Author’s response to reviews

Title: Predicting COPD one year mortality using prognostic predictors routinely measured in primary care

Authors:
Chloe Bloom (chloe.bloom06@imperial.ac.uk)
Federico Ricciardi (f.ricciardi@ucl.ac.uk)
Liam Smeeth (liam.smeeth@lshtm.ac.uk)
Paddy Stone (p.stone@ucl.ac.uk)
Jennifer Quint (j.quint@imperial.ac.uk)

Version: 1 Date: 28 Jan 2019

Author’s response to reviews:

Reviewer #1: Bloom and colleagues report on the development and testing of a prognostic index for one-year mortality in patients with COPD. They derive a 'BARC' index comprising blood biomarkers, age, respiratory parameters, and comorbidities, which performed well in predicting mortality (and superior to other variables derived for longer-term risk prediction).

The study represents an important step at the intersection between respiratory and palliative care. To date shared working and especially referral to palliative care is compromised in part by the lack of prognostic tools that perform sufficiently well to inform practice. The BARC index lacks the simplicity of some current mainstream respiratory indices, e.g. ADO, but here is derived from UK general practice databases, demonstrating 'real-world' utility.

My concerns are mainly minor, with the exception of the current presentation of data in Table 1, where I think there is some error - probably in the % values. As I read it, the sample (54990) has been divided by training and test set in columns 1 and 2, then died and not died in columns 3 and 4. Please check especially Pneumococcal vaccination (% columns 1 and 2), comorbidities asthma and hypertension (% columns 3 and 4), and palliative care received (% columns 3 and 4). I am hereafter assuming the errors do not alter the subsequent modelling.

Reply: Apologies, yes these were typos in the table at the time of making the table only, and have now been corrected.

I did wonder about the added value of the blood biomarkers given they were problematic with regards to missing data and were accepted within 18 months of the annual review. Were they collected around an exacerbation or symptomatic period for example? I appreciate this could be
inherent to the prediction of one-year mortality, but then why the long time scale? Some detail on the additional information they provided, possibility a sensitivity analysis, and discussion this aspect of the BARC index would be welcomed.

Reply: Thank you for your thought-provoking comment. We included blood tests in our predictive model as we included all potential prognostic indicators based on published literature. As blood tests did provide significant predictive value to the model, we felt it was important to include them; this may partially explain why the model performs better than other models. The blood tests included are common and non-specialist and we feel that it is likely that patients, where the GP is considering using such an index, will already have had a blood test in the past 1.5 years. However, if not, a simple blood test is easy to organise within primary care. In consideration of the length of time the blood test was taken before the annual review, only 16% did not have their blood test within one year of annual review, and 58% had the blood test within 6 months of their annual review. In consideration of the timing of the blood tests, only 2% of blood tests were taken within a 2 week period around an exacerbation (GP or hospital managed). We have now added these findings to our results (page 7, line 230-231) and extended the discussion (page 10, lines 313-317) on blood tests around these points.

Palliative care was I think operationalised differently here and in the group's previous ERJ paper (reference #7), so much so around 1% of the sample received it here compared to 20% before using the broader range of Read codes. I cannot see it specified in the prognostic predictors section in the method. Perhaps state in the introduction palliative care of any type.

Reply: The palliative care variable shown is the percentage that received palliative care in 1.5 years before the annual review date. This is the date at which each patient's risk of death in the next year was assessed at. In the actual year before death, 21% received palliative care. Thank you for your comment as we can see now that this variable measured in this way and reported in the first descriptive table is actually somewhat confusing and does not help with the findings so we have removed the variable.

Introduction, line 20: suggest reasonable level of physical function.

Reply: This has now been changed (line 75).

Method, Validation of the risk model, line 48: suggest time to death, as opposed to 'failure times'.

Reply: This has now been changed (line 192).

I do not see an ethics statement in the manuscript.

Reply: This is in the declaration section not the methods, as per the journal’s policy.
For the comparison of observed and predicted mortality, I would be interested to see absolute numbers to understand the number of false positive and false negative predictions at the individual level.

