

Sentinel: Advancing Automated Data Quality Checks in OHDSI Networks

Background: High-quality electronic health record (EHR) data are essential for research, yet manual abstraction can exceed 2,000 hours per study. Sentinel is a framework designed to continuously monitor and enhance EHR data quality. It combines OMOP CDM structure and NLP outputs to monitor completeness, plausibility, and conformance.

Methods: Sentinel applies a standardized, automated workflow to evaluate data integrity within OMOP CDM datasets. It integrates outputs from structured sources and NLP pipelines, highlights discrepancies, and provides benchmarking across sites. Live dashboards visualize metrics to guide data quality improvement at scale. This streamlines ETL processes and reduces manual review.

Figure 1. Overview of dataset quality metrics and concept counts in a Sentinel demo dashboard.

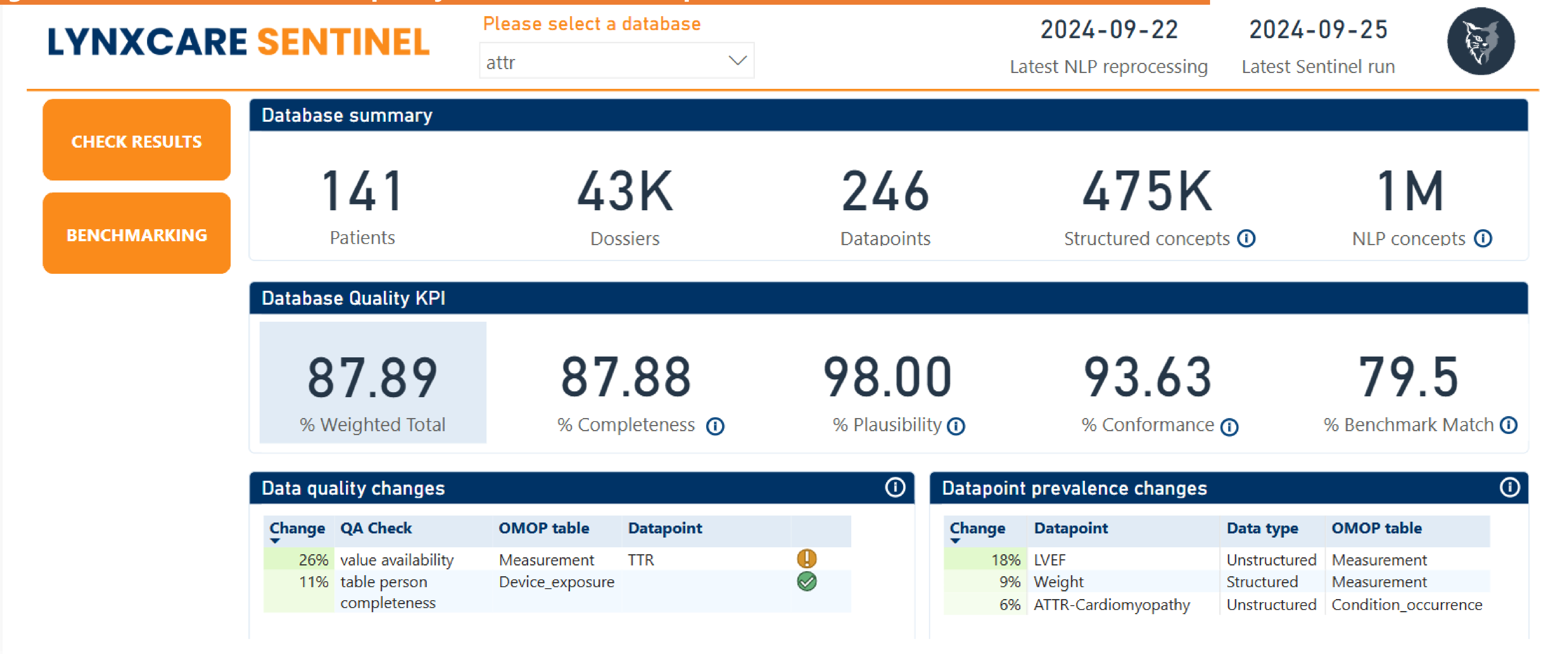
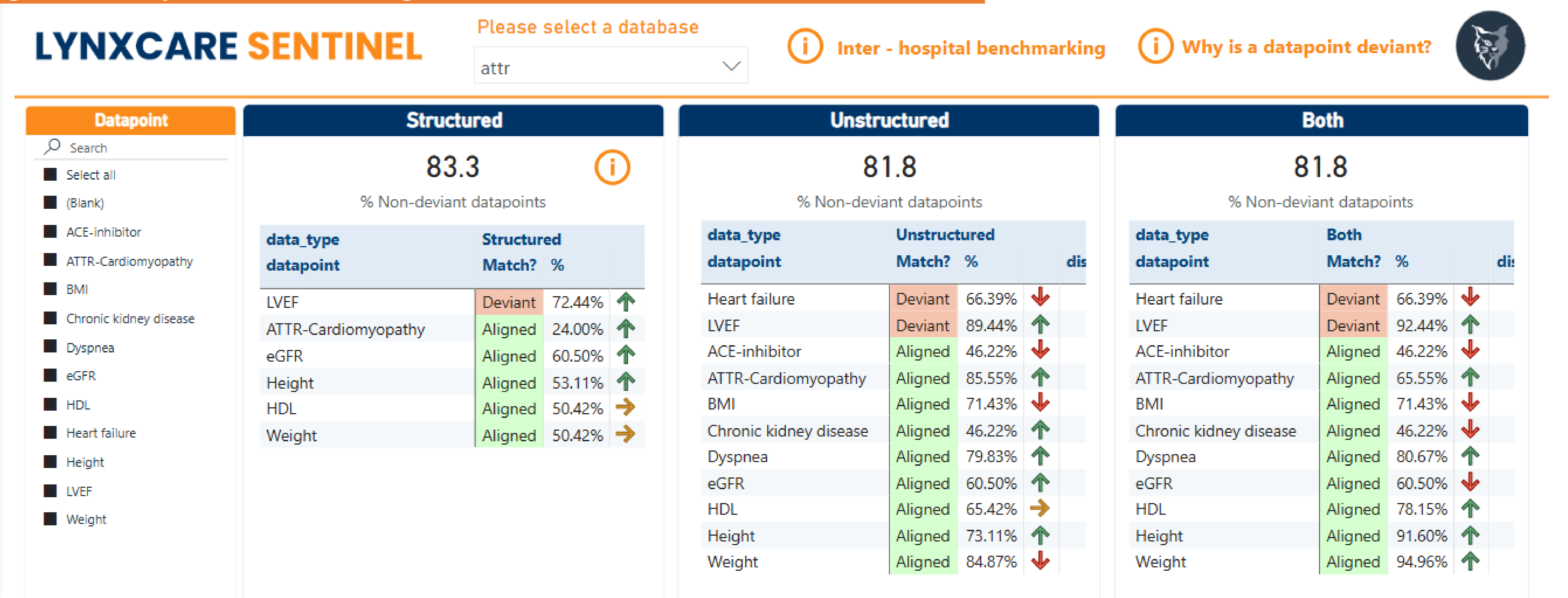


Figure 2. Datapoint benchmarking across structured and unstructured sources.



Conclusion: By automating data quality checks across diverse hospital datasets, Sentinel reduces reliance on time-intensive manual review. Its integration with OMOP and NLP outputs enhances transparency and reproducibility, supporting faster insight generation and advancing the goals of EHDS and OHDSI-aligned research.

