Three Respiratory Trial Strategies that Won't Leave You Breathless

PEER REVIEWED

The biopharmaceutical industry invests billions of dollars annually into clinical research to address a range of debilitating respiratory conditions such as asthma, chronic obstructive pulmonary disease (COPD), cystic fibrosis and idiopathic pulmonary fibrosis, as well as non-respiratory indications with lung safety and disease progression endpoints. Clinical trials that collect respiratory endpoints are among the most expensive to conduct.[1] In such a competitive environment where recruiting patients is increasingly difficult and costly, data from all patients need to actively contribute to the overall success of the study and development programme.

Lung function data are regarded by regulators as a key biomarker for clinical efficacy. However, measurements are inherently variable, and this variability drives the need for greater patient numbers to demonstrate a treatment effect which increases the time and cost of development. This variability also reduces the ability to determine responder patients and can actively reduce the magnitude of observed treatment effect in some situations. There are many examples of excessive variability undermining study outcomes, which result in repeated pivotal programmes. At an earlier stage, high variability increases the chance of a type I or type II error and can impact go/no-go development decisions. The fact that routine clinical care does not require high-quality data, and that patients often find it easier to generate suboptimal data contributes to this problem.

Attempts to reduce variability culminated in the American Thoracic Society (ATS) and European Respiratory Society (ERS) guidance published by Miller in 2005.[2] Despite this guidance, a post-hoc analysis of specialist respiratory tertiary care sites in Belgium revealed that 57% of data failed to meet core ATS/ERS standards and that the majority of respiratory technicians used the highest FEV1 data suggested by the spirometer, without critical appraisal.[3]

This level improves when equipment is standardised and sites are provided basic training at the study start. However, in a recent study, where approximately 3.3 million lung function assessments were analysed, approximately 11% of the tests submitted by investigators still had errors. When inappropriate test data were deselected and more appropriate test results were selected, the average change in FEV1 was approximately 112 mL[4] (Figure 1).

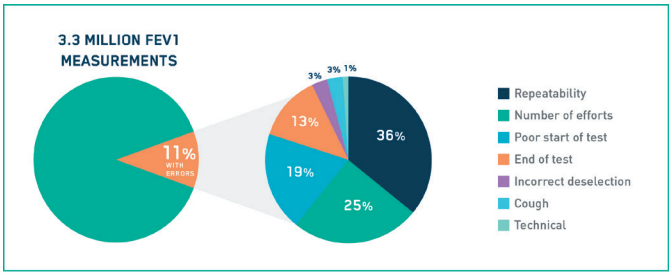

FIGURE 1. RETROSPECTIVE ANALYSIS OF MULTIPLE RESPIRATORY STUDIES HIGHLIGHTS MAGNITUDE AND CAUSE OF INVESTIGATIVE SITE MEASUREMENT ERRORS

Sources and Results of Variability

Variability in lung function testing can originate from many areas. The improper performance of spirometry tests or the inclusion of patients who are unable to master the technique represents the largest source of variability, which can exceed 50% of the test value.

This patient-based variability can preferentially occur at the start of any clinical research, as the patient is not sufficiently experienced at generating spirometry data. Suboptimal efforts can also (counterintuitively) inflate, rather than underestimate, lung function values. Improvements in patient technique due to practice can then actively reduce the lung function values, which counters the potential benefits from treatment and thus underestimates true treatment effect. Variability at any time point will increase the standard deviation of the treatment effect, which reduces statistical power and confounds the determination of a true mean treatment effect.

A good example of this technique-based variation can be found in COPD patients. Within the young healthy lung, as forced expiration occurs, inter-attachments within the lung support the airways and hold these open, allowing the air to escape quickly. As the lung ages, there is a natural destruction of lung parenchyma, which impacts these attachments. This decline is accelerated with advancing pathology and increased levels of inflammation driven by infection or underlying abnormal immune responses as seen in asthma or COPD. In this altered state, the pressures placed on the lung compress the lung and start to collapse the airways. This results in a pressure-dependent collapse of the airways which restricts the flow of exhaled air. In this instance, a good effort will massively restrict the amount of forced exhaled air in one second (FEV1). A failure to exhale forcefully will reduce the level of pressure-dependent airway collapse and actually increase the FEV1. As patients do not like the experience of collapsing their airways and restricting their airflow, their technique is often poor at the start of a trial. Gradual improvement in technique will drive loss of FEV1.

