Paper AS19

Data Diaries Don’t Lie: Stats That Spot ePRO Risks

Nechama Katan, Wicked Problem Wizards, Longmeadow MA, USA
Steven Gilbert, Pfizer, Cambridge MA, USA

ABSTRACT

Electronic Patient-Reported Outcomes (ePRO) are increasingly used in clinical trials to support key efficacy and safety endpoints. Because ePRO data are entered directly by patients, traditional approaches such as record-level review and query management are often not appropriate to confirm that the data are fit for purpose. Sponsors remain responsible for ensuring data integrity, and audit trail metadata provide a practical way to evaluate how ePRO data are generated and whether observed behavior is consistent with independent patient entry.

This paper describes statistical methods for using ePRO audit trails to identify data integrity risk signals, with emphasis on timing- and behavior-based metadata such as open and close times, device identifiers, and masked PIN codes. Common risk scenarios include delayed or clustered entries, unusually fast completion times, data entered at unusual hours, and patterns that suggest site staff may be entering data for patients. Because audit trail data are high volume, highly skewed, and may include both positive and negative timing values, the paper highlights the need for data normalization and visualization before applying outlier detection or monitoring methods.

Building on industry guidance for Audit Trail Review Analytics (ATRA), the paper presents a maturity-based framework that progresses from descriptive statistics and visualization to time-trended monitoring with statistical limits, then to advanced analytics for multi-metric site prioritization, and finally to statistical root cause analysis. Worked examples and figures illustrate how these methods help distinguish meaningful signals from expected operational variability and support risk-proportionate follow-up.

INTRODUCTION

As reliance on ePRO data grows, so does the need for confidence that these data accurately reflect independent patient input and are fit for their intended purpose. Unlike traditional site-entered data, ePRO data are generated through frequent, patient-driven interactions with electronic devices. These interactions create unique data integrity risks that are not always visible through standard data review approaches. Examples include delayed or clustered entries, implausibly short completion times, and patterns that suggest data may be entered by site staff rather than patients. While individual events may be explainable, consistent behavioral patterns across patients or overtime can raise concerns if not identified early.

Audit trail metadata provide a detailed record of how ePRO data are generated, including timing, user context, and device-related information. When analyzed systematically, these metadata offer valuable insight into patient behavior, site workflows, and study conduct. However, the volume, structure, and statistical properties of audit trail data make manual review impractical and simple summary statistics insufficient.

This paper focuses on statistical and analytical methods for evaluating ePRO audit trail data in a scalable and risk-proportionate manner. Building on industry guidance for Audit Trail Review Analytics (ATRA), it demonstrates how data normalization, visualization, time-trended monitoring, and advanced analytics can be applied to distinguish meaningful signals from expected operational variability. The goal is not to detect misconduct, but to support earlier identification, prioritization, and investigation of ePRO-related risks consistent with modern risk-based quality management principles.

REGULATORY AND QUALITY CONTEXT

ICH E6(R3) calls for planned, risk-based review of trial data, audit trails, and related metadata. The amount and type of review should match the risks of the study. In parallel, ICH E8(R1) emphasizes Quality by Design and the need to proactively manage factors that are critical to data quality.

For studies that rely on ePRO as one of the primary endpoints, data integrity cannot be assessed only by checking whether data are present or complete. It must also consider how, when, and where patient-reported data are entered and modified. ePRO data are entered directly by patients, and sponsors cannot query patients in the same way they can query site-entered data. However, this does not remove the sponsor's responsibility to ensure that the data are fit for purpose.

In the authors' experience, the most effective way to assess whether ePRO data are fit for purpose is through analysis of audit trail metadata. Useful audit trail elements for ePRO include diary open and close times, device identifiers, and masked PIN codes. PIN codes should be pseudorandomized or masked so they are not visible to the sponsor. As with all audit trails, records of data changes, reasons for change, and the user who made the change are also required.

In practice, ePRO data values are rarely changed. As a result, greater insight is often gained by analyzing behavioral metadata such as open and close times rather than focusing only on data modifications. Audit trail data provide critical visibility into how ePRO data are generated and how study processes are being followed at the site level.

This paper does not restate regulatory guidance in detail. Instead, it focuses on analytical methods that support proportional, defensible oversight consistent with these regulatory expectations.

EPRO RISK SCENARIOS AND AUDIT TRAIL SIGNALS

ePRO systems generate audit trails that record timestamps, user context, and event types for each interaction. These metadata provide insight into how ePRO data are entered, not just what values are recorded. This information makes it possible to identify patient behavior and system usage patterns that are not visible in traditional datasets.

Frequently observed ePRO-related risk patterns include, but are not limited to, situations where data appear to be entered by site staff rather than by the patient. These patterns may be identified through audit trail indicators such as:

- Time between diary open and save events (how long it took to enter the data)
- Close to Open time The time between a patient closing one assessment and the next patient opening the next This can be negative if there is overlap.
- Device identifiers and PIN codes
- Data entered at unusual times (e.g., outside expected patient waking hours)
- Data entered unusually quickly or with highly similar timing across patients

These patterns may reflect issues with patient burden, site workflows, training practices, or system configuration rather than intentional misconduct. For this reason, audit trail signals should be interpreted as indicators of risk that require further evaluation, not as conclusions on their own.

LIMITATIONS OF MANUAL REVIEW AND MANUAL DATA EXTRACTION

Audit trail data are large by design. As more data are collected in ePRO and similar systems, the volume of audit trail data continues to grow. Traditional review approaches often assume that patient-entered data cannot be queried and therefore do not require systematic review. However, practical experience has shown that meaningful analysis of ePRO audit trails is both necessary and feasible.

Manual review and manual data extraction do not scale for this type of data. Effective analysis requires automated data loading, where possible, and analytical tools beyond spreadsheets such as Excel. It is also important to account for technical factors, such as timestamps recorded in local time for patients or sites, which must be standardized before analysis.

Before visualization or modeling, ePRO audit trail data should be normalized. Normalization supports clearer visualization and allows behavioral patterns to be compared across patients, sites, and time periods. Graphical review is a critical first step in understanding these data and should be performed whenever possible.

Although this may appear simplistic for an experienced analytics audience, many important issues become visible only through visualization. The data presented in this paper are synthetic but are derived from real examples of risks observed in clinical trials.

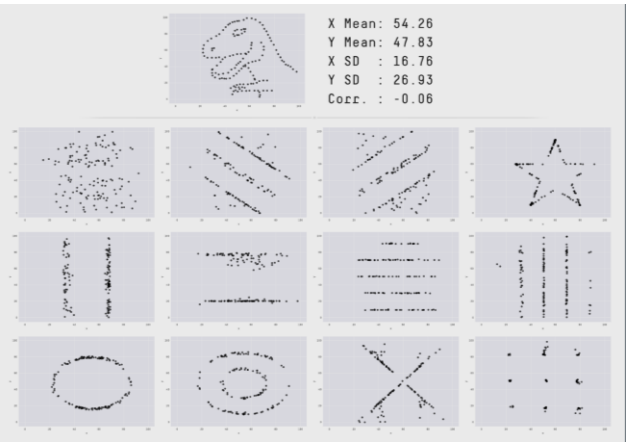

Figure 1 The Datasaurus Dozen, Matejka & Fitzmaurice (2008)

Figure 1 illustrates the importance of data visualization using the Datasaurus Dozen. All 13 datasets shown have identical descriptive statistics, yet their visual patterns are clearly different. Any analytical approach used for ePRO audit trail review should therefore support data visualization as a core capability.

PREPARING EPRO AUDIT TRAIL DATA FOR ANALYSIS

Before applying analytical methods, it is important to understand the limitations of raw ePRO audit trail data and the need for appropriate data normalization. Many audit trail metrics have wide ranges, skewed distributions, or structural constraints that can obscure meaningful patterns if analyzed directly.

OPEN-TO-SAVE TIME

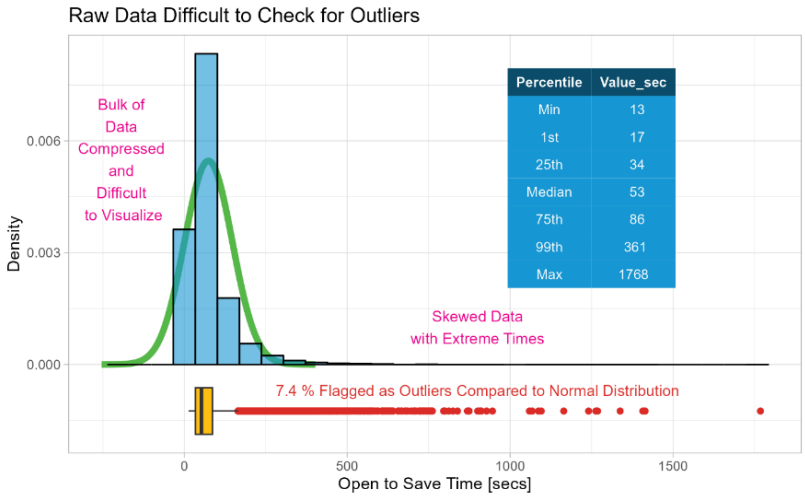

Figure 2 Open to Save times raw

Figure 2 shows the raw distribution of *open-to-save time* for an ePRO device. This metric represents the amount of time a questionnaire remains open while the patient enters responses. In raw form, the data span a very wide range of values, making visualization and outlier detection difficult.

Standard outlier detection methods, such as boxplots, are typically calibrated for approximately normal distributions. Almost all ePRO timing metrics are not normally distributed and should not be expected to be. As a result, direct application of these methods can either over-flag or under-flag potential issues.

