

# Reimagining Clinical Data Exploration: Animated Data Visualizations for Interactive Clinical Trial Storytelling

Tony Mathew, Pfizer Healthcare India Pvt Ltd, Chennai, India

## ABSTRACT

As clinical trials become more complex and data-rich, the way we explore, and present data must evolve beyond static summaries. Traditional outputs like tables, listings, and static plots are powerful for documentation and audit trails but they often fall short in conveying the full dynamics of a trial. The human brain is wired to understand patterns over time and across dimensions, yet our tools rarely harness that strength. There is an opportunity to bridge this gap by transforming clinical data into animated, interactive visual narratives.

This project aims to develop an animated data visualization platform, powered by Python and delivered through a Streamlit interface. The core objective is to convert multi-dimensional, longitudinal data into motion-based visuals that allow clinical teams, statisticians, reviewers, and even non-technical stakeholders to grasp complex trends more intuitively. Instead of seeing a single snapshot, users will be able to observe how data evolves across visits, treatment arms, or patient subgroups—revealing emerging trends, temporal shifts, and unexpected outliers that may otherwise remain hidden in static summaries. Simple yet powerful animated graphs, such as line plots showing lab values over time, bar charts comparing treatment arms across visits, and scatter plots for patient-level changes are some of the examples that can be used to demonstrate changes and patterns more clearly.

Ultimately, this approach seeks to redefine how clinical data is consumed—not just as a record of what happened, but as a living, visual story of the trial itself. By aligning statistical rigor with modern visualization, we can support faster, deeper, and more collaborative data understanding across the clinical development lifecycle.

## INTRODUCTION

Historically, the exploration of clinical trial data has relied on static tables, listings, and figures, which limit reviewers' ability to easily understand how data evolves across different study milestones, interim updates, or iterations of derivation. As datasets progress through multiple transfers and transformations, small changes in distributions, missing data, and overall structure can remain undetected until later stages of validation.

Recent developments in interactive and animated data visualization present a chance to rethink how clinical data is examined and conveyed. By incorporating movement, temporal sequencing, and interactivity, the evolution of complex datasets can be rendered into a clear visual story—facilitating quicker insight generation, earlier identification of anomalies, and enhanced communication across functions.

This paper introduces an animated, interactive dashboard aimed at showcasing the progression of clinical datasets across different versions, allowing stakeholders to observe changes in data over time instead of deducing alterations from fixed outputs.

## RELATED WORK AND BACKGROUND

Clinical trials generate large volumes of complex, heterogeneous, and longitudinal data. Traditionally, clinical data exploration and reporting rely on tables, listings, and static plots. While these outputs are essential for regulatory compliance and auditability, they provide limited support for exploratory understanding, particularly for temporal trends, subgroup variability, and multivariate relationships. As clinical datasets grow in scale and complexity, visualization has increasingly become central to clinical data exploration and insight discovery. Early clinical visualization largely focused on patient-level summaries and static timelines, which were helpful for individual review but lacked scalability and flexibility for cohort-level longitudinal exploration.

Interactive visualization has advanced clinical data review by enabling dynamic filtering, drill-down analysis, and linked views, allowing users to move seamlessly between high-level summaries and patient-level detail. A scoping review in the *Journal of Medical Internet Research* reports increasing adoption of such tools for trend exploration, cohort comparison, and hypothesis generation in healthcare research. Since clinical outcomes evolve over time, temporal visualization is particularly important. Scheer et al. found common use of line plots, timelines, scatter plots, and cohort

comparison views, though many tools still lack sufficient interactivity for robust longitudinal comparison. Empirical findings further show that time-oriented visual representations enable richer insight generation than tabular outputs alone.

More recent work has moved toward visual analytics systems that integrate analytical models with interactive visualization for deeper exploration. DPV combines hidden Markov models with interactive interfaces to explore disease progression pathways and compare patient subgroups, while Trial View introduces AI-supported visual analytics for temporal event exploration in clinical trials. Animation also plays a key role in representing temporal change: Heer and Robertson demonstrated that animated transitions improve change tracking and perceptual continuity compared to static transitions, making them well suited for longitudinal trend understanding. Overall, literature suggests that interactive, temporal, and animated visualization improves insight discovery in complex clinical trial data; however, such approaches remain limited in regulated clinical workflows, motivating platforms that complement traditional statistical outputs with richer exploratory and narrative analysis.

## **MOTIVATION AND SOLUTION OVERVIEW**

Typical review activities involve assessing the effects of updates in derivation, validating that record counts and critical analysis variables function as anticipated, and making sure that no unexpected structural or distributional alterations have been introduced. These activities are often aided by static summaries and pairwise comparisons, which can limit insights into temporal trends and frequently necessitate extensive manual interpretation when multiple versions are considered.

To fill this operational gap, we created an interactive dashboard using Streamlit that facilitates version-aware, animated exploration of clinical trial datasets. Instead of substituting current QC processes, the dashboard is designed to enhance them by allowing swift, exploratory examination of dataset changes before or during formal validation.

The dashboard processes several versions of SDTM or ADaM datasets provided in CSV or Excel formats and offers five targeted analytical perspectives that align with typical programming review requirements.

- Animated variable-level visualizations to assess distributional changes
- Dataset-level metric trends to monitor record counts and missingness
- PCA-based multivariate projections to detect global structural shifts
- Automated summaries of column- and row-level changes across versions
- Metadata-level compliance checks to ensure consistency across dataset versions
- Comparative descriptive statistics to support interim and pre-lock review

This solution is agnostic to the study, adaptable to various formats, and ready for export, allowing animated visualizations to be saved as GIF or MP4 files for use in review meetings, documentation, or interdisciplinary discussions. By focusing on visual consistency and contextual changes over time, this method enables statistical programmers to detect potential problems sooner and communicate changes in datasets more effectively, all while maintaining the simplicity of current workflows.

## **METHODOLOGY AND SYSTEM DESIGN**

An interactive dashboard was developed using a modular framework to visualize the evolution of datasets across different versions. This solution automatically tracks versions of multiple clinical datasets in formats such as CSV, Excel, sas7bdat, and rds, representing consecutive versions of the same domain or analysis dataset. Each dataset version is automatically assigned a unique version identifier, enabling structured temporal analysis

The system integrates:

- Streamlit for the interactive user interface
- Plotly for animated and interactive visualizations
- Scikit-learn for multivariate analysis, including Principal Component Analysis (PCA)
- Image export utilities to generate GIF and MP4 animations for offline use

## **ANIMATED DATA EXPLORATION AND QUALITY ASSESSMENT**

The suggested method positions animation as a key analytical element for reviewing clinical datasets. Instead of considering dataset versions as separate snapshots, subsequent versions are represented as a continuous timeline, facilitating a temporal examination of how the data's structure and content develop through updates and cuts over time. This perspective enables reviewers to observe changes directly rather than deducing them from static, side-by-side comparisons.

## VARIABLE-LEVEL ANIMATED EXPLORATION

At the variable level, we analyze distributions and relationships across different dataset iterations using animated visual displays. Each frame of the animation represents a particular dataset version, while the axis scales remain consistent to ensure comparability. This method allows for a clear visual evaluation of variations in central tendency, dispersion, data density, and category composition over time.

By viewing distributional patterns as a continuous progression, reviewers can more easily spot gradual shifts, sudden changes, or unforeseen variations caused by derivation logic or updates in upstream data. This technique is especially useful for evaluating essential analysis variables and categorical groupings that are important for subsequent analyses.

