All You Need To Know About SMART™ – The Engine That Drives CluePoints

Marc Buyse, ScD IDDI (International Drug Development Institute) and CluePoints, Inc.

1. Central Statistical Monitoring

This section provides an overview of the concept of Central Statistical Monitoring and the methodology used within the SMART™ engine that lies at the heart of the CluePoints solution.

1.1. The concept underlying CSM

To effectively target on-site monitoring visits, sponsors must look at their data in numerous ways. By doing so, it is anticipated that 100% SDV levels will be vastly reduced and studies will be significantly de-risked by improving data quality. In comparison to KRIs, which are highly subjective and only indicate the potential for risk, CSM is based on the actual clinical data. All variables are deemed to be indicative of quality, from lab data, baseline and clinical data, to treatment and patient-reported outcomes; everything is assessed and all data counted as equally important. In a clinical study, everything collected should be worth collecting, and therefore worth checking. CSM ensures the ability to determine data quality and ensure that monitoring efforts focus on sites with statistically atypical data. CSM offers the potential to determine where issues might lie in clinical data during study conduct and before significant problems occur, consequently, helping to avoid any shocks and surprises at the point of regulatory submission. In addition, it significantly decreases the likelihood for further studies having to be performed down the line, should major data issues disqualify or cast doubts on the results of a trial.

To successfully implement CSM, a number of important factors must be considered. In the early stages of a trial or in studies that employ numerous small centers, the volume of data available may be too limited to perform statistical tests to detect abnormalities. In addition, as CSM is reliant on computerized data, the technique may miss some types of fraud or errors that can only be detected during site visits such as evidence provided in hand-written documents or interviews. This considered, on-site visits will be more efficient if monitors perform visits with a priori information about unexpected data patterns identified by statistical monitoring already at their disposal.

1.2. CluePoints Methods

The SMART™ engine forms the basis of CluePoints' Central Statistical Monitoring solution. It employs complex statistical algorithms to drill down into individual patient data to detect issues that could compromise the study and jeopardize successful regulatory submission. This approach necessitates minimal work for the sponsor in gaining objective information in order to optimize on-site monitoring by targeting centers at risk. The protocol is implemented in the same manner in all sites involved in a study using the same CRF. Throughout the duration of the study, variables are grouped by CRF section. These are then grouped by visits, visits are grouped by patient, patients are grouped by investigator, investigators grouped by center, centers grouped by country, and countries are grouped by geographical region.[8]

In the case of a randomized trial, the group allocated by randomization provides another design feature that allows for specific statistical tests to be performed, because baseline variables are not expected to differ between the randomized groups, while outcome variables are expected to differ equally in all centers if the treatments under investigation have a true effect. The ability to compare the distribution of all variables in each study site with other sites, means that abnormal patterns can be identified. The approach rests on the premise that the multivariate structure and time dependence of variables in statistical checks are highly sensitive to deviations and extremely difficult to imitate. Comparisons can be performed with either one variable at a time in a univariate fashion or with several variables, taking into account the multivariate structure of the data, or using longitudinal data when the variable is repeatedly measured over time.[8]Fabricated data will exhibit abnormal multivariate patterns that are detectable statistically. In addition, humans are known to be poor random number generators, meaning that tests on randomness can be used to detect data that have been falsified.[8]

Using a statistical approach necessitates a large number of statistical tests to be performed: tests on proportions of outliers, means, global variances, within-patient variances, event counts, distributions of categorical variables, proportion of week days, proportion of missing values, correlations between several variables, and so on and so forth. These tests generate a high-dimensional matrix of p-values, which can be analyzed by statistical methods and bio-informatic tools to identify outlying centers. An additional benefit of CSM is that sponsors who strategically outsource to CROs are finding increased efficiencies by using the method as an oversight tool to regularly check the quality of their data. As a result, the solution is also proving essential in helping sponsors select the best sites for future trials.

