SWAT 167: Risk-based trial monitoring: Site performance metrics across time

Objective of this SWAT
To investigate how trial management events (e.g. site visits, Data and Safety Monitoring Committee (DSMC) meetings, and interim analyses) impact on site performance.
To improve our knowledge of how site performance metrics behave across time.

Study area: Monitoring
Sample type: Metric data calculated from data collected from sites (central monitoring data), ,
Estimated funding level needed: Low

Background
Trialists need to promote patient safety and well-being and ensure data and trial integrity in their clinical trials. Often, the intensity of a central and site monitoring strategy is chosen based on the level of risk inherent in the trial design or population. Many monitoring strategies involve selecting metrics (e.g. percentage of expected CRFs that are returned) and assessing site performance based on these. If the metrics are poor (e.g. there is a threshold violation for several metrics, or for one important metric, or one extensive violation) for a site, then an escalation action occurs, such as an on-site monitoring visit. This SWAT aims to analyse data from a range of Phase III/ IV trials to investigate changes in such metrics over time, and how these relate to trial management events (including escalation events). With the ICH E6 (R2) 2016 guidelines advocating risk-based monitoring, more trials are now using site performance metrics to assess if an escalation activity is required. Trialists need to find out more about how site performance metrics normally behave. These data will add to our knowledge and contribute to discussions of which metrics to use.

Metrics are numeric measurements, mostly obtained and calculated from data held in the database of the trial, and used to evaluate sites' risk or performance. Thresholds are numeric values that metrics are compared against to identify potential problems at a site. They should be initially set as part of the monitoring plan to reflect potential risk, and then reassessed periodically throughout the trial. Triggers are indicators of whether a metric has violated a set threshold and can be used to assess potential or actual risks and identify site under-performance that requires action. The total trigger score is the total number of metrics whose thresholds were violated (triggers that 'fired') at a particular site and timepoint.

Relevant actions might be:
• trial management events, including routine Data and Safety Monitoring Committee (DSMC) meetings or interim analyses.
• escalation actions, including on-site monitoring visits, and site training or the dissemination of new guidelines based on monitoring findings.

In this SWAT, a current list of metrics and corresponding thresholds will be obtained for each host trial. These metrics will have a threshold level agreed for action, so that a threshold can be deemed to have been violated or not (Yes (1) meaning that action should be considered, No (0) meaning that no action is indicated). Metrics and their thresholds may be simple, such as:
- Have there been any protocol deviations since the last site visit? Yes / No, or complex, for example containing a time element.
- Is the data return rate below 80% and was it below 80% at the previous report? Yes / No.

Until relatively recently, there were no available lists of metrics and trial teams would usually choose their own, risk-based, metrics at trial set up. While suggestions for metrics have now been published, there is still a dearth of evidence for their utility and, therefore, the trials in this SWAT will choose their own metrics and thresholds to reflect their unique assessment of trial risks and we also encourage them to consider published metrics to allow us to test their utility, such as the eight site performance metrics proposed by Whitham et al.[1]

Interventions and comparators
Intervention 1: N/A

Index Type:

Method for allocating to intervention or comparator

Outcome measures
Primary: Outcome measures for relevant action
a) Change in total trigger score after a relevant action
b) Change in total trigger score one year after relevant action for those in which there was an improvement post relevant action
c) Difference in behaviour of different metrics (e.g. are some metrics more sensitive?)
d) Interaction of metrics (e.g. where different metrics may give the same monitoring answers)
Secondary:

Analysis plans
The analysis comprises two parts:
Total trigger score across time:
• Graphs of the total trigger score (vertical axis) across time (horizontal axis) per trial site with vertical lines on each graph to show wherever a relevant action occurred, with all graphs for a trial's sites on one page where possible. Such actions might include site visits or other key events (such as a DSMC, new training initiative or provision of a guidance document to sites).
• Calculate how many sites had a lower total trigger score at the next report date after a specified relevant action (if a site has more than one of the specified relevant actions, one would be randomly selected).
• Calculate how many sites had the same or lower score from post relevant action to one year later (or a suitable time interval for the trial and action).
Individual triggers across time:
• Grids of individual triggered metrics (rows) across time (columns) for each trial site. Show individual metrics as a shaded/ coloured square when they violate a threshold, white where they do not, and textured/grey where the metric has not been measured. Each column represents the regular date when a report on the triggered metrics was produced. All grids for all sites on one page where possible, with vertical lines on each grid (placed between the relevant columns) to show whenever a relevant action occurred.
• Calculate how many metrics remain 'triggered' throughout the trial (or relevant trial phase e.g. a metric about recruitment might only apply during the recruitment phase).
• Calculate how many metrics are never 'triggered' during the trial.
This entire analysis plan can be completed for the chosen trial metrics and any published metrics, where data are available.

Possible problems in implementing this SWAT
We suspect that many trials have not stored this type of central monitoring data but we hope that trialists and trial units interested in this SWAT might consider starting to store this data for future analysis.

References
1. Whitham D, Turzanski J, Bradshaw L, Clarke M, Culliford L, Duley L, et al. Development of a standardised set of metrics for monitoring site performance in multicentre randomised trials: a Delphi study. Trials; 2018;19:557.

Publications or presentations of this SWAT design
None as yet (a manuscript has been submitted for publication)

Examples of the implementation of this SWAT
Yorke-Edwards VE, Diaz-Montana C, Mavridou K, Lensen S, Sydes MR, Love SB. Risk-based trial monitoring: Site performance metrics across time. Trials 2019;20(Suppl.1):P-33.

People to show as the source of this idea: Victoria Yorke-Edwards, Matthew R. Sydes and Sharon B. Love
Contact email address: v.yorke-edwards@ucl.ac.uk
Date of idea: 29/OCT/2020

Revisions made by:
Date of revisions: