

Statistical Assessment Layer for Site Audits (SALSA)

Concept Paper

A statistical approach for data-driven site auditing

Version 2.0, 30 July 2020

Index

1. Executive Summary	3
2. Introduction	3
3. Description of Data Sets and Variables	4
3.1. Dataset Summary Table.....	5
3.2. Variable Specific Evaluations	5
3.2.1. Quantitative Distribution Analysis	5
3.2.2. Tables with 10 smallest values, largest values and most frequent values	6
3.2.3. Digit Analysis	7
3.2.4. Inter and Intra-subject Variability.....	8
4. Real Life Examples from Investigational Site Audits.....	9
4.1. QRS value	9
4.2. Body weights (1).....	11
4.3. Body weights (2).....	13
4.4. Respiratory Rate.....	14
4.5. Blood Pressure	15
5. Added value of statistical assessments for audit support	17
6. Further uses of SALSA in risk-based Quality Assurance	17

1. Executive Summary

In recent years, regulators have expressed increased expectations on sponsors' utilization of statistical methodology for the monitoring of data quality. Such concepts have thus far been used to only a limited extent in the area of auditing.

The Engaged Database has developed a tool which allows a rapid analysis and visualization of selected study data in order to evaluate data patterns for selected sites and individual subjects. This allows identification of unusual data patterns and statistical abnormalities that may point to differences between subject populations at different clinical sites as well as identification of outliers for individual subjects or data points.

Our tool can be used for the selection of sites to be audited as well as for the preparation of individual site audits and adds value to the efficacy and objectivity of auditing processes. SALSA makes the audit "smarter" because we are utilizing an evidence-based rationale for the data we select for review at the site and for the questions we ask site staff and the CRA.

This summary document describes the analysis of a basic data set which consists of parameters which are obtained in almost all clinical trials; additional data can be added on a study specific basis, if required. Examples of real-life analyses are presented where the tool was used during preparation for investigational site audits.

Finally, additional potential uses of the SALSA tool as part of a sponsor's risk-based Quality Assurance toolkit are discussed. These include providing data-based evidence on the overall data quality and integrity of a site to aid in risk evaluation and to support sponsors' decisions when changes to the annual audit plan become necessary.

2. Introduction

The design and conduct of clinical trials have become more and more complex during the last 15 years. An increased number of trials are conducted in an international, multi-regional setting, using an increased number of external service providers and collecting a large amount of data from various sources, including central laboratories, central imaging reading providers, central ECG assessment teams and patient reported outcome tools. This increase in the complexity of trials and the amount of data collected has become a challenge for clinical Quality Assurance (QA), as the resources allocated to ensuring quality and compliance in clinical trials normally cannot be increased proportionately to these developments.

At the same time, the majority of trials now use electronic data capture (EDC) techniques for at least part of the data being collected, making this data available in real time and remotely without the need to travel to the sites. One way to address the challenges for clinical QA is to use the available data to support QA related activities and decision making.

In a traditional QA approach, the quality of data at investigational sites is assessed during on-site audits by verifying a limited number of CRF data against the original source data. Such an approach is actually below the standards suggested by regulators, who expect that statistical methodology should be used to monitor data quality (e.g. EMA Reflection paper on risk based quality management in clinical trials).

We have developed a process together with biostatistical experts with the aim to analyze data exported from an EDC system (Medidata Rave) using standardized statistical methods to identify meaningful information that can be used in the QA services we offer to our clients.

A set of the standard variables used in most clinical trials has been selected for which data is analyzed on individual subject level as well as on site level using tables, graphs and statistical tests to identify implausible data, unexpected values, unusual data patterns and trends, and data extremes (outliers). Although these data points are not the key data in most clinical trials, they serve as an indicator for data quality and integrity and are at the same time relevant for subject safety and potentially for dosing calculations for investigational medical product (IMP). An analogous approach was used by authorities in several cases in which study data was considered invalid due to issues with ECG data, even if these were not the key data of the respective trials (e.g. data manipulation within a Phase I Unit in Hyderabad, India, identified in 2014).

The intention is to use the outcome of the analysis as supportive information for QA activities, specifically auditing. The analysis can support the process of site selection or auditor preparation prior to conduct of an on-site investigator audit or a remote investigator site audit (where audit sampling has to be even more tailored) to help the auditor to select subjects and data to be reviewed during the audit, equipment and equipment documentation to be scrutinized and processes and procedures to be discussed in detail with staff. Information gathered from SALSA therefore provides an additional and efficient insight into the data quality and integrity and makes audit results more robust.

It is important to note, however, that there is no universally applicable approach to identify data patterns and outliers and non-compliances of sites are seldom confirmed by statistical evidence alone. Unexpected data can arise for multiple reasons and caution needs to be exercised to not over-interpret unusual data patterns, specifically when obtained using small sample sizes. Data analysis therefore serves as a support for further investigation, in our case during audits, and is not intended to replace other methods of compliance and safety monitoring like on-site monitoring visits or central and medical monitoring or safety signal detection processes implemented by the sponsor or CRO(s).

3. Description of Data Sets and Variables

Study data exported from the EDC system (typically in MS-Excel format by Data Management or the auditor who is granted access to the EDC system for an audit) is analyzed and presented in a statistical report. The data extracted includes at least that from one (1) participating investigational site but ideally contains data from several other sites for reference and comparison.

The report generated can include data for a total of sixteen (16) standard variables which are collected in most clinical studies. In addition, study specific variables can be included in the analysis on a study specific basis (e.g. data related to the primary endpoint. The variables are organized into five data sets:

- A vital sign data set
- A demographic data set
- An ECG data set
- An adverse event (AE) data set
- A concomitant medication data set

The vital sign data set is comprised of the variables height, weight, body temperature, respiratory rate, systolic and diastolic blood pressure and pulse rate. The demographic data set includes the variable

gender and the ECG data set includes the duration of the QRS complex and the QT duration (raw data and corrected QT). The adverse event data set includes the variables number of AEs, seriousness and relationship. The concomitant medication data set includes the number of concomitant medications administered.

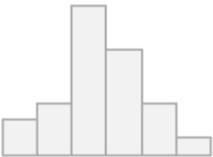
The report allows selection of data to be displayed first by investigational site and then by one of the five data sets described above. The variables contained in the selected data set are then displayed for the selected site. Data is always presented for one site at a time.

Each data set is presented in a data set summary table, variable-specific evaluations and a data appendix including AE and concomitant medication dataset: as these variables are not measured on a metrical scale no variable specific evaluations can be calculated. The data appendix functionality allows that any data point shown in one of the analysis results can be easily traced back to an individual subject or subjects.

3.1. Dataset Summary Table

Data Summary Tables include variable names, central tendency and dispersion measures for each variable (min, max, mean, median, inter quartile range coefficient of variation), the number of distinct values included in the data set of the selected site and a graphical representation of the data.

Example: Summary Table for the variable “height” from the Demographic Data Set Summary

Variable	Label	Stats / Values	Freqs (% of Valid)	Graph
VSHT_RAW_cln [numeric]	Height	Mean (sd) : 164.9 (6) min < med < max: 154 < 164 < 180 IQR (CV) : 6.5 (0)	16 distinct values	

The Vital Sign Data Set Summary Table is used to obtain a first impression of the data set and provides an overview “at a glance” of the dispersion of data and outliers.