Reply: We are afraid we cannot give the figures you’re asking for. In fact, what can be predicted by our model is not the death status at 1 year, but the probability of death at one year. These probabilities are those given (using the mean probability of death for patients separately from each one of the four groups selected according to BARC quartiles) in the bar graph in the manuscript. Hence, false positive and false negative rates for prediction cannot be given, as they require a binary outcome (death/alive) that does not coincide with that of our model. There may be a way to circumvent this, by defining an arbitrary threshold for the BARC score and dichotomising the outcome according to this threshold, but we think this would weaken the rationale of using a cox model to predict one-year survival of COPD patients, which is a significant strength of our study.

For the reporting and in Figure 1, consider swapping the order so the prediction precedes the observation.

Reply: We have now changed accordingly.

Are the observation proportions rather than probabilities?

Reply: These are probabilities of death within the year after the index date.

Discussion, line 44: I do not fully understand the sentence concerning 20% of the cohort not having an annual review due to a shortened follow-up period. Is this study-follow up i.e. death or a clinical follow-up visit?

Reply: Patients that were eligible for the cohort, except that they did not have an annual review date in the specified time period (1/1/2010-31/6/2015), were 20%. This was mainly because those patients did not have enough research quality data, or were not seen in a CPRD-GP practice long enough, to have a year of eligible data and an annual review date. The patients were all followed up until death or censored. This sentence has now been clarified in the discussion (page 9, lines 301-303).

Figure 2 - BODEx not BODE in legend.

Reply: This has been corrected.

Supplementary Figure 2 - would benefit from a legend and the colour for the dataset third from the top displayed very light.
Reviewer #2: This study addresses a major unmet need in a large population, and is highly relevant to current clinical challenges and practice. Compared to inoperable lung cancer, patients with severe COPD have a greater symptom burden but much lower access to palliative care and advance care planning. As the authors highlight, this largely reflects the uncertainty of prognosis in COPD. Consequently improved prognostication should support selection for, and use of, these services. The manuscript is very clearly presented.

1) I strongly recommend excluding conditions that in isolation would commonly trigger consideration of palliative care and advance care planning from the primary analysis, notably lung cancer and pulmonary fibrosis.

Reply: Thank you for your point. We feel that it is not appropriate to exclude lung cancer from the main model as firstly many patients with lung cancer are not within 1-year of death therefore specialist knowledge on their cancer would be required, secondly many patients with lung cancer are not referred to palliative care until very advanced disease, thirdly other indexes do not exclude lung cancer (including CODEX and the Charlson Index – which in fact gives metastatic cancer the highest score). However, we do agree this should be investigated and as such we have carried out a sensitivity analysis to see if excluding patients with lung cancer altered its performance compared to other scores. BARC still performed well and showed a better performance than the other indexes (page 6, lines 202-203 and page 8, lines 261-262; and additional supplementary table 2).

The variable pulmonary fibrosis included all forms of lung fibrosis (e.g. sarcoidosis) – this is now clarified in the methods (page 4, lines 136-137), many of which do not have a life-limiting prognosis. However, regarding interstitial pulmonary fibrosis (IPF), which does typically have a poor prognosis, unfortunately palliative care is actually very poorly provided to these patients (equivalent or worse than for COPD), therefore, including them is essential.

2) The large cohort available is a particular strength. However randomised split into training and test sets is likely to favour the derived tool in the test set. The TRIPOD Statement, Ann Intern Med 2015, would not consider the current approach external validation. Non-random selection of a geographically and, if possible, temporally distinct cohort is a better approach; can this be achieved? This would allow robust external validation and unbiased comparison of the performance of BARC with other tools. The terminology (training and test sets) may be unfamiliar to most clinicians, but is clearly explained.