2. A brief history of the development of SMART (Statistical Methods Applied to Research Trials)

This section provides a brief overview of the development of SMART™, along with relevant bibliographic references (some of which are currently undergoing peer review).

2.1. Statistical methods for the detection of fraud

In 1997, a clinical trial reported one of the most remarkable cases of fraud of all times: a single site had fabricated all of their 438 patients, none of whom was ever questioned in spite of intensive on-site monitoring visits (The ESPS2 Group, 1997). Eventually the fraud was established beyond a reasonable doubt using statistical methods. This and other prominent cases of fraud motivated a group of statisticians to publish papers on the use of statistical methods for the detection of fraud in clinical trials (Buyse *et al.*, 1999; Buyse and Evans, 2001; Buyse, 2010). The same group carried out a survey of the membership of the International Society for Clinical Biostatistics (ISCB), which showed that the prevalence of fraud in clinical research was, in all likelihood, not as low as is commonly assumed (Ranstam *et al.*, 2000).

2.2. First software version (S+)

In 2000, the International Drug Development Institute (IDDI, Belgium), CluePoints' sister company, undertook to write a software application based on the concept that fraud, and more generally errors in clinical data, can be detected using statistical methods. The software was written in S+ and was intended to be used in a randomized clinical trial sponsored by the Association Européenne de Recherche en Oncologie (AERO) and funded by Bristol-Myers Squibb and Chugai. The trial randomized 873 patients in 78 centers in France. Centers participating to the trial were randomized

between two monitoring methods: one group of sites would receive routine on-site visits, the other would be monitored centrally using the statistical software. An early analysis showed no difference between the randomized groups of centers in terms of patient recruitment, quantity and quality of data submitted to the trial coordinating office or duration of patient follow-up (Liénard *et al.*, 2006). Unfortunately, the randomized question of monitoring was discontinued after initiation visits at the request of the participating centers.

2.3. **Second software version (S+)**

In 2006, IDDI undertook to rewrite the software in SAS and add capabilities, in particular an upload module that could read SAS datasets and automatically detect the type of each variable, and hence the appropriate statistical tests. A collaboration was undertaken with Sanofi-Aventis to re-analyze a number of clinical trials with known fraud. The software was very slow (with each run taking several hours). Besides, even though it did detect atypical data patterns, it did not appear to have enough sensitivity and specificity to be recommended for routine use. [2]

2.4. **Third software version (R)**

In 2009, IDDI obtained funding to rewrite the software once again, under the name SMART™ (Statistical Monitoring Applied to Randomized Trials). The R open source environment was chosen for reasons of efficiency, versatility and flexibility. The funding was granted by the Walloon Region in Belgium to a Consortium of two industrial and two academic partners: IDDI (project leader), GlaxoSmithKline Vaccines (GSK), and the Universities of Brussels (ULB) and Louvain-la-Neuve (UCL). The software was developed, tested and validated on simulated trials at IDDI, ULB and UCL; it was further validated on actual trials at IDDI and GSK. Two major breakthroughs made this version of the software a powerful tool for central monitoring of data quality: the use of random effects models and the addition of a scoring method to rank centers and "flag" those centers whose data appear atypical when compared to all other centers (Venet *et al.*, 2011; Desmet *et al.*, 2013). A collaborative agreement was signed with the Clinical Trial Service Unit at Oxford University (UK) for the validation of SMART™ in large academic trials in cardiovascular disease.

