

Ensuring the quality of your data in Respiratory trials: Data management from a statistical standpoint

1. Abigail Fuller, Statistician, Veramed Limited
2. Nick Cowans, Statistician, Veramed Limited



1. Introduction

Large, global late phase studies inevitably involve huge amounts of data of varying quality. Data frequently needs cleaning up prior to locking the database, a responsibility typically lying with data management. The ability to look at multiple extracts of data while the study is ongoing and blinded has enabled us to develop novel methods and processes for increasing the confidence in data quality.

Forced expiratory volume in one second (FEV₁) is the volume of air expelled from the lungs in one second, this is measured in millilitres or litres depending on the equipment. Outliers in FEV₁ can commonly be caused by a decrease in subject effort, illness or equipment failure. These values would not be considered valid and including many subjects with these values in the analysis can cause variations that may not represent the true treatment differences. Looking at these in a visual way emphasises the importance of ensuring that these outliers are genuine data points.

A respiratory exacerbation is an event, such as pneumonia or a COPD exacerbation that effects the airways and hence the patients ability to breathe. Respiratory tract exacerbations may present over a period of a few days with symptoms progressing. Due to this, sometimes these events can be recorded as two separate exacerbations when they are actually the same event progressing over time. When rates of respiratory tract exacerbations are an endpoint, recording duplicate or overlapping events will alter results.

The need for identifying these two different situations is apparent. Previously, clinicians would spend time looking through vast amounts of data, however, these Patient Profile review tools have made clinical review a quick and easy process stressing the importance of these data on our endpoints.

2. Measures of Interest

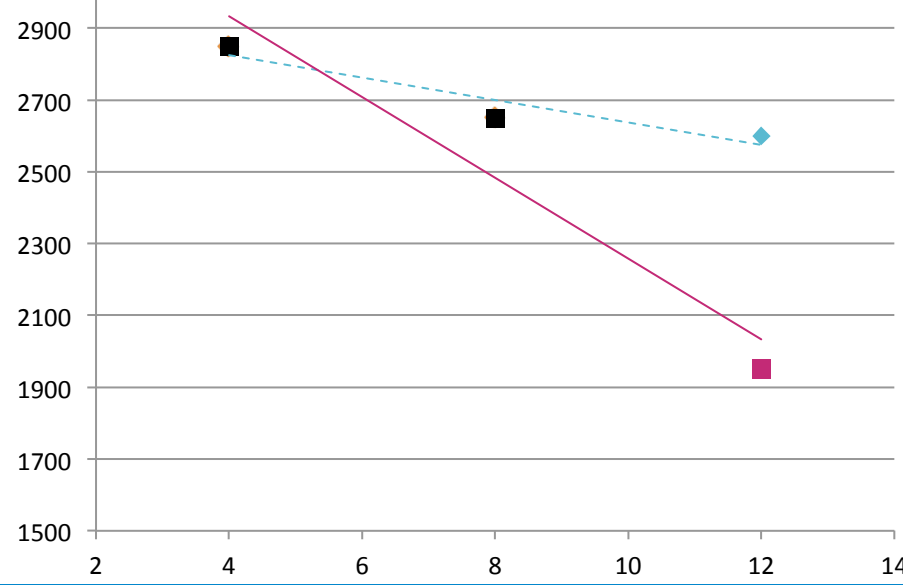
FEV₁ Rate of Decline

Interest

Rate in decline FEV₁ is an endpoint commonly used in respiratory trials. Rate of decline is measured in millilitres per year and a raw rate of decline can be calculated simply using linear regression.

When calculating the slope extreme values of FEV₁ caused by respiratory events or equipment malfunctions can lead to differences in the rate of decline that don't represent the true rate of decline of the subject. This is illustrated here in this example plot of FEV₁ (mL) against Time (weeks). As illustrated in the figure, changing the final data point to be an outlier (red) changes the raw rate of decline quite substantially.