This process does not generate just acceptable or unacceptable data at two extremes but a continuum of effort-dependent collapse. It is possible for patients to generate repeatable but still suboptimal data and as such, even when the exhalation meets minimum standards established by the ATS/ERS criteria, it is possible to inflate FEV1 data twice the level seen within a normal treatment effect.

A core technique to counter this issue is for sites to focus on the flow loop geometry for each patient to ensure that this is consistent between visits. To address variability within the data, it is necessary to adhere to few basic steps:

1) Assure Quality Before First Patient In

Standardise devices. Standardising equipment across investigative sites minimises the variability originating from different equipment models. Sites use a variety of equipment for clinical use, resulting in an inability to determine the level of adherence to ATS/ERS minimum standards, sensor accuracy and calibration. Without standardisation, sites can adopt varying approaches to configuring

devices and recording patient data; variability in test results will increase. Increased numbers of outliers will undermine trust in the data during regulatory review.

Apply best practices. The most effective clinical trials follow ATS/ERS standards as a baseline and then apply additional best practices to reduce other areas of variability that can influence data quality and reliability. These practices can include software workflow reinforcement for study protocols, checks for inclusion/exclusion criteria, alerts to highlight out-of-range values or data acceptability, and even reminders to check if patients have adhered to restrictions on highly-caffeinated drinks or tobacco prior to testing.

Consistently train. The core objective for site training is the generation of higher-quality research-grade data. While it is possible to provide overread feedback based on deviation, this only allows the ability to select the best efforts based on data that has already been generated. New, more optimal data cannot be generated once the patient has left the site. Also, it is not possible to remove suboptimal data from the primary analysis in most studies. It is critical that sites understand what drives variability and actively manage the patient to ensure that the primary data generated is optimal for each patient visit.

Effective education and training needs to move beyond hardware and software operation, so that sites understand how variability in data quality occurs and what the impact of this will be on the study outcome. Without this, sites are more likely to return to the standards followed in normal clinical practice.

While training may add incremental time and expense to trial start-up, the benefits are potentially significant – a greater percentage of acceptable data can be collected, resulting in an array of downstream benefits.

Require device proficiency testing. A proficiency verification step serves to assess the site staff's core understanding of the equipment and data transfer process. Software and hardware should restrict access to those staff who have met these basic requirements. Because of patient consent issues, the certification process is normally performed on fellow research staff rather than patients with underlying pathology; therefore, certification does not ultimately test the suitability of staff to adequately test research patients. As a result, early critical assessment of the first patients analysed provides the best opportunity to assess competency and address remaining training issues. It is important to link each test procedure to a specified technician to drive optimal quality. Password control of spirometry systems does not prevent different technicians from progressing with testing once the initial sign-in process is complete. A fingerprint sensor with verification prior to each test is a quick and efficient way to allocate each test to a specific technician.

Select sites based on past performance. Future site performance related to data quality can largely be predicted from past performance. Core data which look at the percentage of various QC grades in recent studies give a good indication around the overall level of proficiency of a specific site. More enhanced algorithms can be generated to use a combination of technical quality indicators along with behavioural and performance data that score a site's historical performance. This can give an indication around both the quality of data and the core abilities of the site to meet recruitment and operational responsiveness targets. This evidence-based risk approach allows the implementation of mitigation strategies and the selection of sites that are most suited to specific study requirements.