2.5. **CluePoints, Inc. (Cambridge, MA) and CluePoints, S.A. (Belgium)**

In 2012, a patent application was filed for SMART™ in the US. The company CluePoints was founded in November 2012 to commercialize centralized statistical monitoring services using SMART™ in the US and Europe (www.cluepoints.com). The number and variety of trials analyzed by CluePoints so far provides real-life evidence that centers flagged by SMART™ may have data quality issues that are worth investigating further. Typically, centers identified as atypical fall in 3 categories: (A) those in which the data discrepancies can be explained and require no further action; (B) centers in which the data discrepancies can be resolved centrally after contacting the center and reviewing the findings; and (C) centers in which the data discrepancies appear so serious that an on-site audit visit is indicated. CluePoints has also been run on completed trials about to be published, with no major findings (see, *e.g.*, the report on a large trial in gastric cancer to be published by Tsuburaya *et al.*, 2013). The methodology used in SMART™ is under constant improvement, supported by methodological research carried out by the Consortium of IDDI, ULB and UCL (see, *e.g.*, Desmet *et al.*, 2014). Development is carried out by CluePoints, and includes a SaaS (Software as a Service) solution

that will be available in the first quarter of 2014 to enable trial sponsors or other interested parties to carry out their own analyses of data quality. [3]

3. What is SMART™ ?

This section develops the concept and statistical principles underlying SMART™, a system that was developed and perfected over a number of years *to check data consistency* in clinical trials.

3.1. The concept underlying SMART™

In clinical trials, data are collected using the same (electronic) case report form in the centers that enter patients in the trial. SMART™ is a software that checks if the data collected are *consistent* across all centers, and if they are not, it identifies those *centers that differ* substantially from the others. All centers participating to a clinical trial use the same protocol, hence the data collected are expected to be consistent across all centers, except for obvious differences in patient characteristics or other regional factors. Moreover, data collected in clinical trials are generally complex and diverse, so they are ideally suited to elaborate statistical checks.

3.2. Modules of SMART™

SMART™ consist of the following modules (written in "R"):

- A **data upload module** that reads the database and identifies the type of all variables collected (*e.g.* binary such as gender, continuous such as blood pressure, dates such as visit dates).

- Several **statistical test modules** that perform as many statistical tests as possible on all variables collected (*e.g.* tests for differences in means or variances, proportions of missing values, distributions of first digit). Table 1 shows the statistical tests that are currently implemented in SMART™. Each statistical test compares one center to all others in a specific study or suite of identical studies. The results of all statistical tests are kept in a large matrix of *P*-values having as many rows as centers and as many columns as statistical tests. For a clinical trial in which data are collected on 300 variables at 100 centers, a total of 150,000 statistical tests would be carried out (assuming an average of 5 statistical tests per variable). Note that the number of statistical tests performed depends on the size of the case report form and the number of centers, but not on the number of patients.

- A **scoring module** that calculates a "*Data Inconsistency Score (DIS)*" for each center. The *DIS* of a center is analogous to an average *P*-value, and as such it quantifies the probability that the differences observed between the data from that center and the data from all other centers were due to chance (*i.e.* random variation) alone. Centers that have small *DIS* (small *P*-values) may be in need of further scrutiny.

- A **reporting module** that generates a report to be edited by a statistical analyst. The report provides a table of *DIS* for all centers, and details on data issues in centers having small *DIS*.

Tag name	Corresponding test	Compared to all centers, an example of an atypical pattern is:
Reporting	**Tests for Number of Reported Values**	
Count	Count of non-missing records	Higher than expected number of reported adverse events.
Missing	Proportion of missing values	Heart rate not recorded in a too high number visits.
Proportion	**Tests for Proportion of a Category of a Categorical Variable**	
Proportion	Proportion of a category of a binary variable	Larger than expected proportion of serious adverse events.
	Proportions of the categories of a factor variable	Different than expected proportions of three medication routes.
Days	Proportion of Saturdays	Too high number of Saturdays amongst the visit dates.
	Proportion of Sundays	Too high number of patients randomized on a Sunday.
Distribution	**Tests for Distribution of a Numeric Variable**	
Outliers	Proportion of global outlier values	Too high number of outliers for weight measurements.
Mean	Mean value of continuous measurements	Lower than expected mean temperature of the patients.
Variability	Between-patient variability	Too similar height measurements for all patients.
Sequences	**Tests for Sequences of Repeated Measurements**	
Propagation	Proportion of propagated values	Too many replicated values along sequences of successive weight measurements per patient.
Variability	Within-patient variability	Too small range of variation within sequences of successive measurements of the blood pressure per patient.
Outliers	Proportion of sequence outliers	Too high number of unexpected jumps within the sequences of successive weight measurements per patient.
Transition	Sequence transition 1 –> 0	When the category is TRUE in a sequence of measurements, a too high number of observed transitions to the category FALSE.
	Sequence transition 0 –> 1	When the category is FALSE in a sequence of measurements, a too small number of observed transitions to the category TRUE.
Start	Starting value	Too high proportion of sequences of measurements starting with the category TRUE.