Subjects with outliers can be found easily, but the influence on the results is more difficult to explain without a visual/graphical option.



Rate of exacerbations

The rate of exacerbation events can be reduced by taking certain medications and hence this rate is also often used as an endpoint.

This can be easily calculated as the number of exacerbations divided by the time of treatment. For example, if one subject has one exacerbation in 6 months, the rate would be 2 exacerbations a year. This means that if an exacerbation event is recorded as two events instead of one, the subjects exacerbation rate per year can be as much as doubled. If this happens with a lot of subjects this can influence the results.

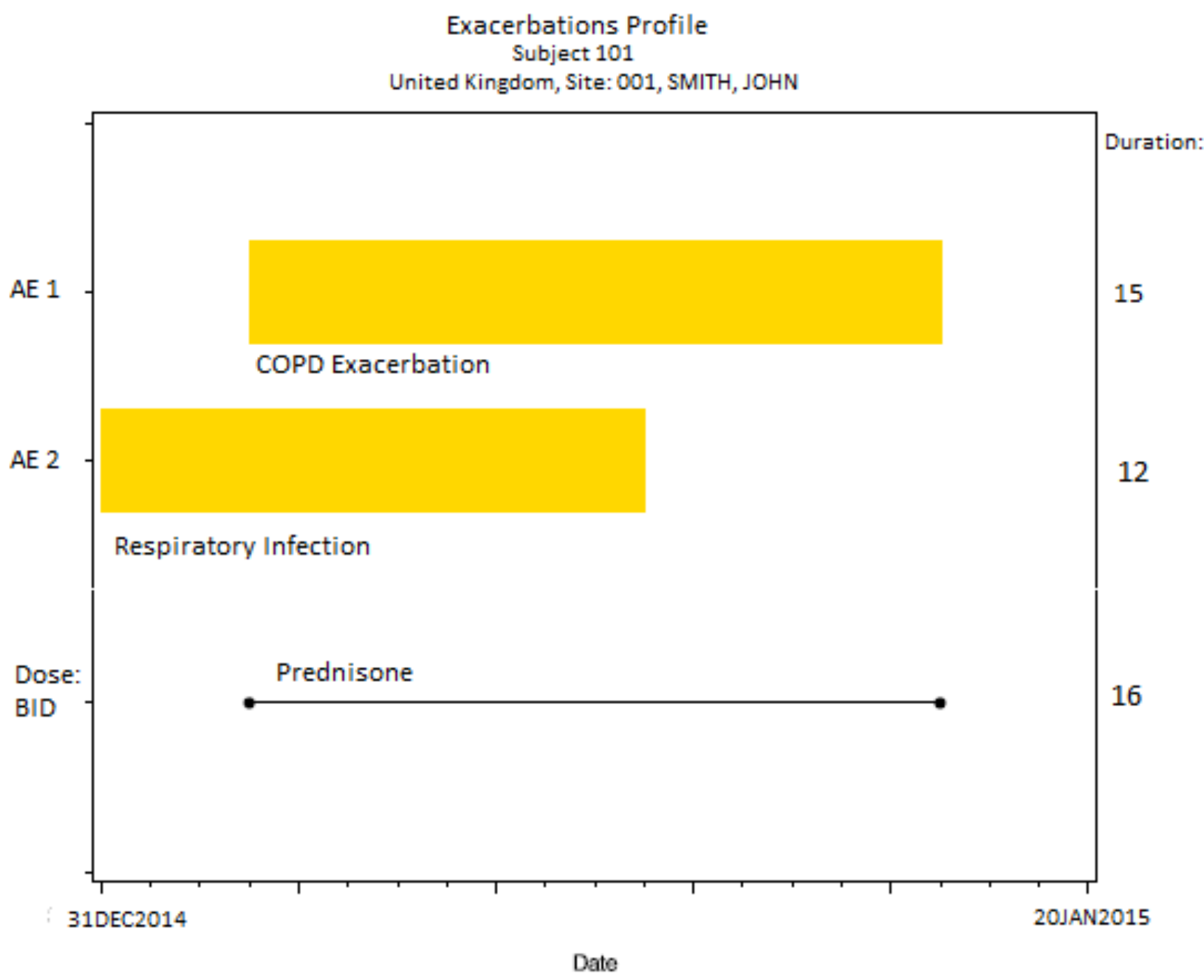
This data would traditionally be reviewed at a subject level, subjects with multiple respiratory events could be programmatically identified but data would be reviewed on a case by case basis, an inefficient and time consuming process.