To address this, a shifted log-normal transformation was applied. Reaction time data can be modeled with a shifted log-normal distribution, making this approach appropriate for open-to-save times. In this example, a transformation of the form:

$$\log(Y - 11)$$

was used, where 11 seconds represents the minimum plausible time required to open and close the questionnaire. This transformation compresses extreme values while preserving relative differences and allows unusually short completion times to remain visible for further evaluation.

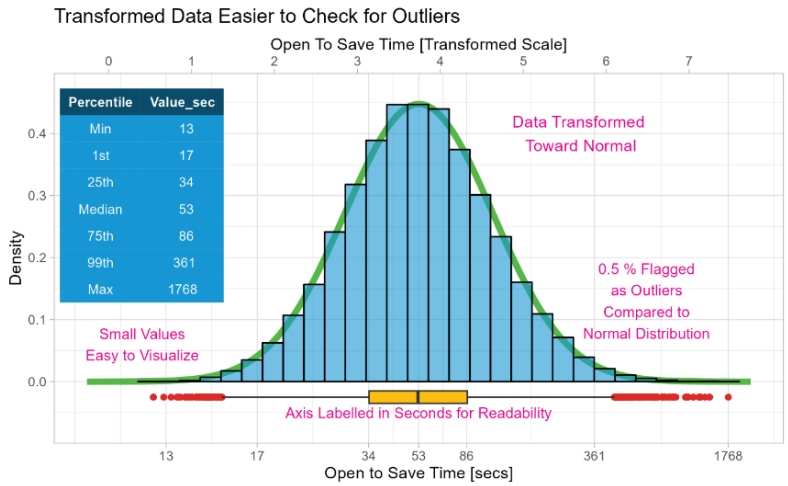

Figure 3: Log Normal Transformation

Figure 3 shows the transformed data. After normalization, the distribution is more centralized, and outliers are more clearly separated, supporting both visualization and downstream analysis.

CLOSE-TO-OPEN TIME BETWEEN DEVICES

A second timing metric of interest is *close-to-open time*, defined as the time between one subject closing a questionnaire and another subject opening theirs at the same site. Under typical conditions, where patients complete assessments independently at home, this metric is expected to show a mix of positive and negative values, reflecting overlapping completion times.

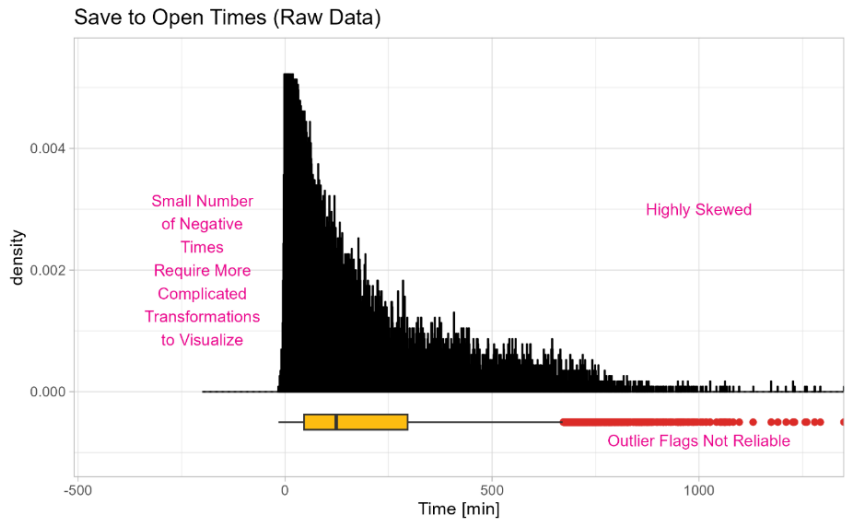

Figure 4 Closed to Open Times

In contrast, sites where staff assist multiple patients in completing ePRO assessments often show highly consistent and unusually regular gaps between close and open events. This pattern has been observed across multiple trials and sponsors.

The raw distribution of close-to-open times is typically highly skewed, particularly at sites with few patients or without fixed assessment windows. Overlapping assessments produce negative values, which cannot be transformed using standard methods such as logarithmic or square-root transformations.

To address this, a pseudo-log transformation was applied, which supports both positive and negative values. After transformation, previously hidden structure becomes visible. In this example, a small negative peak and a positive peak around two hours can be clearly observed. Very short gaps (for example, less than five minutes) are uncommon and provide a useful starting point for deeper investigation.

