the explainable variations by each predictor (roughly speaking, relative importance of each predictor), and also use GBM to re-classify the population into three groups with high, medium, and low risk of seeking avoidable ED visit. This is appropriate, but the authors should describe in details how GBM was set up, and how the relative importance of predictors was assessed, and how these three groups were classified. Without sufficient details, it is difficult for readers understand the method behind the paper.
-> We thank the reviewer for the thoughtful comment. We have now expanded the section introducing the GBM analysis:
“We next tested whether we could improve our ability to predict which 111 calls would result in avoidable ED attendances by running a gradient boosted tree model (GBM). A GBM models the outcome measure as the result of a series of decision trees. Each tree attempts to identify areas where the others make poor predictions and correct for that, resulting in relatively strong predictive performance even in the presence of complex nonlinear relationships or interactions between the predictors and the outcome (a situation in which other techniques such as OLS may not perform so well). It is one of the best-performing predictive algorithms for tabular data. We built the GBM using a training subset of the data (a random selection of 80% of the 10,954,783 rows) and evaluated its out-of-sample predictions using a test subset (the remaining 20% of rows).

When we introduce the idea of variable importance, we have changed the explanation to hopefully be clearer how that concept is defined. We have replaced “... in terms of the overall predictive power (measured in terms of deviance, a generalised notion of residual sum of squares” with “The importance of a variable is defined as the improvement in log likelihood which is attributable to each decision in the decision trees which is made using the variable in question. These are then renormalised to sum to 100% to give the relative importance.”

8) Although this is not a methodological paper, the authors compared the area under the curve (AUC) between logistic regression and GBM, and found there was only a small improvement in predictability. It is not unexpected. Therefore, is it possible present the logistic regression results only?
-> If it is tolerable to the reviewer, we would prefer to retain the results of the GBM model in the paper. This is because:
(i) it was surprising to us that the GBM offered such limited incremental predictive power compared to the logistic output. This may also be surprising (and therefore informative) to other researchers, particularly given that machine-learning methods like GBM are occasionally assumed to always provide dramatically better predictions.

(ii) we consider the disaggregation of the explainable deviance in the outcome measure across the different independent variables to be a complementary addition to the inferential results provided in the regression table. Given that we want to present the GBM for reason (i) above, we think it is then logical to explore the predictive properties of the same GBM.

We are nonetheless appreciative of the reviewer's suggestion for a more concise results section.

Overall, this paper addressed an interesting and important questions about the effectiveness of 111 triage call, and potential health care cost saving and waste in the nation.
-> We thank Dr. Yu for her review and for this positive assessment of the paper.

VERSION 2 – REVIEW

REVIEWER	Janette Turner University of Sheffield UK
REVIEW RETURNED	21-Jan-2020

GENERAL COMMENTS	Thank you for your consideration of previous comments. These have all been addressed appropriately in this revision. Two minor points. There is much use of parentheses in the text. Some of these are appropriate but others - e.g page 3 para 1, page 11 para 2 are not as they enclose proper sentences that should be used instead (picky I know). I think "summary" would be better than "decomposition" which makes me think of graves in the title for table 2.

REVIEWER	Xinhua Yu University of Memphis
REVIEW RETURNED	22-Jan-2020

GENERAL COMMENTS	The authors addressed the reviewer's concerns.