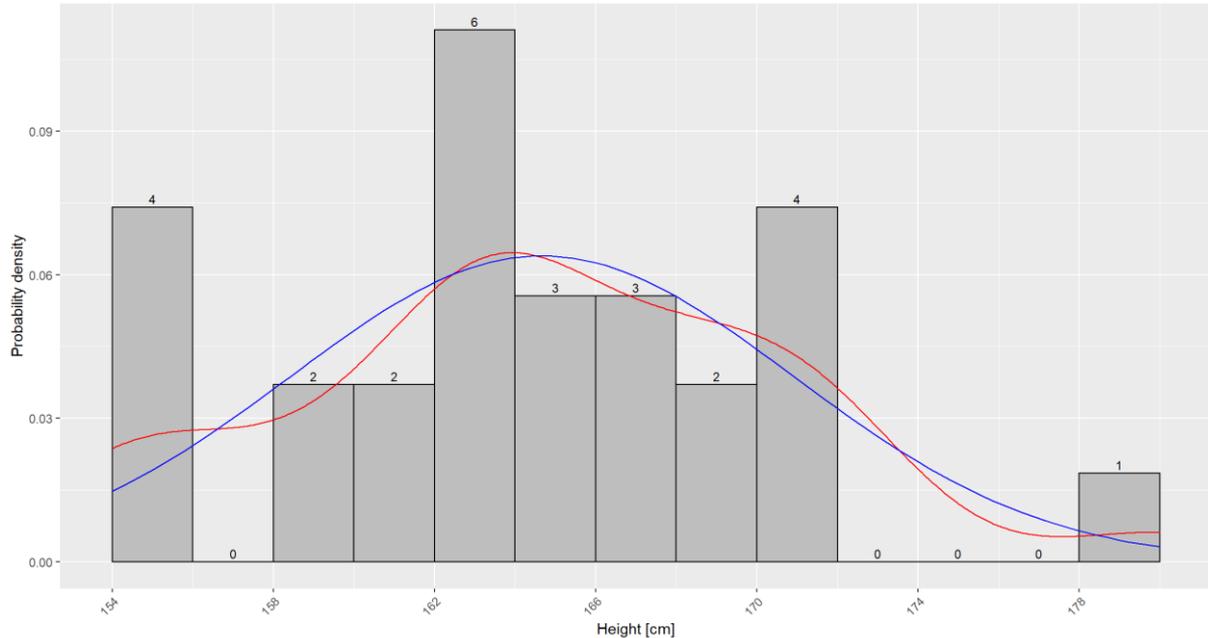
3.2. Variable Specific Evaluations

3.2.1. Quantitative Distribution Analysis

For each variable of the selected data set a graphic display of data distribution is provided which includes a graph showing a normal distribution with mean and standard deviation calculated from the data set (blue line) and a density estimation of the data set (red line). In addition, a check for normal distribution of the data is provided for the variables height and weight using three different statistical tests (Shapior-Wilk test, Anderson-Darling test and Lillefors test).

Example: Data distribution for variable “height” including normal distribution (blue line) and density as derived from the data set (red curve) and results of three (3) tests for normal distribution.

Distribution of Height



Each bar represents an interval of width 2 cm. Numbers on top of bars indicate counts in respective interval.
 Red curve: Density as estimated from data.
 Blue curve: Normal distribution with mean and standard deviation of data
 Probability to observe data if population distribution is Gaussian (N =27):
 0.57 (Shapiro-Wilk test)
 0.63 (Anderson-Darling test)
 0.72 (Lilliefors test)

This representation of the data is used to review the dispersion of the data in detail, identify outliers and check for a normal distribution of the data for variables where a normal distribution is expected. Whether the data is expected to follow a normal distribution depends not only on the variable but also on the population being reviewed and the informative value of the statistical tests depend on the sample size which equals the number of subjects enrolled by the site for which the data is displayed. Information can be derived by analyzing individual sites or comparing data from different sites. See below for an example.

3.2.2. Tables with 10 smallest values, largest values and most frequent values

Example: ten smallest values of variable height (largest values and most frequent values not shown):

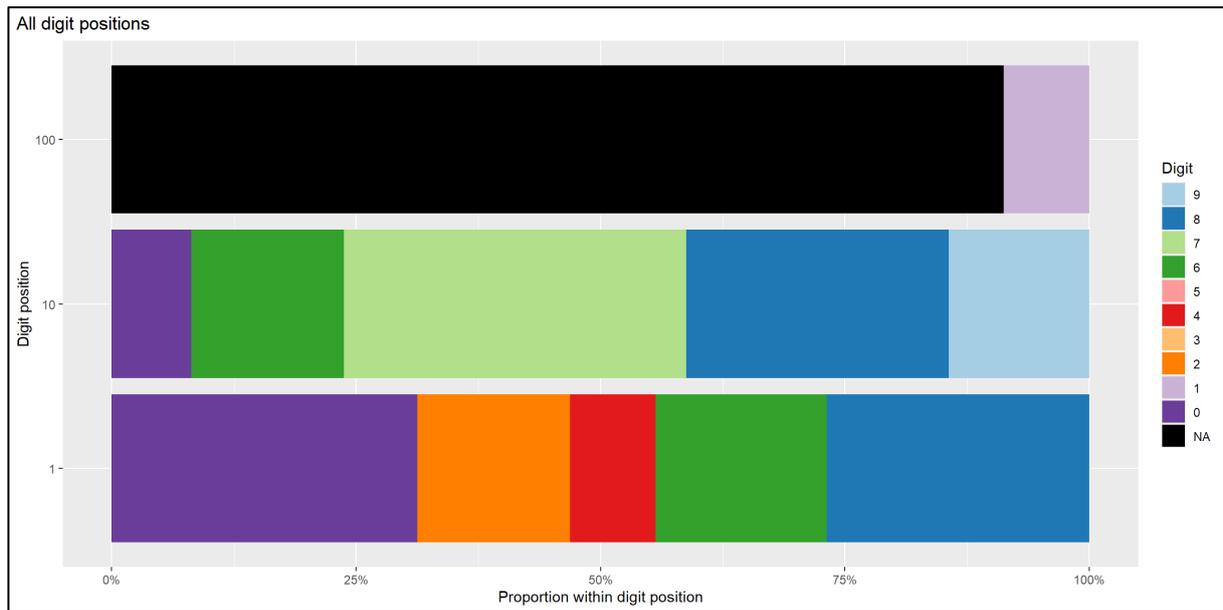
Value	Frequency	Relative Frequency
154	2	7.41%
155	1	3.70%
156	1	3.70%
159	2	7.41%
161	1	3.70%
162	1	3.70%
163	4	14.81%