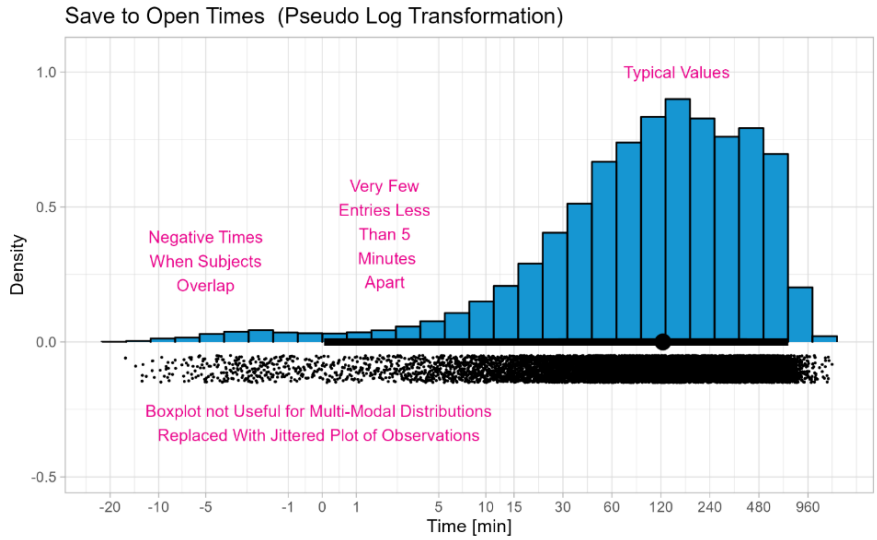

Figure 5 Pseudo Log Transformation

In **Fgure 5**, data are shown using both a jittered scatter plot and a boxplot. The median and interquartile range are highlighted to provide additional context. Negative values are naturally bounded by the time required to complete the assessment, which in this example does not exceed approximately 20 minutes.

WHY DATA NORMALIZATION MATTERS

These examples demonstrate that raw ePRO audit trail data are rarely suitable for direct analysis. Timing metrics often have structural limits, skewed distributions, and mixed positive and negative values. Without appropriate normalization, important behavioral patterns may remain hidden or be misinterpreted.

Data normalization is therefore a critical step in ePRO audit trail analysis. Appropriate transformations improve visualization, support more reliable outlier detection, and enable meaningful comparison across patients, sites, and studies. While no single transformation is suitable for all metrics, thoughtful normalization, guided by the nature of the data, provides a necessary foundation for the analytical methods described in the following sections.

ANALYTICS MATURITY MODEL FOR EPRO AUDIT TRAIL REVIEW

EPRO audit trail data are high volume, granular, and highly context dependent. Extracting actionable insight from these data requires statistical methods that scale beyond static summaries and manual review. This paper adopts a maturity-based analytics framework aligned with Audit Trail Review Analytics (ATRA), as described in the eClinical Forum ATR position paper (2021 and second paper February 2026). The framework is organized into four analytical capability levels:

- Level 1 – Descriptive Statistics and role of visualizations
- Level 2 – Time-Trend and Threshold Monitoring
- Level 3 – Advanced Analytics and Risk Prioritization
- Level 4 – Root Cause Analysis and Feedback

These levels represent analytical capabilities rather than sequential steps. Depending on study risk and data availability, multiple levels may be applied in parallel.

To illustrate how this framework operates in practice, the following case study focuses on one common ePRO risk scenario: data entered at unexpected times.

CASE STUDY: DATA ENTERED AT UNEXPECTED TIMES

As clinical trials increasingly rely on ePRO data to inform primary endpoints, patient compliance has become a key operational metric. Sites are often measured and compared based on how consistently patients complete required assessments. In some situations, pressure to maintain high compliance can lead sites to intervene by entering data on behalf of patients, either for all participants or for those who are struggling.

This behavior cannot be easily detected by reviewing individual records in isolation. It becomes visible only when the aggregate behavior of all patients at a site is examined over time. Audit trail metadata, particularly timestamps, provide the necessary context to identify these patterns.

In many protocols, patients are given reminders or alarms to prompt diary completion. Under normal conditions, this results in some clustering of entries around expected times, but with natural variability across patients and days. In contrast, site-entered data often exhibit unusually consistent timing patterns, such as repeated entries during clinic hours or tightly clustered completion times across multiple patients.

The following sections apply the maturity-based analytics framework to this scenario, demonstrating how descriptive summaries, time-based monitoring, and comparative analytics can be used to identify, prioritize, and investigate these behaviors in a proportionate and defensible manner.

DESCRIPTIVE STATISTICS AND THE ROLE OF VISUALIZATION

Several descriptive metrics are useful for evaluating ePRO data entry behavior at the site level. In general, the goal is to identify patterns that appear too regular and not consistent with natural, independent patient behavior.

OPEN-TO-CLOSE TIME

Open-to-close time measures how long a questionnaire remains open for each patient. This metric provides insight into how much time is spent completing the ePRO. Extremely short completion times may indicate that responses are being entered without adequate consideration, while very long completion times may suggest interruptions, or usability issues. While both extremes can represent potential data quality concerns, short entry time is more of a concern.