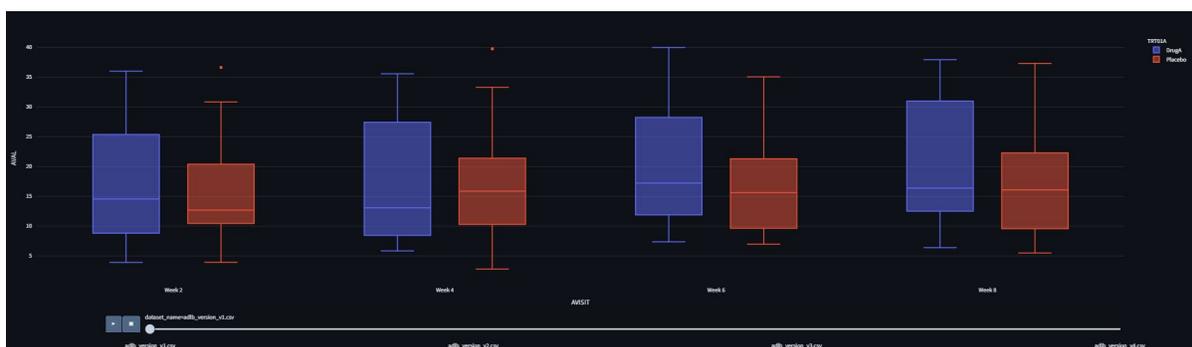


Figure 1: Variable- Level Animated Exploration

## DATASET-LEVEL TREND ASSESSMENT

In addition to exploring variable-level details, characteristics of the dataset are compiled and analyzed across different versions to give an overarching perspective on the dataset's development and reliability. Metrics including record counts, distinct identifier counts, patterns of missing data, and specific summary statistics are collected for each version and assessed within a temporal framework.

Visualizing these metrics across versions reveals abrupt changes that could be missed in static summaries. This approach facilitates the early detection of potential issues related to filtering, joins, or data ingestion, and aids in the swift validation of expected trends in dataset growth or stabilization across intermediate stages.



Figure 2: Dataset- Level trend Assessment

## MULTIVARIATE STRUCTURAL EVOLUTION

To evaluate changes involving multiple variables at once, the method employs multivariate dimensionality reduction through Principal Component Analysis (PCA). Numeric variables are standardized, and missing data is addressed via imputation to maintain consistency across different dataset versions. The PCA projections are then analyzed across these dataset versions as a developing structure rather than a fixed projection.

This multivariate approach allows for the identification of overall structural changes that might not be evident from individual variable assessments alone. Automated summaries detailing the variance accounted for, contributions of

variables, and overall dataset movement within the reduced space offer further insights for discerning whether the observed changes are anticipated or potentially concerning.

## **AUTOMATED COMPARATIVE QUALITY ASSESSMENT**

Alongside exploratory analysis, this method incorporates automated comparative evaluations to aid in formal quality assessment. Successive iterations of the dataset are analyzed to detect structural modifications, such as the addition or removal of variables, as well as changes at the row level based on user-specified key identifiers.

These analyses yield succinct summaries that capture how datasets evolve, enhancing traceability and minimizing dependence on manual reconciliation. By merging automated comparison with animated exploration, this method promotes both systematic validation and a clear understanding of how datasets change.

Comparative descriptive summaries additionally provide context for the changes observed by measuring differences in key statistics between versions. Collectively, these outputs assist in performing sanity checks during interim reviews and enable effective communication about dataset behavior among statistical programmers, statisticians, and clinical reviewers.

## **COMPARATIVE STATISTICAL INSIGHTS ACROSS DATASET VERSIONS**

In addition to structural and exploratory analysis, the suggested method facilitates a comparative statistical evaluation of numerical variables across different versions of the dataset. For chosen metrics, descriptive statistics are generated for each version, allowing for a methodical comparison of summary behavior as the data changes with interim updates or derivation modifications.

Analyzing trends in essential statistics like means provides an extra layer of validation that enhances both visual analysis and automated quality control checks. These summaries aid in regular sanity checks by enabling reviewers to verify that the changes observed align with expectations stemming from increased data volume, modifications in derivation logic, or updates from upstream data. Significant shifts in summary statistics between versions may suggest broader alterations in data composition or derivation processes that require additional scrutiny.

By showcasing these comparisons with an awareness of version differences, this method enhances clear and high-level communication regarding dataset behavior among statistical programmers, statisticians, and clinical reviewers, especially during interim analyse and pre-lock review stages.

## **ANIMATED CLINICAL TRIAL DATA STORYTELLING**

A key contribution of this work is the establishment of animated clinical data storytelling as an additional approach for reviewing and communicating datasets. Animation allows for the interpretation of different dataset versions as a sequential timeline, enabling reviewers to directly observe the progression, stability, or disruption in data behavior.

Exportable animated visualizations offer a way to distill the complex evolution of datasets into brief visual narratives. These narratives can be incorporated into reports, presentations, or review documents, minimizing the necessity for extensive verbal or written explanations. For stakeholders without programming expertise, animation acts as an approachable medium for grasping how data evolves over time without the need for intricate technical details.

By transforming dataset review from a solely static validation task into an exploratory and explanatory activity, animation fosters a shared understanding and enhances overall efficiency.

## **BENEFITS AND PRACTICAL USE CASES**

The suggested framework presents unique advantages for various roles in clinical trials while conforming to established validation methods.