Value	Frequency	Relative Frequency
164	2	7.41%
165	2	7.41%
166	1	3.70%

3.2.3. Digit Analysis

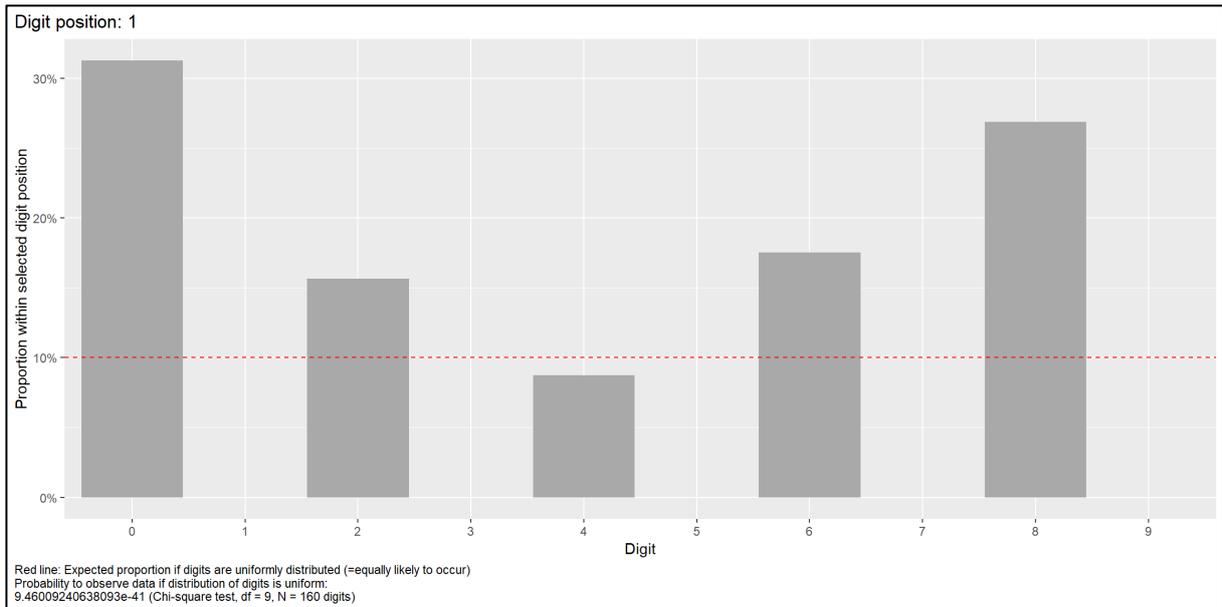
A graphical representation of the frequency of each digit occurring for the values of each variable is displayed showing the proportional representation of each digit from 0 to 9 as a percentage.

Example: Proportion of digit positions 1, 10 and 100 for the variable QRS



In addition, the distribution of the lowest digit position is provided, including a Chi-Square analysis of the probability to observe the data under the assumption of a uniform digit distribution.

Example: Proportional distribution within digit position 1 of the variable QRS complex (position 1 is the lowest digit position for the variable QRS complex)



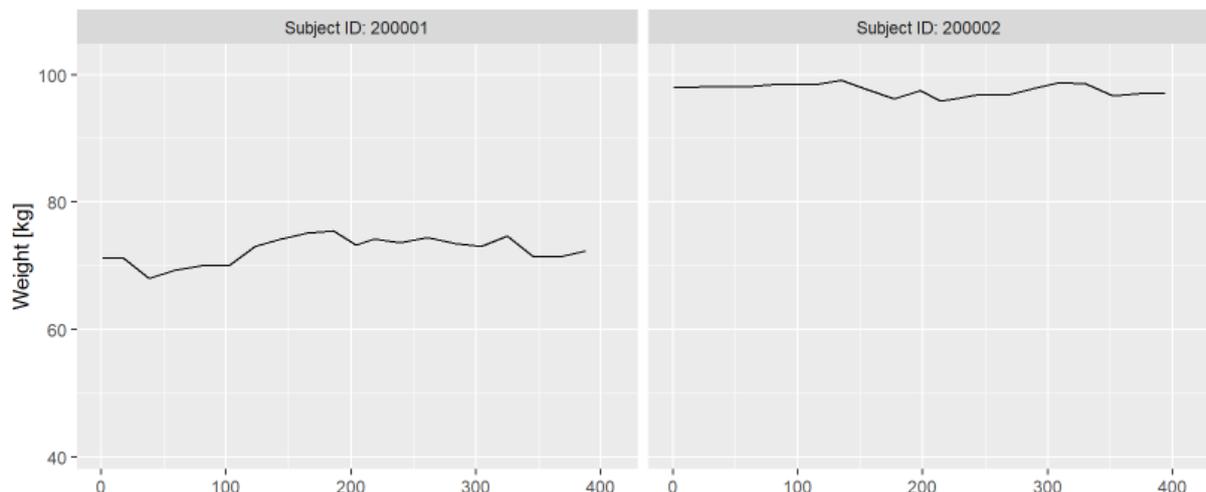
This data is used to identify unusual data patterns. For certain variables (e.g. weight) a uniform distribution is expected for the final digit. Unusual data patterns can be indicative of an imprecise measurement process (e.g. rounding) or inadequate equipment. For the interpretation of this analysis details of the instructions given to the site are to be taken into consideration (e.g. if sites are instructed to round results according to a specific algorithm) as well as standard approaches in clinical practices (e.g. rounding blood pressure results when measuring manually) and artefacts introduced by the equipment used. See below (section 4) for examples.

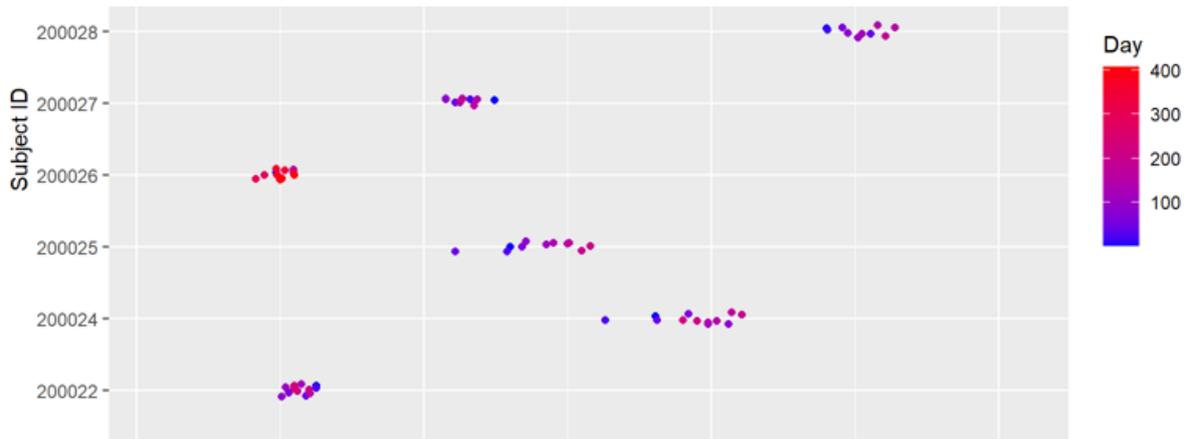
3.2.4. Inter and Intra-subject Variability

A statistical evaluation of variance including the inter-subject and intra-subject variability is provided. Data points of all variables for all visits of all subjects of a site are shown in comprehensive graphics (individual subject curves and in a dot plot graphic). This allows a visual impression of site and subject specific data variability and variations over time.