CLOSE-TO-OPEN TIME

Close-to-open time measures the interval between one ePRO questionnaire being closed and the next one being opened at the same site. This metric is useful for identifying non-overlapping or highly regular data entry patterns. When patients enter data independently, close-to-open times are typically irregular and may include negative values when entries overlap. As the number of subjects at a site increases, overlapping entries become more likely, and additional variability is expected. These features represent normal operational noise.

Figure 6 shows an example of close-to-open time distributions consistent with independent patient behavior. The variability reflects differences in subject schedules, entry timing, and occasional overlap between assessments.

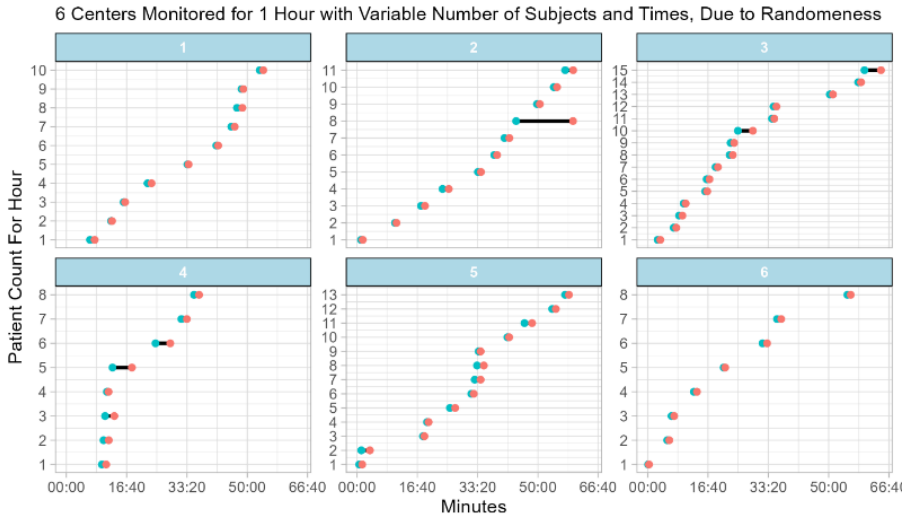

Figure 6 closed to open distributions

In contrast, **Figure 7** shows a pattern that is too regular to be explained by independent patient activity. Evenly spaced intervals and the absence of overlap suggest systematic data entry, such as site staff entering data sequentially for multiple patients. These patterns can be more difficult to detect when sites enter data for only a subset of patients rather than for all participants.

Reviewing audit trail data from studies with confirmed issues provides important context for identifying similar patterns in future studies. Comparative visualization across sites remains a key tool for distinguishing expected variability from behavior that warrants further investigation.

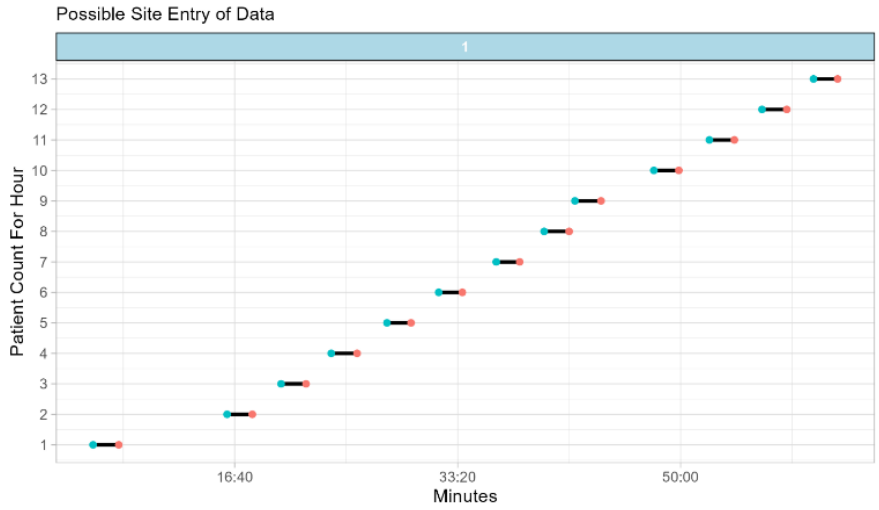

Figure 7 Too Regular Closed to Open

OPEN-TO-OPEN TIME

Open-to-open time measures the time gap between consecutive ePRO entries at a site. The short times are often indication of patients entering data after an alert. At sites where patients enter data independently, this distribution shows irregular and variable gaps. In contrast, sites that enter data on behalf of patients often display unusually consistent gaps, such as repeated intervals of approximately 2 minutes, see **Figure 8 open to open.** In addition to the average gap, the variation in these intervals is an important indicator and should be monitored over time.