Reply: Thank you, we have taken on board your comment, it would take some time to redo the whole study again from the beginning and we think the current predictive model is good. Therefore instead we obtained an external dataset, some of which is temporally distinct (2004 to 2015), as well as a distinct patient phenotype (patients that did not have an annual review during the study period 2004 to 2015). We used one year after study eligibility to be the index date.
(instead of an annual review date). We then tested the validity, predictive probability and area under the curve of BARC in the external dataset. BARC still showed reasonable performance and compared favourably to the other risk scores. We have now added the methodology results and discussion of this external dataset to the manuscript (abstract: lines 29-31, 35, 40, 42-51; methods: page 4, lines 116-125, page 5, lines 179-181, line 188, page 7; results: page 7, lines 220-224; page 7, lines 244-248, page 8, line 256; discussion: page 9, line 268; additional table 1b and additions to table 3 and 4).

3) Tools consistently performed better within their derivation cohort than an external validation cohort. The latter is closer to expected performance in clinical practice and therefore should be the headline result reported (this can be alongside performance in the training/ derivation cohort). Table 4 test cohort AUROC BARC = 0.757; abstract BARC = 0.81.

Reply: Thank you, we agree. We have corrected this.

4) Other investigators have estimated the Charlson Index from CPRD data: if possible, include CODEX in the comparison of tool performance.

Reply: CODEX was derived from COPD patients after hospitalisation, therefore, some of the variables require more specialist knowledge (this is discussed in the discussion, page 9, lines 283-288) including AIDS diagnosis, severity of liver disease, severity of renal disease and hemiplegia. However, assuming any liver disease noted in the records was moderate to severe, any renal disease noted was moderate to severe, any diagnosis of HIV was AIDs, and hemiplegia only existed if noted, we obtained a CODEX score and compared to all the other indexes. Our score still compared favourably to CODEX. As we were not able to truly derive a score on all patients due to the lack of specialist information we have not included this in the manuscript.

First cohort: ADO= AUC:0.675, 95% CI(0.655-0.694)
First cohort: BODE= AUC:0.481, 95% CI(0.452-0.510)
First cohort: DOSE= AUC:0.588, 95% CI(0.564-0.611)
First cohort: CODEX= AUC:0.675, 95% CI(0.655-0.694)
First cohort: BARC= AUC:0.780, 95% CI(0.762-0.800)
External: ADO= AUC:0.568, 95% CI(0.541-0.595)
External: BODE= AUC:0.413, 95% CI(0.379-0.447)
External: DOSE= AUC:0.515, 95% CI(0.485-0.546)
External: CODEX= AUC:0.565, 95% CI(0.538-0.592)
External: BARC= AUC:0.691, 95% CI(0.667-0.715) 

(see table in pdf)

5) Generalisability:

The CPRD dataset covers 11 million, only 54,990 patients with COPD could be included and 21% died during follow up (median follow up period = 2.7 years). The mortality rate seems high for a general practice COPD cohort. What was the COPD-specific and all-cause mortality rate within this cohort (54,990) compared to the wider UK population diagnosed with COPD?

Reply: The all-cause crude mortality rate was 78.6 per 1000 person years and COPD-specific was 27.0 per 1000 person years. The crude rate, rates by age category, and rates by gender, are all in keeping with previous COPD mortality data (Morales et al, Respiratory Research, April 2018; and a recent mortality study carried out by our research group, Gayle et al, Thorax accepted Dec 2018). The similarity to other cohorts is now commented upon in the discussion section (page 9, line 298).

The tool includes FEV1 and MRCD; values for both are frequently not recorded on our local GP databases; missing data rates were <5% and 20% respectively in this study. Has data completeness in the CPRD been compared to other routine clinical GP systems?

Reply: We are not aware of published comparisons with other databases but because the training and test sets had to have an annual review this also explain the low amount of missing data. In the patients without an annual review there was 15% missing FEV1 and 50% missing MRC score. This is now commented upon in the discussion section (page 9, lines 307-310).

Exacerbations of COPD are not accurately coded, some require only antibiotics or only steroids, and self-management presents challenges. I appreciate an algorithm has been developed to identify exacerbations.

Reply: Thank you we agree and the algorithm looked at exactly that issue and found that including patients with both antibiotics and steroids on the same day, or either one (antibiotics or steroids) but also with two out of three symptoms (cough, dyspnoea or wheeze) recorded gave the highest PPV (86%) for exacerbations. This study is referenced in the paper.