Example: body weight over time (as days in the study) is displayed for each subject (only two subjects shown in the line plot)

Weight versus Day by Subject





In the dot plot all data points for each subject are shown in a two dimensional representation. Each line shows the data points for one individual subject. The color allows a visual impression of the sequence in time, i.e. the dots are blue at the beginning of the study and become more red over subsequent visits.

4. Real Life Examples from Investigational Site Audits

The data presented is taken from phase 3 study sites. The analyses were used by the auditors in the context of preparation for investigational site audits for pre-selected sites. Site selection had been conducted prior to the statistical analysis. The data of the auditee sites was compared with data from other sites in the study to assess which data patterns might be due to site- or country-specific effects as opposed to study-specific data patterns which might be explained by study indication or the population under study.

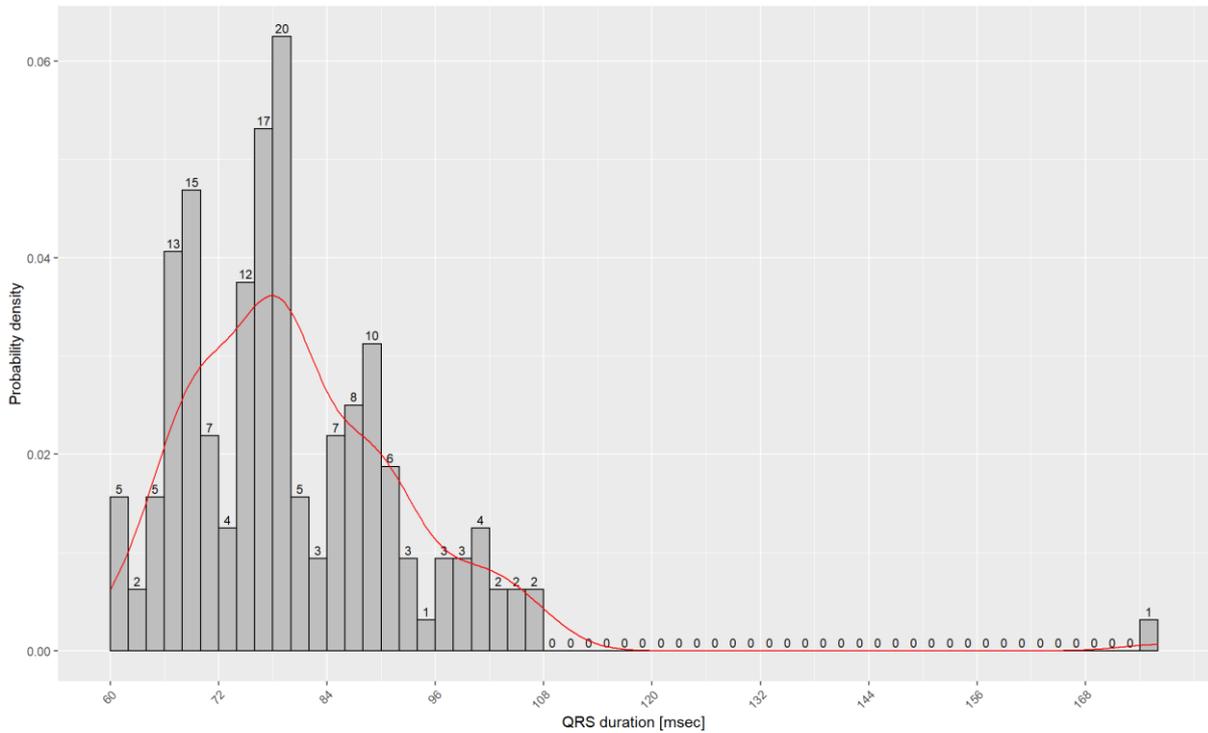
Unexpected data patterns were observed during initial data review of both the sites to be audited and the other sites included in the review for comparison. Results from review of the sites to be audited were used for audit preparation; unusual data patterns observed at other sites not among the sites to be audited were forwarded to the sponsor for further follow-up. This included a question raised due to the distribution of body temperature, which was within a very small range at some sites (36,0°C to 36,9°C) compared to other sites (35,3°C to 37,8°C).

Based upon the review of the data of the auditee sites, the audit sample was selected, i.e. patients with data that were unexpected or should be confirmed for any other reasons were included and methods and equipment used to obtain results were scrutinized during the audit. The following examples contain analytical results from the test SALSA site audits:

4.1. QRS value

One outlier for QRS values (176ms with the next value 106ms) was identified in the data set.

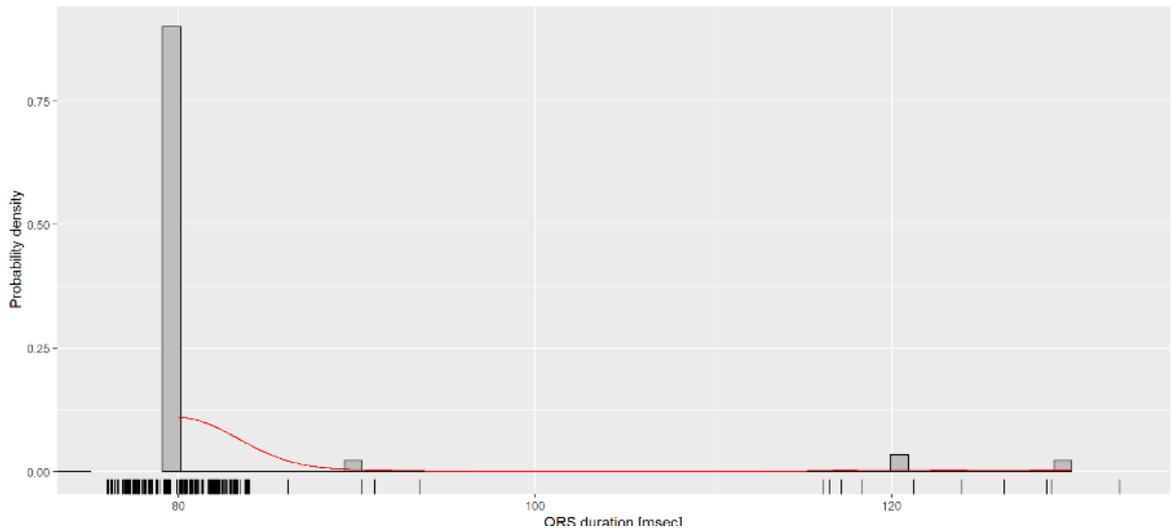
Distribution of QRS



Each bar represents an interval of width 2 msec. Numbers on top of bars indicate counts in respective interval. Red curve: Density as estimated from data.

On the basis of this analysis, the patient was selected for source data (SD) evaluation during the audit, and it was confirmed with the Study Coordinator (SC) that a) the value was indeed printed on the ECG, and b) that a comment had been added to the print-out by the investigator that the ECG was 'normal'. Subsequent checks confirmed that the eCRF system had auto-queried this value (according to the protocol, follow-up of elevated QRS results were required); however, the SC entering the data had answered the query 'confirmed value, result is normal'. Further, the CRA had source data verified (SDVed) the entry without taking any additional action, e.g. reporting as protocol deviation that no follow-up action had been taken. Processes for quality control of data and monitoring were subsequently reviewed and recommendations to strengthen the processes to avoid future undetected protocol deviations were made.