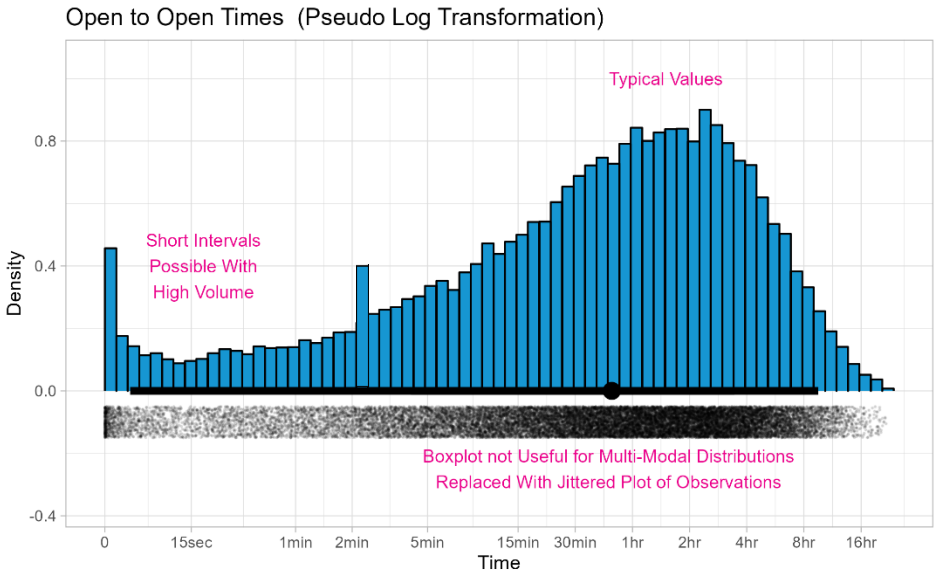

Figure 8: Open to Open time gaps with an unusual site interval at 2 min

Such spikes in the patterns are consistent with situations where site staff enter data for patients, either for all participants or for a subset of patients who have difficulty completing assessments independently, such as elderly patients. Regardless of the underlying reason, a normalized distribution makes these unusual patterns visible and allows them to be flagged for further

investigation. These behaviors are often masked when relying on simple descriptive statistics alone, reinforcing the need for appropriate transformation and visualization.

OPEN TIME

Open time reflects when a questionnaire is opened relative to the reminder alarm. Under normal conditions, patient entries cluster loosely around the alarm time, with natural variation before and after the reminder. Some patients enter data early, while others respond later. This variability is expected when patients complete assessments independently. When data are entered by site staff, different patterns may emerge, such as a shifted start time or a longer tail of entries, consistent with sequential entry for multiple patients.

These patterns are best understood through visualization rather than summary statistics alone. **Figure 9** shows an example of open time measured in minutes after the hour. In this example, data entry at Site 1 (shown at the bottom) appears unusually regular, with entries occurring at nearly fixed intervals.

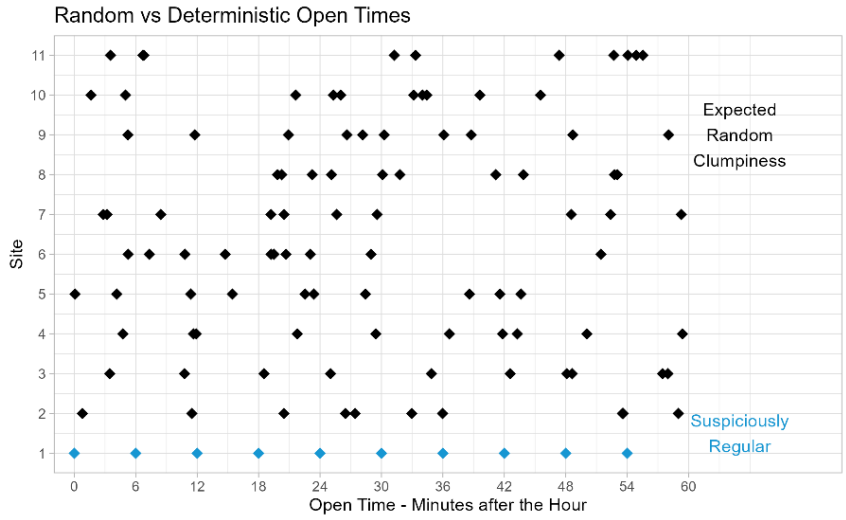

Figure 9 example of Open Time in minutes after the hour

Random behavior does not produce evenly spaced patterns. A useful analogy is raindrops falling on the ground rather than marks placed at fixed intervals on a ruler. Overly regular timing, such as entries occurring every six minutes, may indicate that data are being entered systematically by site staff rather than independently by patients.

Descriptive statistics are a necessary first step for evaluating ePRO data entry behavior, but they are not sufficient on their own. Metrics such as open-to-close, close-to-open, open-to-open, and open time help characterize how data are entered and highlight patterns that may be inconsistent with independent patient behavior.

