Author’s response to reviews

Title: Can routine data be used to support cancer clinical trials? A historical baseline on which to build: retrospective linkage of data from the TACT (CRUK 01/001) breast cancer trial and the National Cancer Data Repository

Authors:
Lucy Kilburn (lucy.kilburn@icr.ac.uk)
Maria Aresu (m.aresu@imperial.ac.uk)
Jane Banerji (Jane.Banerji@icr.ac.uk)
Peter Barrett-Lee (peter.barrett-lee@wales.nhs.uk)
Paul Ellis (paul.ellis@gstt.nhs.uk)
Judith Bliss (judith.bliss@icr.ac.uk)

Version: 1 Date: 01 Aug 2017

Author’s response to reviews:

TRLS-D-16-00786: Can routine data be used to support cancer clinical trials? A historical baseline on which to build: retrospective linkage of data from the TACT (CRUK 01/001) breast cancer trial and the National Cancer Data Repository.

Response to reviewer comments

Reviewer #1:

With respect to the multivariable logistic model - more detail (i.e. candidate covariates, selection methods) would add to the clarity

Details regarding the multivariable logistic model have been added to the manuscript (p9 tracked version).

With respect to Overall Survival - Whilst there is very little difference in the two distributions presented in Figure 3 - there is some small evidence of divergence after 8 years. There might be some merit in performing landmark analysis (using landmarks of 5,6,7 & 8 years perhaps) to see if this divergence is more pronounced.

We would be cautious to put too much emphasis on the small divergence in the most recent data available (8-9 years post randomisation) in the Kaplan Meier curve for overall survival due to the smaller number of events/numbers at risk available at this time. Follow-up beyond 5 years for
both TACT and routine practice in general tend to be less frequent. This, combined with the lag in reporting and processing of routine data would make this part of the Kaplan Meier curve the most unreliable, therefore we do not think a landmark analysis would be appropriate for this comparison.

It is mentioned in the discussion that there was considerable amount of cleaning required of the data that was obtained - this raises a couple of questions. Firstly - if this approach was adopted in a future clinical trial, to what extent would the ‘cleaning’ have to be verified. Secondly, presumably there is considerable resources that go into the collection, cleaning and organisation of the data - is it possible to make any comment on the extent to which the resources incurred using this approach compare to standard data collection approaches?

We thank the reviewer for the comments stated above. If routine data was adopted in a future clinical trial in the place of centre-based follow-up we would recommend that data ‘cleaning’ would need to have an appropriate level of quality control checking attached to it as part of the trial’s standard data monitoring plan. This needs to be efficient and proportionate to the risk of the trial. For example, trials of an investigational medicinal product used in an unlicensed indication would require a higher amount of checking. Similarly, fields related directly to primary endpoint evaluation would require more validation. Data monitoring plans should be agreed by the trial team as early as possible. For routine data we recommend both “in house” verification of data as part of central statistical data monitoring whilst allowing the possibility of direct contact with centres for clarification of significant suspected systematic issues with data quality. We have added these recommendations to the discussion section of the manuscript (p15-16 tracked version).

In response to the second question, we agree that considerable resources go into the collection, cleaning and organisation of data. If routine datasets were to replace the gold standard of individualised data collection to facilitate long-term follow-up this should reduce the burden on research teams in hospitals, freeing up their time to concentrate on higher risk patients. In addition, patients should benefit by avoiding unnecessary follow-up visits. The resource saving may be less clear for clinical trials units. While switching to routine data use may reduce costs, the amount of time required to clean, process and merge routine data (with either in-house data collection or even datasets from other counties) may increase the workload substantially for the trials unit data managers and statisticians. Although implications of resource use was mentioned in the introduction and discussion section of the manuscript, we have added further detail based on the questions raised by the reviewer to add clarity (p16 tracked version).

Reviewer #2:

Registries and other sorts of 'routine' data collection routes may in general be rich sources of data, but there are inherent biases if one chooses to use this as a primary source of data collection. The authors should explain this issue in greater detail.

Missing data within datasets (in particular important details regarding relapse) and the time lag to receive data may contribute towards bias in using routine data as the primary source of data
collection. Of note the population we are investigating using routine data in is a pre-specified clinical trial population and we are not trying to extrapolate to the general population. This has been discussed in the manuscript (first paragraph p11).

The term 'characterises' in the background section of the abstract should be replaced with something less vague. It is unclear what this really means.

The term ‘explores’ has been used as an alternative, that is, “This project explores the potential for routine data to inform cancer clinical trials.” (p2 tracked version)

Several statements require more evidence/specific references to provide greater strength. For example, there should be a reference or greater evidence that there is a tendency to curtail follow-up after five years. Is this anecdotal or is there a specific reference or references that allow the authors to make this claim?