In another case, nearly all QRS values documented in the eCRF showed a result of 80ms.

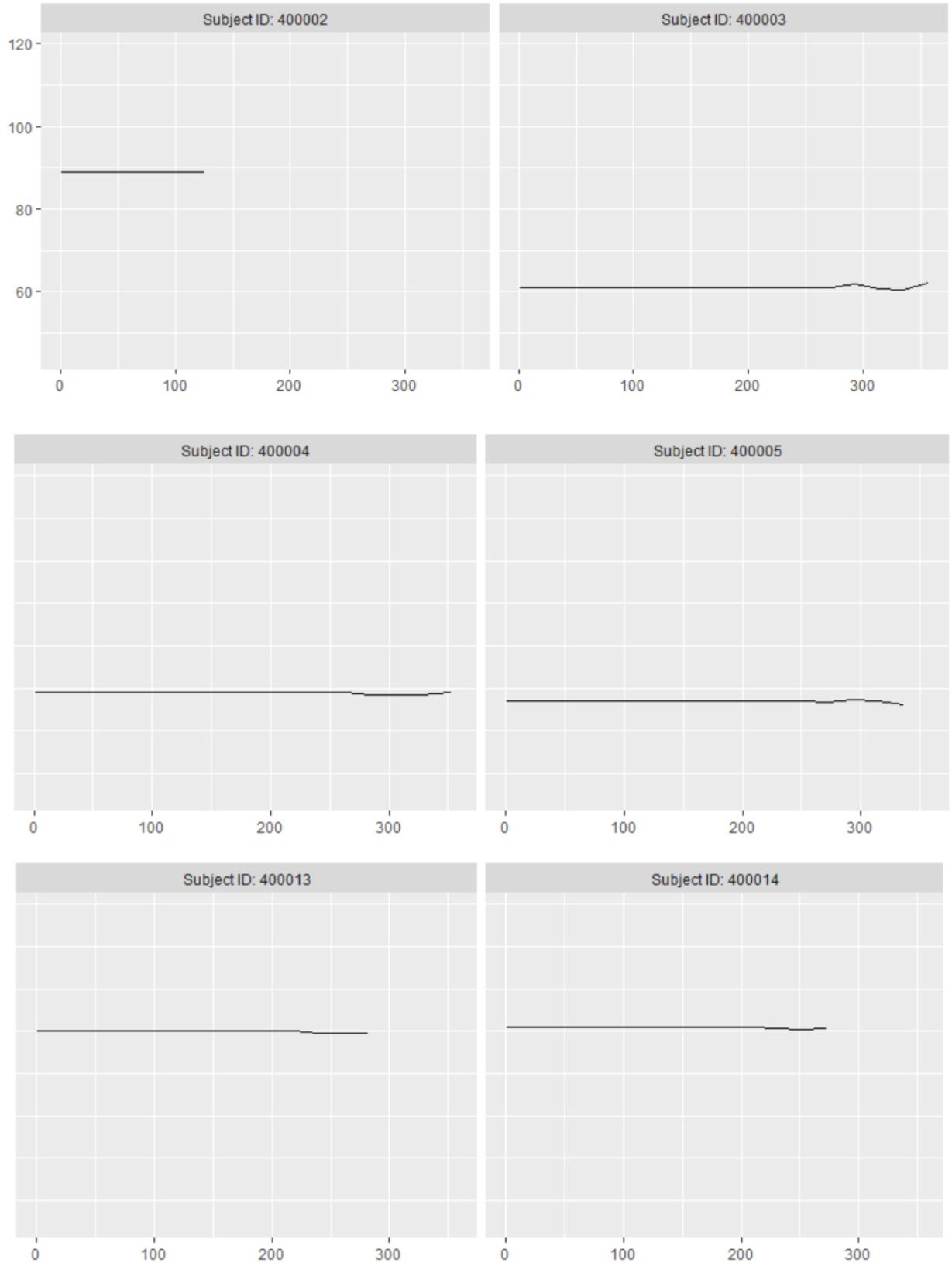


ECG curves were available for all subjects assessed during the audit and showed different results. No automatic assessment of the ECG values had been performed. This case was identified as an indication of poor manual evaluation during the audit. Due to the safety relevance of this parameter, further action was indicated.

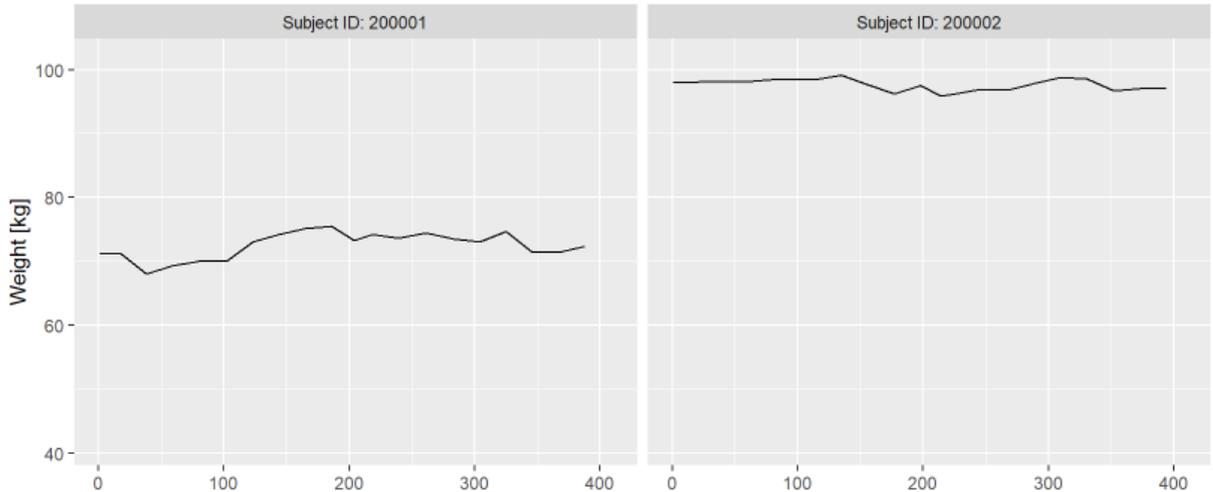
4.2. Body weights (1)

For all subjects (more than 15), no change of body weights was observed at all during their entire study participation (overall study duration more than 1 year), although the IMP was known to result in significant changes of body weights (up to 10 kg at other sites).