For statistical programmers, this method enables quicker exploratory quality checks, early identification of derivation or structural problems, and enhanced confidence before formal validation takes place. Statisticians enjoy clear visibility into the readiness and stability of datasets throughout interim analyses, whereas clinical reviewers receive a transparent perspective on the evolution of data during the study. At the study team level, animated visual outputs encourage alignment by offering a shared visual reference for discussions about the data.

Practical use cases include:

- Interim data review and ongoing study monitoring
- Database lock readiness assessments

- Validation of vendor data transfers across successive deliveries
- Regression testing of ADaM derivations following specification updates

## FUTURE ENHANCEMENTS

Future improvements could involve the automatic identification of SDTM and ADaM variable roles, and the automatic generation of narrative summaries to highlight key findings. Another key direction is AI-assisted narration using domain-adapted language models to generate contextual summaries of dataset shifts beyond rule-based statements. Additionally, the ability to analyze semi-structured datasets should be integrated, along with the inclusion of a "chat-with-data" bot for interactive querying and exploration.

Tackling these issues would bolster the effectiveness of animated data exploration as a standard element within clinical data review processes.

## CONCLUSION

This work demonstrates that animated and interactive visualization can significantly improve how evolving clinical trial datasets are reviewed, beyond what is possible with static tables, listings, and conventional plots. By treating successive SDTM/ADaM versions as a temporal sequence rather than isolated snapshots, the proposed Streamlit-based dashboard enables reviewers to directly observe shifts in distributions, missingness, record volumes, and multivariate structure across interim transfers and derivation iterations—supporting faster, more intuitive insight generation while remaining compatible with regulated programming workflows.

The platform integrates complementary analytical views including animated variable-level exploration, dataset-level trend monitoring, PCA-based drift detection with interpretation outputs, automated QC summaries for column/row changes with drill-down change localization, metadata consistency checks, and comparative descriptive statistics. Together, these capabilities facilitate earlier detection of unexpected dataset changes, strengthen traceability during interim and pre-lock review, and improve cross-functional communication through exportable GIF/MP4 visual narratives. Overall, this work positions animated clinical data storytelling as a practical extension to traditional QC, enabling clearer understanding and more efficient validation of dataset evolution in clinical development.

## REFERENCES

- [1] V. L. West, D. Borland, and W. E. Hammond, "Innovative information visualization of electronic health record data: A systematic review," *Journal of the American Medical Informatics Association*, vol. 22, no. 2, pp. 330–339, 2015. Available: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC4394966/>
- [2] E. Kukafka et al., "Interactive visualization applications in population health and health services research: Systematic scoping review," *Journal of Medical Internet Research*, vol. 24, no. 2, 2022. Available: <https://www.jmir.org/2022/2/e27534/>
- [3] M. Scheer et al., "Visualization techniques of time-oriented data for patient and cohort comparisons: Scoping review," *Journal of Medical Internet Research*, vol. 24, no. 10, 2022. Available: <https://www.jmir.org/2022/10/e38041/>
- [4] R. Ledesma et al., "Health timeline: An insight-based study of timeline visualization of clinical data," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, 2019. Available: <https://bmcmidinformedecismak.biomedcentral.com/articles/10.1186/s12911-019-0885-x>
- [5] B. C. Kwon et al., "DPVis: Visual analytics with hidden Markov models for disease progression pathways," *IEEE Transactions on Visualization and Computer Graphics*, 2019. Available: <https://arxiv.org/abs/1904.11652>
- [6] Y. Li et al., "TrialView: An AI-powered visual analytics system for temporal event data in clinical trials," *arXiv preprint*, 2023. Available: <https://arxiv.org/abs/2310.04586>
- [7] J. Heer and G. Robertson, "Animated transitions in statistical data graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1240–1247, 2007. Available: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/InfoVis2007-DynaVis.pdf>
- [8] OpenAI. (2026). ChatGPT (20 Jan version) [Large language model]. <https://chat.openai.com>

## CONTACT INFORMATION (HEADER 1)

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Tony Mathew

Company: Pfizer Healthcare India Pvt.Ltd.

Address: New No.237, Anna Salai, Thousand Lights, Chennai, Tamil Nadu 600014, India

Email: [tony.mathew@pfizer.com](mailto:tony.mathew@pfizer.com)

Website:

Brand and product names are trademarks of their respective companies.