As part of the initial work we had conducted on challenges with long term follow-up in cancer trials we undertook a review of protocols and case report forms for all relevant trials listed in the NIHR breast cancer trials portfolio at the time (Kilburn et al. Trials 2014). This showed that, as the primary endpoint for most breast cancer clinical trials at the time was at five years after randomisation, most trials sought more frequent follow-up until this point. Thereafter, follow-up reduced in frequency, most often, to annual contact or, in approximately one third of trials reviewed, stopped altogether. We have added this reference to the statement as suggested by the reviewer. (p4 tracked version)

The authors should define 'routine' sources of data capture.

Routine sources of data capture are those regarding individual patients’ cancer diagnosis, treatment and outcomes collected by individual hospitals. These sources include data submitted by NHS providers such as patient administration systems, multidisciplinary team (MDT) software, pathology full-text reports and imaging systems; as well other data sources including cancer screening programs, Hospital Episode Statistics (HES) and Healthcare Quality Improvement Partnership (HQIP) commissioned national cancer audits.

These data sources have not been designed specifically for clinical trials use however may provide a potentially cost-effective alternative to hospital-based clinical trial follow-up. This has been added to the manuscript (p5 tracked version).

The authors should explain 'data cleaning, resolution of incomplete and incorrectly formatted information..." (page 6, line 41)...what was the process of data cleaning and how were incomplete and incorrectly formatted data 'resolved'? What are examples of inconsistencies? Were the decisions and ultimate final dataset(s) agreed upon in some objective way or via study team consensus?
We have revised the paragraph relating to the methodology in light of the reviewer’s comments to provide further details on how we arrived at one observation per patient and provide examples of the challenges faced when data cleaning (p7 tracked version).

I have a similar comment regarding decision of a suitable ‘proxy’ event to identify recurrences. What were candidate proxy events and how were they determined/agreed upon?

It is known that a patient will only have oncological intervention following their primary treatment if there is evidence of recurrence and therefore this ‘proxy’ was agreed upon. We have added this further detail in the manuscript (p8 tracked version).

There is reference to an exploratory multiple logistic regression model, but these results are not reported. Again, detail surrounding these methods are missing. Which software was used, what levels of significance were used, etc. Further, it would seems that rather than simply presenting % agreement, more descriptive measures such as Kappa statistics (with confidence limits) and correlation coefficients/Bland-Altman results, etc. would be more informative.

Further to the answer given for Reviewer 1, further detail has been added with regard to the logistic regression model and a table of the model’s results has been added to accompany the text already in the results section (p11 tracked version).

With respect to the descriptive measures suggested by the reviewers for assessing agreement, using statistics such as kappa in this large dataset could suggest high levels of agreement whilst still having a discordance that would be significantly large to be of concern within a clinical trial. In addition correlation coefficients would not be appropriate here as we could have high correlation but low agreement for example if tumour size was measured in cm in one dataset, but mm in another. Our primary focus was on clinically meaningful levels of discordance and missing data thus we do not think these statistics would be useful in this context.

It should be clearer as to what the authors mean by 'one row per patient' - One would assume this means duplicates were removed. How did this process occur? What makes one record the 'correct' one?

Further to the answer provided above, we have expanded the methodology section in the manuscript to discuss the process of removing duplicates and how we arrived at the decision of which observation to keep (p7 tracked version).

The discussion makes reference to 'good' quality in the first sentence. This term seems subjective and/or a clear definition of 'good' should be added.

‘Good’ has been replaced with ‘reliable’ quality i.e. of sufficient quality to be considered for clinical trial follow-up purposes (p11 tracked version).

The conclusion makes reference to 'the goal' - whose goal is this? The authors should explain and elaborate on these points. In addition, the last sentence regarding 'working with NCRAS...to
validate routine data...' should lay out more concrete steps using less vague terminology. How do the authors plan to 'work with NCRAS' and what is the method of validation?

The overall aim for trialists, hospitals and patients is for clinical trials to run more efficiently with a reduced resource burden. The NCRAS prospective validation study has now been explained in more detail in the discussion and briefly re-discussed in the conclusion (p13 tracked version).

The authors should consider discussion of strengths and limitations of this project along with wider applicability to other fields or datasets.

The strengths and limitations of this study have been considered. The multicentre aspect of the study gave a realistic impression of the quality and variability of the available data. However, the newer datasets were not available for use at the time of this study and therefore investigation of the newer datasets will be required. These points have been included in the manuscript (p13 tracked version).