3. Rate of Exacerbations

For Rate of Exacerbations, the data was presented in a patient profile bar plot, with one plot per overlapping event. Information on medications taken at the time of the respiratory exacerbation was included as well as duration of the event in days.

Presenting this information in a visual way, it is apparent that the 2nd Adverse event is not a second event but actually the same exacerbation event which probably started on 31st December. This event would increase the rate of exacerbations for this subject. In this example, the case is easy to deal with, however, when events are separated by a few days or even a week, the case is not so simple, and having the medications taken by the subject is more important.

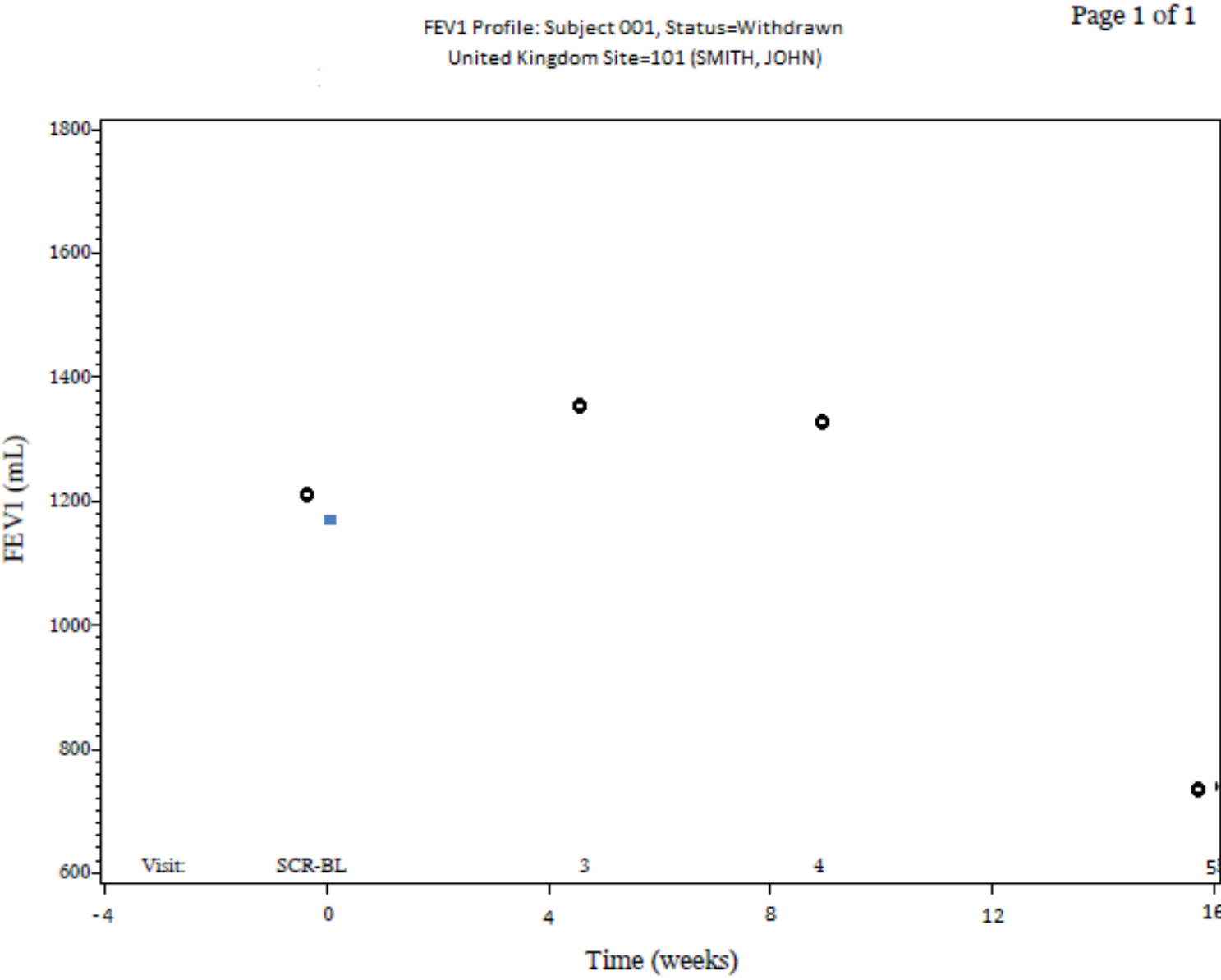
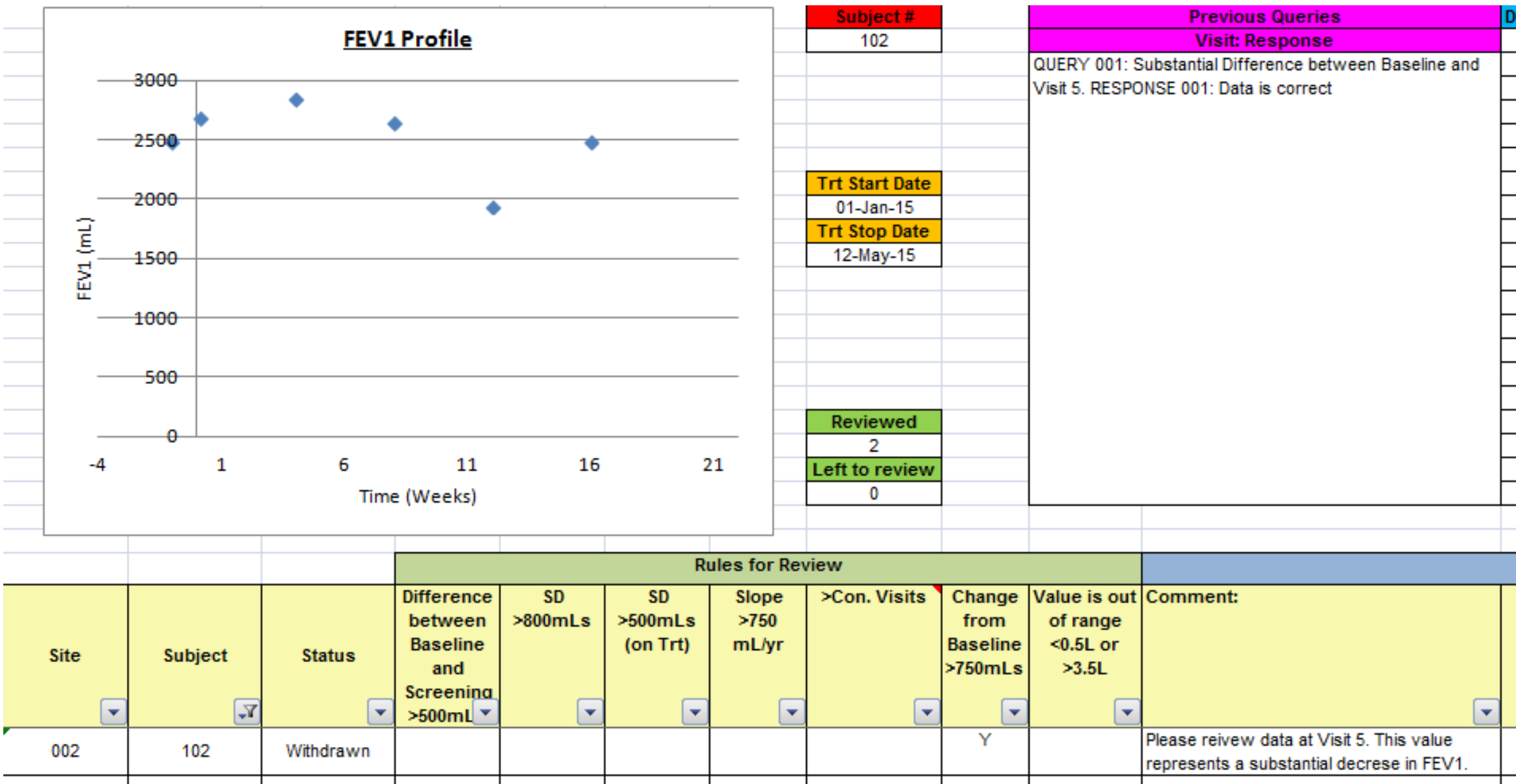
These profiles were reviewed by the clinical team and further action was taken of querying the data or sending the exacerbation profile to the site and asking them to clarify the data. When the site can visually look at the profile, it is immediately clear if they have made any misrepresentation of what happened in the recorded data, and if not they can explain the differences for the clinical team to review.