Table 1: Tag Names for the Different Tests and Groups of Tests Performed by SMART™

References

1. Baigent C, Harrell F, Buyse M, Emberson J, Altman D. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials 2008; 5: 49-55.

2. Buyse M, George SL, Evans S, Geller N, Ranstam J, Scherrer B, Lesaffre E, Murray G, Edler L, Hutton J, Colton T, Lachenbruch P, Verma B, for the ISCB Subcommittee on Fraud. The role of biostatistics in detection and treatment of fraud in clinical trials. Statist Med 1999; 18: 3435-3451.

3. Buyse M, Evans S. Fraud in clinical trials. In: Biostatistics in Clinical Trials (Colton T, Redmond C, *eds.*), Wiley, New York, 2001.

4. Buyse M. Fraud detection. In: Encyclopaedic Companion to Medical Statistics (Everitt B, Palmer C, *eds.*), Arnold Publishers Ltd, London, 2010.

5. Desmet L, Venet D, Doffagne E, Timmermans C, Burzykowski T, Legrand C, Buyse M. Signal detection and power in central statistical monitoring of multicenter trials. Submitted, 2013.

6. Desmet L, Venet D, Doffagne E, Timmermans C, Burzykowski T, Legrand C, Buyse M. Use of the beta-binomial distribution in central statistical monitoring of multicenter trials. In preparation, 2014.

7. European Medicines Agency. Reflection paper on risk based quality management in clinical trials. EMA/INS/GCP/394194/2011. February 2011. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/WC500110059.pdf (Accessed October 31, 2013).

8. Liénard JL, Quinaux E, Piedbois P, Jouhaud A, Decoster G, Buyse M. Impact of on-site initiation visits on patient recruitment and data quality in a randomized trial of adjuvant chemotherapy for breast cancer. Clin Trials 2006; 3: 486-492

9. Ranstam J, Buyse M, George SL, Evans S, Geller N, Scherrer B, Lesaffre E, Murray G, Edler L, Hutton J, Colton T, Lachenbruch P, for the ISCB Subcommittee on Fraud. Fraud in medical research: an international survey of biostatisticians. Controlled Clin Trials 2000; 1: 415-427.

10. The ESPS2 Group. European Stroke Prevention Study 2. Efficacy and safety data, J Neurol Sciences 1997; 151 (Suppl): S1-S77.

11. Tsuburaya A, Yoshida K, Kobayashi M, Yoshino S, Takahashi M, Takiguchi N, Tanabe K, Takahashi N, Imamura H, Tatsumoto N, Hara A, Nishikawa K, Fukushima R, Nozaki I, Kojima H, Miyashita Y, Oba K, Buyse M, Morita S, Sakamoto J. Adjuvant paclitaxel followed by oral fluorinated pyrimidines for locally advanced gastric cancer: The SAMIT randomized trial. Submitted, 2013.

12. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring. August 2013. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM269919.pdf (Accessed October 31, 2013).

13. Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, Genevois-Marlin E, Becker U, Bee V, Wilson V, Legrand C, Buyse M. A statistical approach to central monitoring of data quality in clinical trials. Clinical Trials 2012; 9: 705-713.