Examples of body weights at the affected site (note: the small variations at the right originate from data points obtained after the audit):



Examples of body weights at another site in the same study:

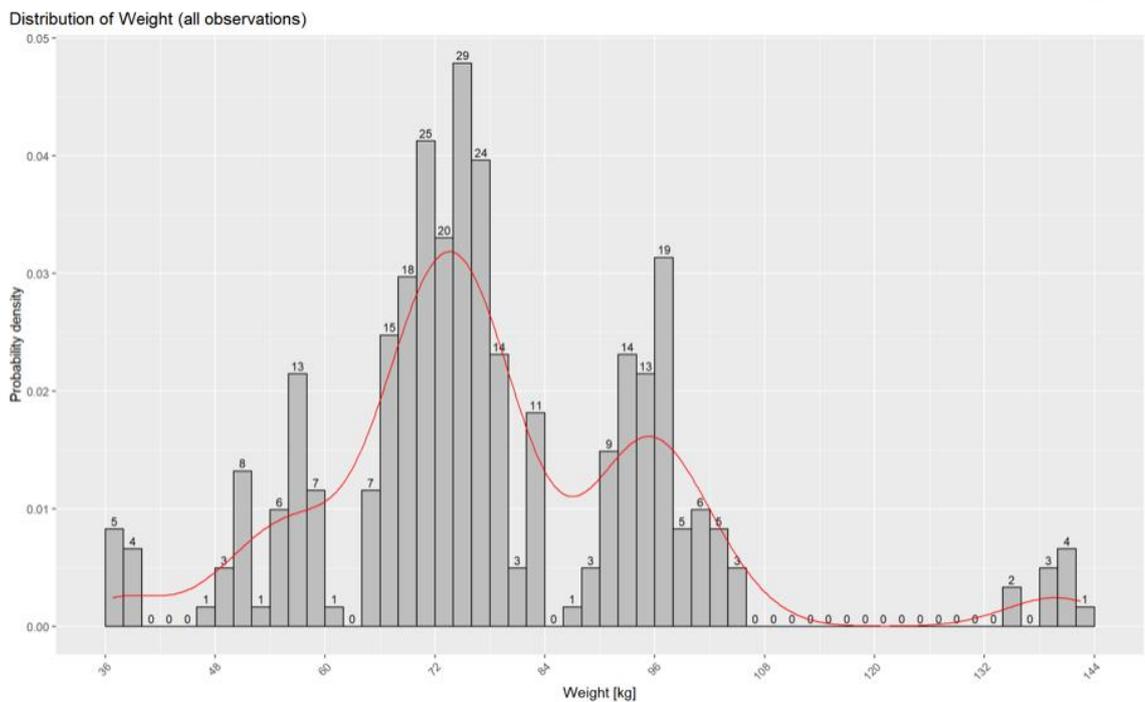


As dosing was weight dependent, this was potentially relevant for subject safety and efficacy. No plausible explanation could be provided by site staff.

From a statistical perspective, the probability that these data are accurate is negligible. Therefore, action was recommended to the sponsor, which – due to the time point the issue was raised – allowed the sponsor to make decisions related to risk evaluation.

4.3. Body weights (2)

Outliers in the distribution of body weights allow the auditor to focus on subjects for which this parameter can have a higher level of medical relevance.

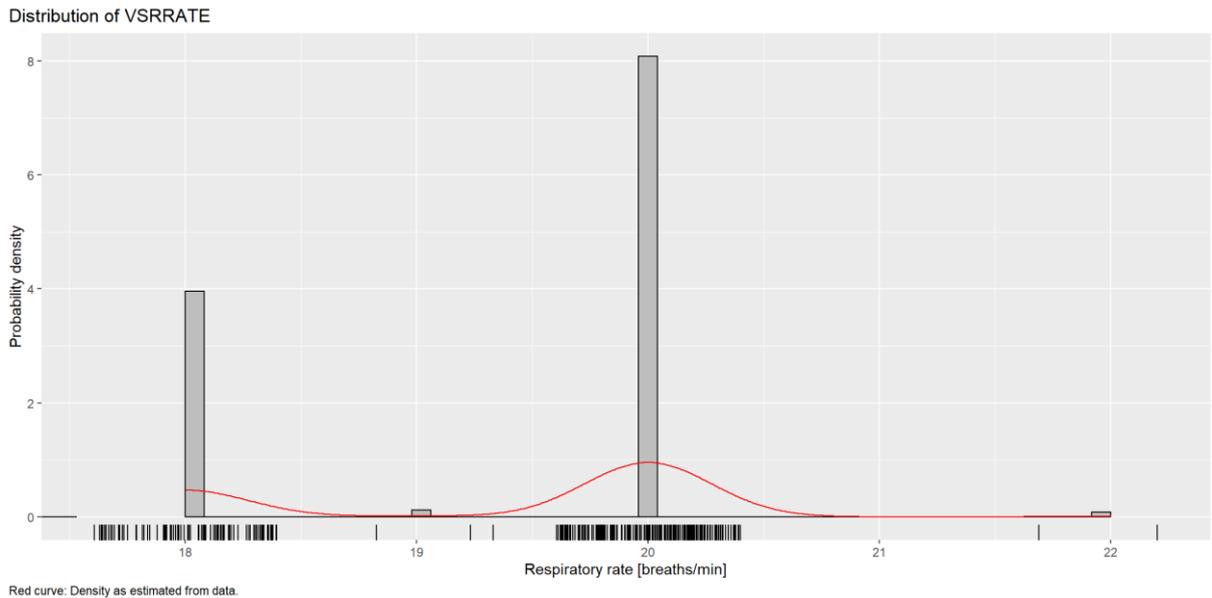


In this example, one (adult) subject had an extremely low body weight (all data points between 36 kg and 37 kg originate from this subject). This is a typical trigger for subject selection for the audit sample. After confirmation of the accuracy of these data, the auditor assessed whether or

not this low body weight was due to any co-morbidities, and if this might be relevant for inclusion/exclusion criteria. After the subject was confirmed to be eligible for the study, Adverse Events were thoroughly evaluated to confirm that Adverse Events were not related to the low weight of the patient as baseline condition.

