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.

LYNXCARE SENTINEL

Please select a database: attr

2024-09-22 Latest NLP reprocessing

2024-09-25 Latest Sentinel run

CHECK RESULTS

BENCHMARKING

Database summary

141	43K	246	475K	1M
Patients	Dossiers	Datapoints	Structured concepts	NLP concepts

Database Quality KPI

87.89	87.88	98.00	93.63	79.5
% Weighted Total	% Completeness	% Plausibility	% Conformance	% Benchmark Match

Data quality changes

Change	QA Check	OMOP table	Datapoint
26%	value availability	Measurement	TTR
11%	table person completeness	Device_exposure	

Datapoint prevalence changes

Change	Datapoint	Data type	OMOP table
18%	LVEF	Unstructured	Measurement
9%	Weight	Structured	Measurement
6%	ATTR-Cardiomyopathy	Unstructured	Condition_occurrence

Figure 2. Datapoint benchmarking across structured and unstructured sources.

LYNXCARE SENTINEL

Please select a database: attr

Inter - hospital benchmarking

Why is a datapoint deviant?

Datapoint

Search

Select all

(Blank)

ACE-inhibitor

ATTR-Cardiomyopathy

BMI

Chronic kidney disease

Dyspnea

eGFR

HDL

Heart failure

Height

LVEF

Weight

Structured

83.3

% Non-deviant datapoints

data_type / datapoint	Structured Match?	%
LVEF	Deviant	72.44%
ATTR-Cardiomyopathy	Aligned	24.00%
eGFR	Aligned	60.50%
Height	Aligned	53.11%
HDL	Aligned	50.42%
Weight	Aligned	50.42%

Unstructured

81.8

% Non-deviant datapoints

data_type / datapoint	Unstructured Match?	%	dis
Heart failure	Deviant	66.39%	
LVEF	Deviant	89.44%	
ACE-inhibitor	Aligned	46.22%	
ATTR-Cardiomyopathy	Aligned	85.55%	
BMI	Aligned	71.43%	
Chronic kidney disease	Aligned	46.22%	
Dyspnea	Aligned	79.83%	
eGFR	Aligned	60.50%	
HDL	Aligned	65.42%	
Height	Aligned	73.11%	
Weight	Aligned	84.87%	

Both

81.8

% Non-deviant datapoints

data_type / datapoint	Both Match?	%	dis
Heart failure	Deviant	66.39%	
LVEF	Deviant	92.44%	
ACE-inhibitor	Aligned	46.22%	
ATTR-Cardiomyopathy	Aligned	65.55%	
BMI	Aligned	71.43%	
Chronic kidney disease	Aligned	46.22%	
Dyspnea	Aligned	80.67%	
eGFR	Aligned	60.50%	
HDL	Aligned	78.15%	
Height	Aligned	91.60%	
Weight	Aligned	94.96%	

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.

OHDSI

Jana Van Canneyt[1], Nicolas De Roover[1], Karen D'Haeseleer[1], Dries Hens[1], Clara L. Oeste[1]
[1]LynxCare Inc., Leuven, Belgium.

LYNXCARE