4. FEV₁ Rate of Decline

The need to have a visual tool to ease the clinical review was identified, with an emphasis on user ease and content needed to reach a clinical decision on further action.

Using a combination of excel and SAS, the following tool was developed. Some clinical review rules for FEV₁ rate of decline were identified including large differences between screening and baseline values and calculated individual rate of declines greater than 750mL/year. Utilising these rules, a review spreadsheet was created identifying the subjects programmatically in SAS. Filtering on subject number gives the individual subject information on one page, including treatment information, previous queries, responses and any previous review information.



NOTE: Blue square represents baseline value

Comments:
Please review data at last visit. Reduction in FEV1 is not consistent with normal rate of decline for subject.

PI Comments & Signature:

Upon reviewing the spreadsheet, decisions can be made as to whether to take any further information, if a query has been answered previously with an appropriate clinical reason for the discrepancies in FEV₁ between visits then the decision may be made to not take any further action with the data, as it represents the true values. Previously, upon reviewing the data, query history for subjects would not have been available to view as easily.

Once all subjects have been reviewed and decisions whether to issue a profile made the spreadsheet can be read into SAS and a PDF version of the patient profile is created and issued to the site for review. This document contains the clinical review comment, the graphical representation of the FEV₁ data and space for the site investigator to comment to explain any large variations. Placing all of the information into one document makes the anomalies in the FEV₁ data visually clear to the investigator at the site, whilst providing anymore information they need to investigate.

The PDF profiles can be issued to sites for their review. If they see any immediate data entry issues with the FEV₁ data, it is assumed these will be corrected and the subject will no longer be picked up by the clinical rules set initially. Alternatively, the Principal investigator can return and comment on the discrepancies in the box supplied. These comments can also be entered into an electronic data capture system to track. The comments from the site should explain the discrepancies seen in the data and can be reviewed by a clinician to make sure they make sense medically.

5. Summary

- Reviewing study data in the traditional way can be a time consuming process.
- Data can be difficult to review when displayed in typical dataset standards such as listings. Having unique tools can help our clinical colleagues keep track of their review more easily.
- Visual tools can assist with understanding of impact on endpoints for both clinical reviewers and investigators at sites.
- Seeing an outlying on an FEV₁ plot, or seeing clinical events overlapping in front of you in a diagram highlight the effect these would have on analysis .
- Using more advanced programming approaches to identify and display individual subjects can have benefits for all departments, data management, clinical and statistics and programming.
- With increased communication between departments and providing these visual tools created from a statistical standpoint to data management, we can help increase data quality and hence contribution towards endpoints.