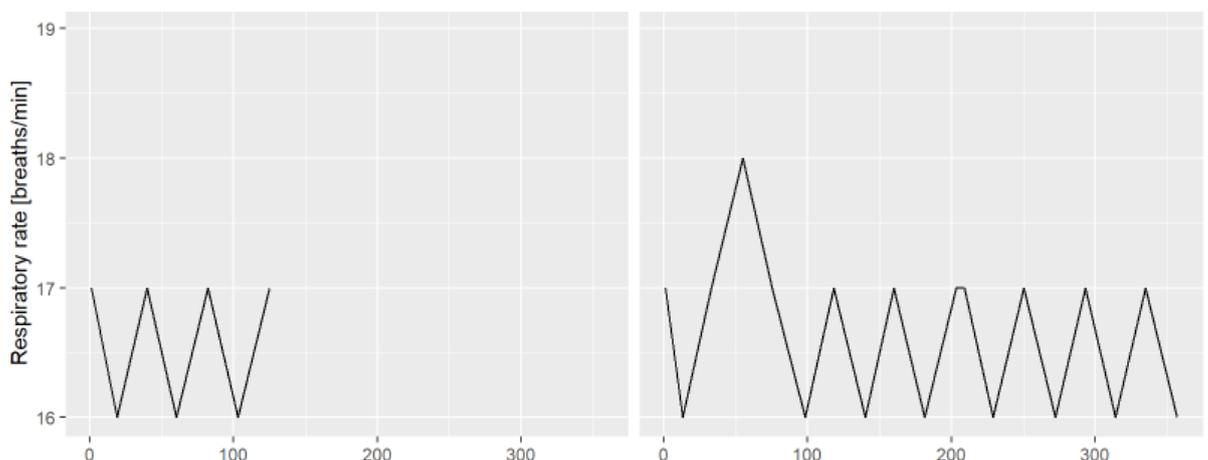
4.4. Respiratory Rate

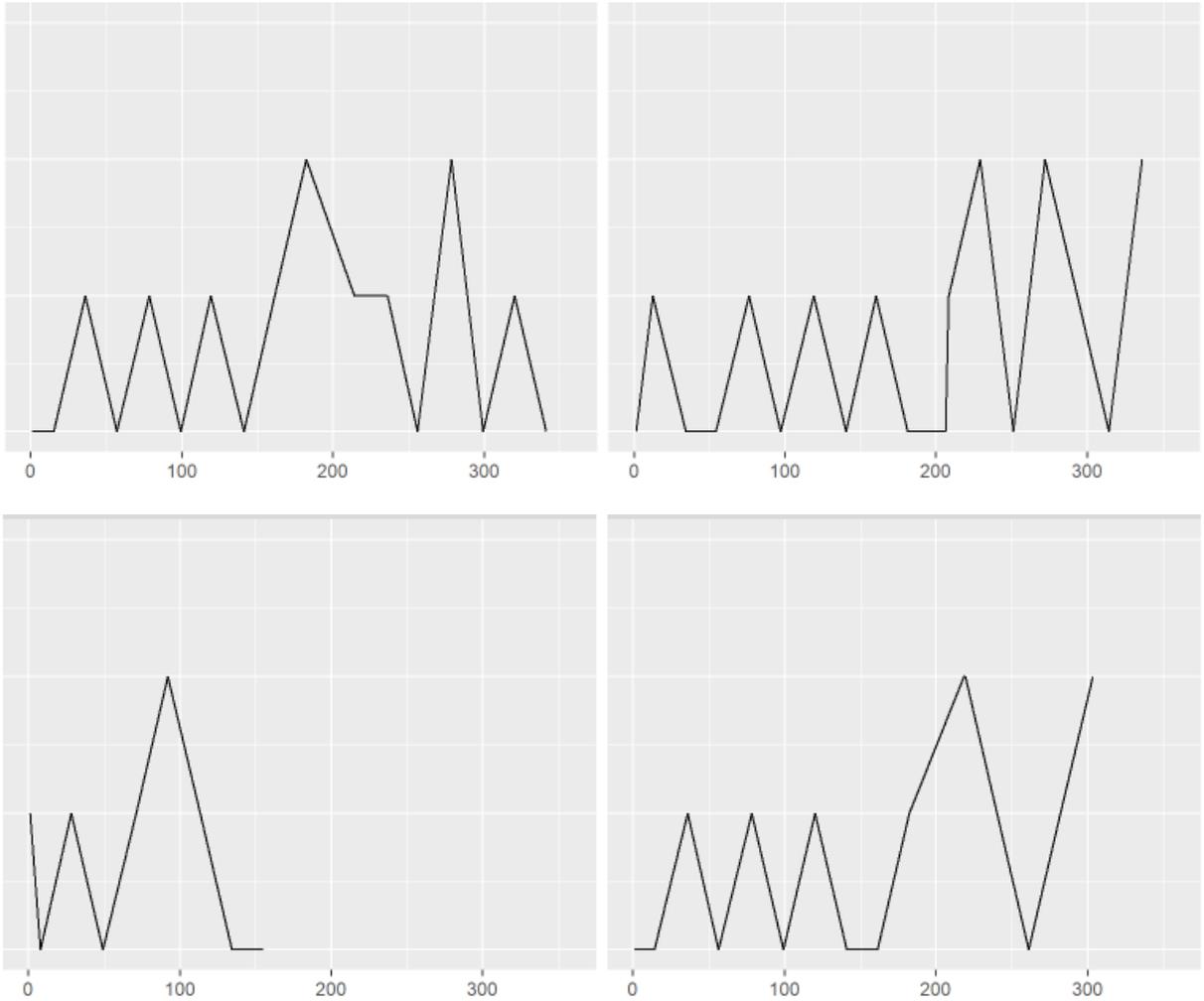
In this case, almost all respiratory rates had the values 18 (n=97) or 20 (n=198) breaths per minute, but only in three (n=3) cases was the value 19 recorded.



This evaluation allowed the auditor to identify a potential methodology issue in the RR examination (i.e. measurement for 30 seconds and the result multiplied by two (2) to calculate the rate per minute, instead of measurement for a full minute). This was followed up during the audit to see whether persons involved in the conduct took different approaches in measuring the RR.

In another case, no abnormalities in the overall distribution of values were detected, but a striking pattern in the sequence of data was noted:





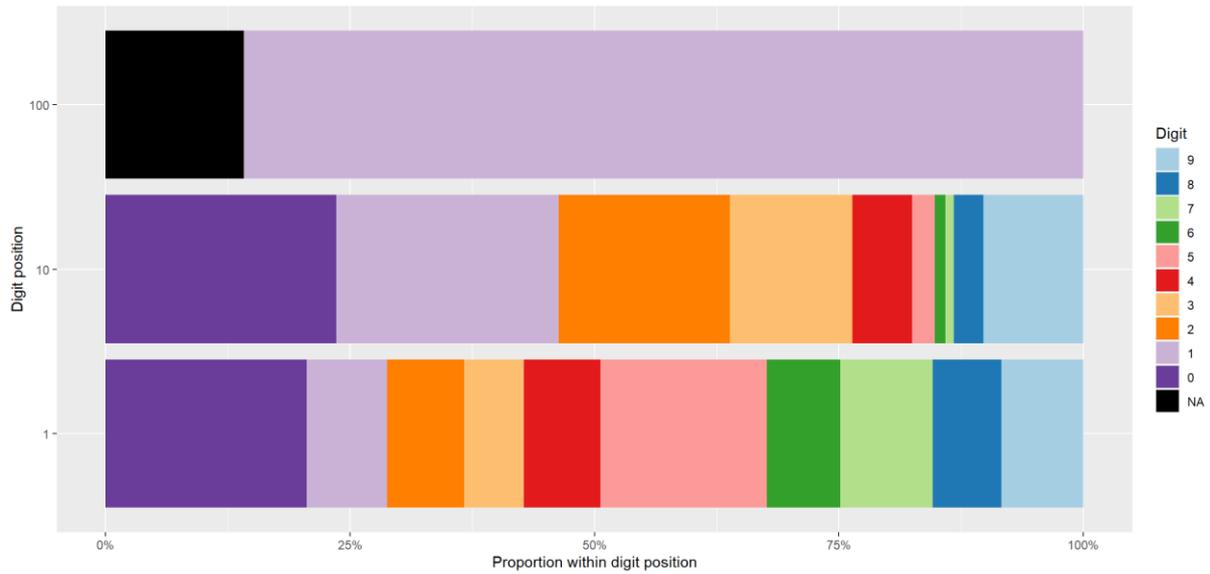
This was a clear signal that this data should be followed up during audit.