2) Optimise Data Collection During the Study

Enforce minimum acceptability standards. Minimum ATS/ERS standards regulate equipment performance criteria and measurement procedures (e.g., environmental restrictions, patient position, instruction and coaching) to ensure reliability and consistency. Following these standards ensures that a minimum quality standard is reached. Moving beyond minimum data standards is possible if sites ensure that data is optimal for each patient. A core concept of this is the focus on flow loop geometry. The flow loop geometry represents the underlying lung pathophysiology and acts in some respects like a fingerprint for each patient. Flow loop geometry should not change between visits in most circumstances. Significant changes in flow loop geometry are normally driven by poor technique, which act as a marker for excessive variability.

Sites should review the flow loop geometry prior to patient testing so that they start the session with an understanding of what the patient is capable of achieving. It is only then that technique-based variability will be captured whilst a patient is still at site and there is time to influence the quality of data.

In addition, any unexplained jumps in lung function values should be explored with an open question to the patient around any changes in their normal routine since the last assessment. This can often uncover issues which the patient has forgotten to mention, which allows for rescheduling of assessments, if appropriate.

Conduct calibration checks. Many forms of equipment require calibration at the start of each test session to adjust measurements to ambient pressure humidity and temperature and to improve the measurement accuracy of the spirometers.

The introduction of pre-calibrated sensors on spirometry devices can reduce site burden and remove potential variability based on poor calibration. These devices also offer the benefit of automated sensors that adjust for temperature, pressure and humidity conditions, further reducing variability and site burden.

Perform ongoing best test review overread. Review by qualified and intra-/inter-reader variability tested respiratory specialists identifies noncompliant values to the ATS/ERS guidance and deselects them. This allows the selection of the highest technically acceptable effort, reducing the level of unacceptable data to 1-2%. Errors contribute to overall variability, reduce study power to show a mean drug effect and can potentially change the responder status of individual patients.

3) Centralise Data Analysis and Risk-Based Monitoring

Perform centralised data quality review. Checking and correcting data against minimum ATS/ERS standards will drive improvements in variability and enable greater determination of the drug effect; however, levels of variability often remain that are not initially detected by focusing on each assessment in isolation. Review of changes in lung function data over time helps to validate the appropriateness of individual data points by comparing these with repeat measures. This process allows the assessor to consider what optimal data quality looks like for a patient and to discover transitions in lung function data that are biologically implausible. Review of outliers based on the transition from baseline parameters at each visit helps to identify potential errors with either the baseline data or subsequent test sessions. The magnitude of the outlier determination needs to focus on the likely drug effect being tested and the specific indication under assessment. Once outlier limits are set, any change in lung function above this level should

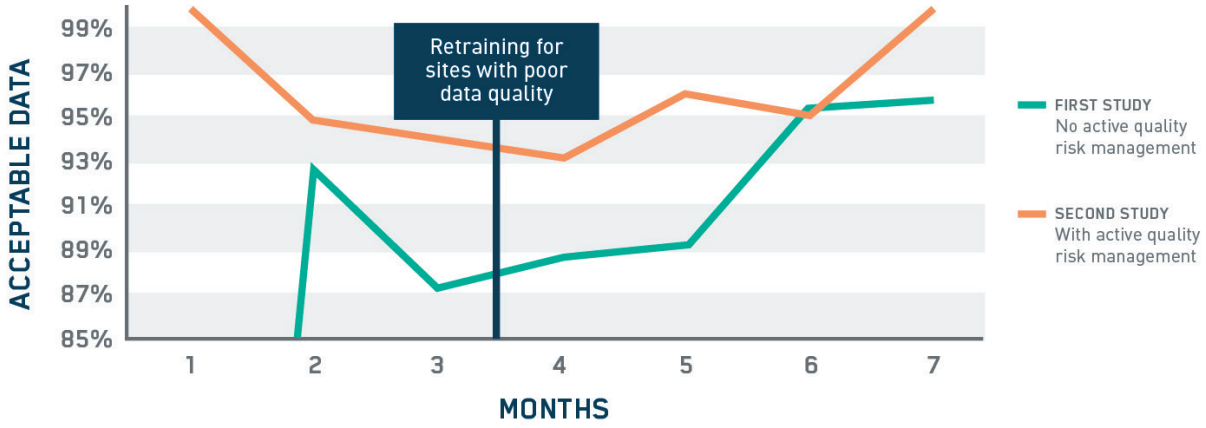

Figure 2: Comparison of Monthly Respiratory Data Quality in Two Studies

prompt a review of all test sessions regarding the transition in numerical parameters and flow loop geometry.