Across these metrics, the key signal of interest is over-regularity. Independent patient activity is naturally variable, while site-entered or assisted data often appear unusually consistent. Visualization, combined with appropriate normalization, is therefore essential to reveal timing structure and patterns that would be masked by simple summary statistics alone.

TIME TRENDS WITH STATISTICAL LIMITS

Descriptive statistics are useful for understanding overall behavior, but they lack the detail needed to show when changes occur. Time-trended analysis with statistical limits adds this context by revealing how behavior evolves during the study. This approach requires additional aggregation of the data across time.

As shown in earlier sections, lack of variability can be an indicator of potentially problematic behavior. For time-trend analysis, a measure of variability is therefore required. In this example, range was used rather than standard deviation to reduce sensitivity to differences in site size and to limit noise from sites with few observations.

A key question is how to define *typical behavior*. This requires the use of statistical limits. Limits may be derived from completed studies, from early pooled study data, or adjusted dynamically as more data are collected. Adaptive limits have the advantage of being wider during study startup and narrowing as sites stabilize. For clarity, the examples shown here use fixed limits representative of mature site behavior.

For open-to-open time, the average gap is not always the most informative metric for time-trended analysis. Instead, the variation in open-to-open times is more useful, particularly because unusually low variability, rather than high variability, may indicate site-driven data entry. Since the objective is to assess site behavior rather than individual patient behavior, site-level timing is evaluated.

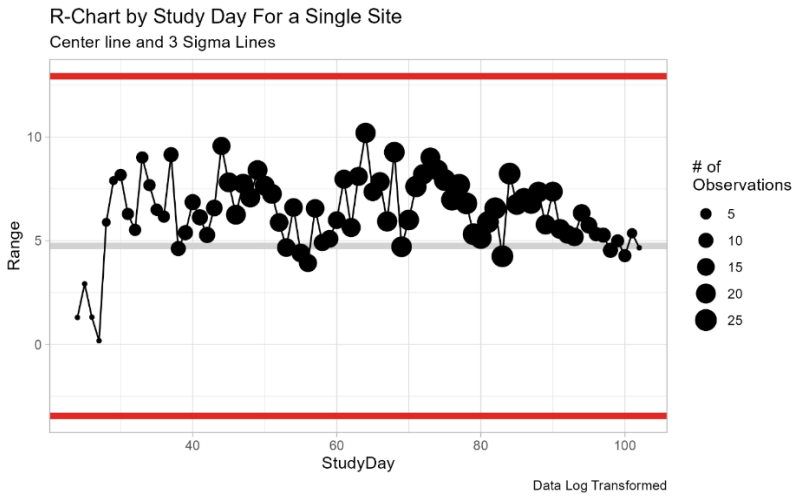

Figure 10 Random Range of Open to Open time

In the case of **Figure 10**, attention is focused on low variability. A three-sigma control limit is applied, with only the lower limit considered relevant. High variability is generally not interpreted as a site behavior issue, except in specific protocol designs. The period of unusually low variability shown in the chart would not be apparent from descriptive statistics alone.

The data are log-transformed due to strong skewness, which is well described by a gamma distribution. An R chart, using Range as the aggregated value, is used rather than a standard deviation-based chart because even large sites may have very few observations on a given study day.

Figure 11, A second example illustrates a site with suspicious behavior. In some cases, site behavior changes over time—for example, a site may initially perform as expected and later begin assisting patients with data entry. Time-trend analysis is particularly powerful in these situations because it allows identification of whether an issue is persistent, historical, or newly emerging. This level of insight cannot be achieved through overall descriptive statistics alone.

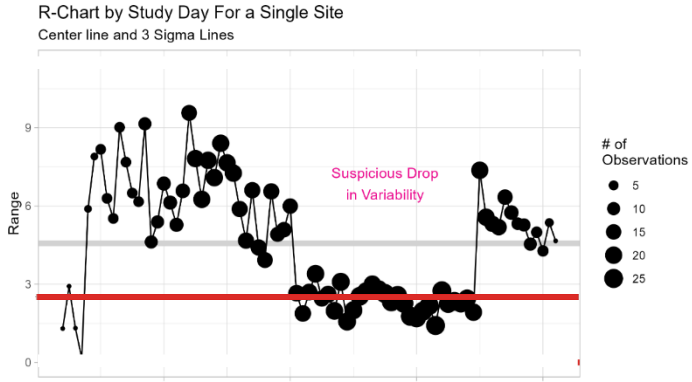

Figure 11 R-Chart by Study data Open to Open

ADVANCED ANALYTICS

Advanced analytics help define what expected behavior looks like based on historical data, what is observed in current studies, and how to identify behavior that is meaningfully different. This provides a consistent, data-driven way to distinguish normal variation from potential risk.

ePRO systems generate many metrics, making it necessary to evaluate site behavior across multiple dimensions rather than one metric at a time. An overall difference score can be used to identify sites that behave unusually across several measures. This approach makes oversight manageable, as it is not feasible to review large numbers of metrics across many sites manually.