Whereas respiratory rate is normally not part of the key dataset of clinical studies, such information can be used together with other information to get a better understanding of the overall data quality at a site.

4.5. Blood Pressure

A preference of the last digits 0 and 5 for blood pressure values (ca. 20% for 0 and 17% for 5) was noted, whereas the expected value based on statistical probability would be 10%.

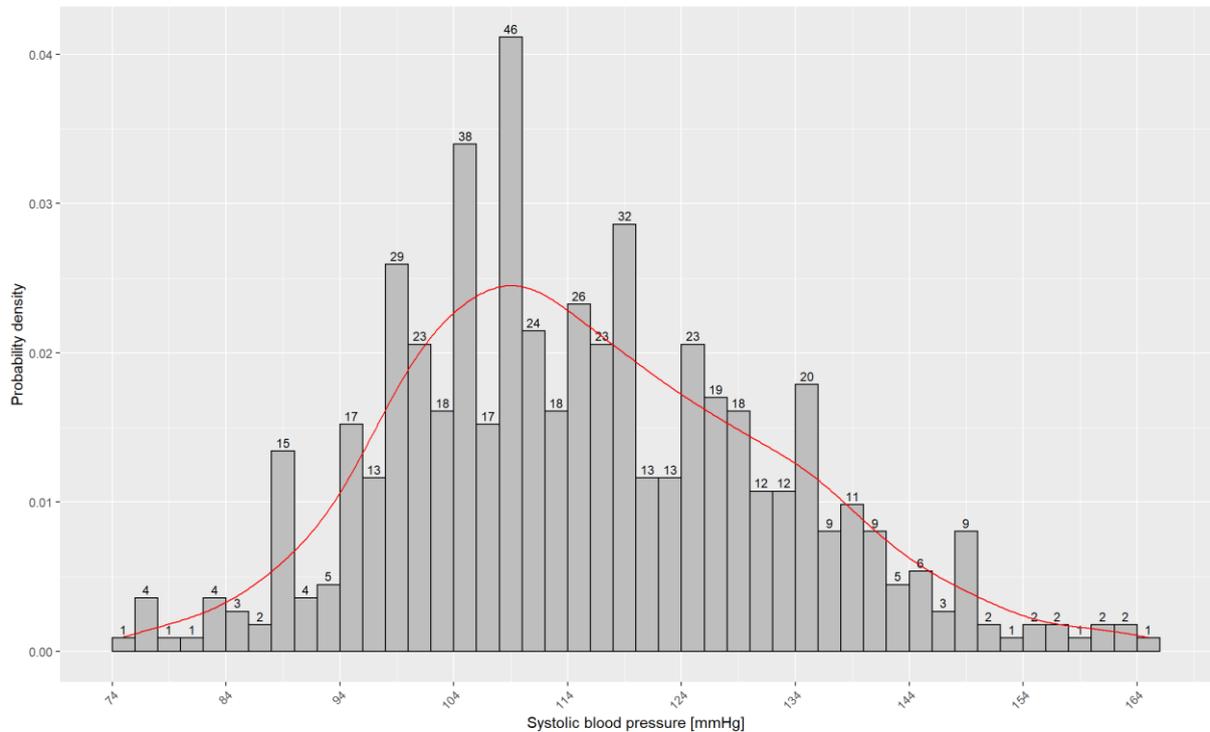
All digit positions



During the audit the use of different devices, and the conduct of the measurement by different persons with potential rounding was evaluated.

In another example, the proportion of blood pressure values in the hypertonic range was significantly higher at one site than at other sites.

Distribution of SBP



Each bar represents an interval of width 2 mmHg. Numbers on top of bars indicate counts in respective interval. Red curve: Density as estimated from data.

The auditor took this result as the basis for evaluation of medical history to confirm that a high number of patients had hypertension at the start of the study, and that this was entered in the medical records and eCRF, respectively.

5. Added value of statistical assessments for audit support

Using data analysis as supportive information for decision making during the audit process provides an additional layer of confidence in the process, ensuring a more robust basis for decision making before, during and after the audit. Decisions about the sites to be selected for auditing can be made based on additional and valuable information. Instead of a random selection of subjects and data to be reviewed during the audit, audits can be focused on previously identified areas of concern and risks ensuring efficient use of limited auditing resources.

In addition, the comprehensive review of cumulated data reveals data patterns which would not likely have been identified with the traditional audit methodology and which support the selection of methods and equipment reviewed during audits as well as topics to be discussed during interviews.

As such data driven approaches are increasingly used by regulatory authorities it is important for sponsor companies to implement such approaches into their routine QA and audit process to ensure that the sponsor can anticipate questions and queries from authorities derived from such analysis and can develop adequate response strategies or consider mitigation activities already during trial conduct.

The approach shown above demonstrates the use of standard variables common to most clinical trials, but the approach is not limited to such parameters. The time and effort needed to add additional parameters to the data sets described above is low. This would allow the addition of specific datasets to the analysis (e.g. primary endpoint data) without significant increase in cost.

6. Further uses of SALSA in risk-based Quality Assurance

SALSA is a versatile tool that can be utilized by sponsor's risk-based QA approach in multiple ways – basically any time information on data quality and integrity is needed to support decisions on what sites to prioritize for audit.

Comparing SALSA results from multiple sites gives an indication of the relative data quality and integrity. This information can be utilized by sponsor's QA as a supplemental tool in their risk assessments and can support their decisions to make changes to their annual audit plans in exceptional circumstances. The ongoing COVID-19 pandemic and the accompanying restrictions on travel and on-site visits are just one example of such circumstances. The decision on how to handle each individual audit, whether it be postponement, removal from the audit plan altogether, or conduct of a full or abbreviated remote site audit must be justified and documented by QA. As the reason(s) for selecting the site for inclusion in the annual audit plan in the first place (e.g. high enrolment, inexperienced site, high monitor turnover, etc.) will probably not have changed since the inclusion of the audit in the audit plan, the decision to modify the original plan may be additionally backed-up by evidence of acceptable data quality and integrity at the site. SALSA can be used to review and compare the data from several sites to prioritize those with potential issues. Based on the analysis, some sites may be prioritized, some may be taken off the schedule altogether, some may be selected for deferment to a later time point, and some may be considered good candidates for a full or abbreviated remote audit.

Once these decisions have been made, SALSA can also be used as part of a risk-based, stepwise approach to remote auditing which aims to reduce the burden caused by the audit to the site staff. A “stepwise RISA” begins with a remote review of the site’s documentation and data, without involvement of the site, and continues with interviews with CRO and site only if and to the extent necessary to clarify issues that could not be resolved during document/data review. Utilizing SALSA reports during this first step yields a more in-depth look at the data and helps to pinpoint questions to be resolved. The general indication of data quality and integrity provided by SALSA can, along with other signals, give a level of confidence in the GCP compliance of the site, without requiring conduct of a full on-site or remote audit in circumstances in which these are not possible.

Please also refer to GXP Engaged’s strategy paper: *A Risk-Based Approach to Auditing Implementing RISA and SALSA* for further information on these innovative approaches.