This process identifies early detection of specific site issues and allows for early remedial action.

Engage in quality risk management and risk-based monitoring. With new regulatory guidance, sponsors and CROs are increasingly engaging in quality risk management and risk-based monitoring (RBM) strategies for maximum programme quality and efficiencies.[5] Studies show that RBM can reduce source data verification and monitoring costs, which account for an estimated 25–50% of overall trial costs.[6,7]

An effective risk management programme includes an endpoint-specific monitoring plan for the investigative site that assesses the initial protocol risks, needs and metrics and then analyses the endpoint data in near real time to identify potential data quality risks. Such a programme enables sponsors and CROs to proactively focus their resources on those sites that need support or retraining to enhance data quality.

These centralised study data can also be integrated into electronic data capture (EDC) and other eClinical systems for enhanced trial oversight. Part of this approach requires the upfront determination of risk for each site and a variable approach to identify risk trends to allow intensification of oversight. Vendors with detailed site feasibility data allow historical quality metrics to be assessed for individual sites to help focus the initial risk assessment. Certification of site staff and bespoke training at study start permit some level of baseline risk determination. Central reporting of current and historical quality helps to identify outliers and determine the current risk of endpoint data generation. More focused training and oversight can be implemented and current or recent quality metrics can be compared with historical performance since trial onset.

As an example, in Figure 2, two different respiratory studies are compared, one with active quality risk management to identify sites with poor data quality, and one with no active risk management. Data quality improved by 4% with enhanced oversight and retraining of targeted sites.

Move from acceptable to optimal data

Since the introduction of the combined guidance in 2005, the standardisation of spirometry systems and use of centralised overread services have significantly improved lung function quality and reduced variability. However, variability still exists with the minimum standards in the ATS/ERS criteria and can reduce the ability to identify patients who respond to therapy. A move to optimal research-grade data can increase the determination of the true drug effect; however, focused training and proactive management are required to achieve optimal results.

Conclusion

Active management of data quality should reduce variability in the outcomes, producing greater confidence in the drug effect, greater ability to identify the true responders, faster drug development, enhanced potential for approval and reimbursement and potentially greater first-line drug use. Adhering to these three strategies will help sponsors cut the time and cost of their trials while improving data quality – and that may just give them the edge they need to gain a competitive advantage in this market.

REFERENCES

1. https://aspe.hhs.gov/report/examination-clinical-trial-costs-and-barriers-drug-development
2. https://www.thoracic.org/statements/resources/pfet/PFT2.pdf
3. Dieriks, B., Use of spirometry software in two major Belgium hospitals, European respiratory journal Vol28 S50 2006; 202S, P1215
4. https://www.ert.com/standards-risk-based-monitoring-executing-successful-spirometry-trial/
5. http://www.appliedclinicaltrialsonline.com/preparing-ich-e6-r2-addendum-3-part-series-0
6. http://www.ncbi.nlm.nih.gov/pubmed/15864237
7. Redfearn S. Risk-Based monitoring slow to catch on with industry. CenterWatch Monthly, May 2012, Vol 19(5).

Phil Lake, PhD

Respiratory Solutions Architect, ERT

Phil Lake has held a variety of roles in drug development for nearly 20 years, predominantly focused on respiratory trials. Prior to ERT, Phil held positions at SmithKline Beecham and GlaxoSmithKline where he worked on a number of anti-inflammatory agents, dual and triple combination therapies, monoclonal antibody studies, anti-infectives and some of the largest mechanistic studies looking at biomarkers within sputum and biopsy samples. For the last 10 years, he has concentrated extensively on rare respiratory diseases, including Cystic Fibrosis and Idiopathic Pulmonary Fibrosis, covering a variety of drug mechanisms and medical devices.