Different analytical methods may highlight different signals. A practical strategy is to apply multiple analyses and then focus on sites that consistently appear unusual across methods. Simple steps, such as removing highly correlated metrics to prevent one group of measures from dominating results, can improve interpretability. More advanced techniques, such as clustering analysis, may also be applied but are outside the scope of this paper.

STATISTICAL ROOT CAUSE ANALYSIS

Dashboards and initial analyses can identify potential issues, but they do not explain why those issues occur. Statistical root cause analysis is used to explore and evaluate plausible explanations for observed differences in behavior.

This step often requires broader analysis of the data. For example, patterns may vary by day of the week if sites periodically enter data on behalf of patients rather than waiting for patient entry. Differences may also be associated with device type, such as site-provided devices versus bring-your-own-device (BYOD), or with patient characteristics such as age or other demographics. Investigating these possibilities requires returning to the raw audit trail data, integrating multiple data sources, and performing analyses that account for covariates and other sources of variability.

While statistical root cause analysis can help rule out alternative explanations, it cannot establish causation. Statistical findings must therefore be interpreted with caution. The only way to determine what actually occurred at a site is through direct follow-up, such as discussion with the study team and, when necessary, targeted site investigation.

Once corrective actions are taken, trended data should be reviewed to confirm that behavior has improved and that the issue does not recur. This feedback loop is essential for demonstrating that identified risks have been effectively addressed.

OPERATIONAL USE OF ANALYTICS OUTPUTS

Analytical insight alone does not improve study quality unless it is integrated into routine oversight workflows. For analytics outputs to be actionable, they must be interpretable, reproducible, and aligned with existing monitoring and issue management processes.

Key principles include consistent metric definitions, pre-specified thresholds where feasible, and clear traceability from summary metrics back to the underlying audit trail records. These elements support transparency and allow findings to be reviewed, explained, and reproduced when needed.

Analytics should be used to prioritize review and guide decision-making rather than to replace clinical or operational judgment. Outputs are most effective when embedded within established workflows, such as central monitoring reviews, risk review meetings, and issue escalation processes. In this way, audit trail analytics function as an input to risk-based oversight rather than as isolated analyses.

Embedding analytics into routine workflows ensures that signals are assessed consistently, actions are documented, and follow-up can be tracked over time. This integration supports scalable oversight while maintaining clear accountability and regulatory defensibility.

CONCLUSION

As ePRO data are increasingly used to support primary and secondary endpoints, ensuring their integrity requires more than traditional record-level review. Audit trail metadata provide critical insight into how patient-reported data are generated and whether observed behaviors are consistent with independent patient entry. However, these metadata are high volume, highly skewed, and structurally constrained. Without appropriate data normalization and visualization, meaningful behavioral patterns may be hidden or misinterpreted.

This paper demonstrates that a maturity-based analytics framework, combined with thoughtful data normalization, enables scalable and defensible oversight of ePRO data. Descriptive statistics and visualization establish baseline behavior, time-trended analysis with statistical limits identifies when behavior changes, and advanced analytics support prioritization across multiple metrics and sites. Normalization is a prerequisite at each stage, allowing timing metrics with wide ranges, skewed distributions, and mixed positive and negative values to be compared meaningfully across patients, sites, and studies.

Analytics alone do not replace operational follow-up or clinical judgment. Statistical methods help identify and prioritize potential risks, but confirmation and resolution require engagement with study teams and, when necessary, direct site investigation. When normalized audit trail data and analytics are embedded into routine oversight workflows, organizations can move beyond reactive review toward proactive, risk-proportionate management of ePRO data, strengthening confidence in patient-reported outcomes and supporting regulatory expectations throughout the clinical trial lifecycle.

REFERENCES

Justin Matejka and George Fitzmaurice. 2017. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1290-1294.

ChatGPT (OpenAI, Version GPT-5.2) was used to improve grammar and sentence structure in the paper. The authors reviewed and are fully responsible for the final content.

eClinical Forum. *Audit Trail Review: Key Tool to Ensure Data Integrity*. Version 1.0. Public Release. eClinical Forum; 2021. Available at: https://eclinicalforum.org/downloads/public-release-audit-trail-review-key-tool-to-ensure-data-integrity-version-10

eClinical Forum. *Audit Trail Review: Strategies for Implementing Audit Trail Review*. Industry Position Paper Public Release eClinical Forum; 2026.

CONTACT INFORMATION

Contact the authors at:

Author Name: Nechama Katan
Company: Wicked Problem Wizards
Email: nechama@wickedproblemwizards.com
Website: www.wickedproblemwizards.com
Linked In: linkedin.com/in/nechama

Author Name: Steven Gilbert
Company: Pfizer
Email: steven.a.gilbert@pfizer.com
Website: www.Pfizer.com
Linked In: **linkedin.com/in/steven-gilbert-110913b**