

Improving Data Quality in Oncology Electronic Health Record (EHR)-Derived Databases for Research and Regulatory Decision-Making

Zhaohui Su, PhD; Lisa Herms, PhD; Aniketh Talwai, MBA; Jessica Paulus, ScD; Amy K. O'Sullivan, PhD; Sarah Spark, RN; Jennifer Frytak, PhD; Robyn Harrell, MS; Jay Hugalavalli, BS; Janet Espirito, PharmD — Ontada, Boston, MA, USA

Objectives

- Implement the FDA-led initiative on the oncology quality, characterization, and assessment of real-world data (RWD) (QCARD¹) to evaluate the relevance, reliability, and external validity of EHR data.
- Facilitate the use of robust RWD and real-world evidence (RWE) in oncology.

Methods

The RWD quality dimensions in the FDA QCARD initiative were implemented using Ontada's ON.Genuity RWD platform, which integrates the iKnowMed® EHR data from ~500 US community oncology clinics with claims and external mortality data.

Results

- Common challenges to RWD were mitigated by integrating data from multiple sources (**Table 1**).
- The RWD quality dimensions defined in the FDA QCARD initiative were implemented (**Figure 1-3**).
- Assessment of relevance revealed consistent high availability of 250+ standardized variables across 20+ clinical domains.
- Mortality and recurrence-free survival data demonstrated high consistency with the external National Death Index (NDI) data source and clinical trial data (placebo arm) (**Figure 4**).

Figure 1. Relevance

Availability (data elements)
<ul style="list-style-type: none">250+ standardized variables across 60+ cancersOncology-specific EHR data
Feasibility (representative patients)
<ul style="list-style-type: none">Largest network of community oncologists in the US with 2,700+ providersNearly 1 in 4 U.S. cancer patients treated in The Network15-year longitudinal near real-time linked data

Figure 2. Reliability

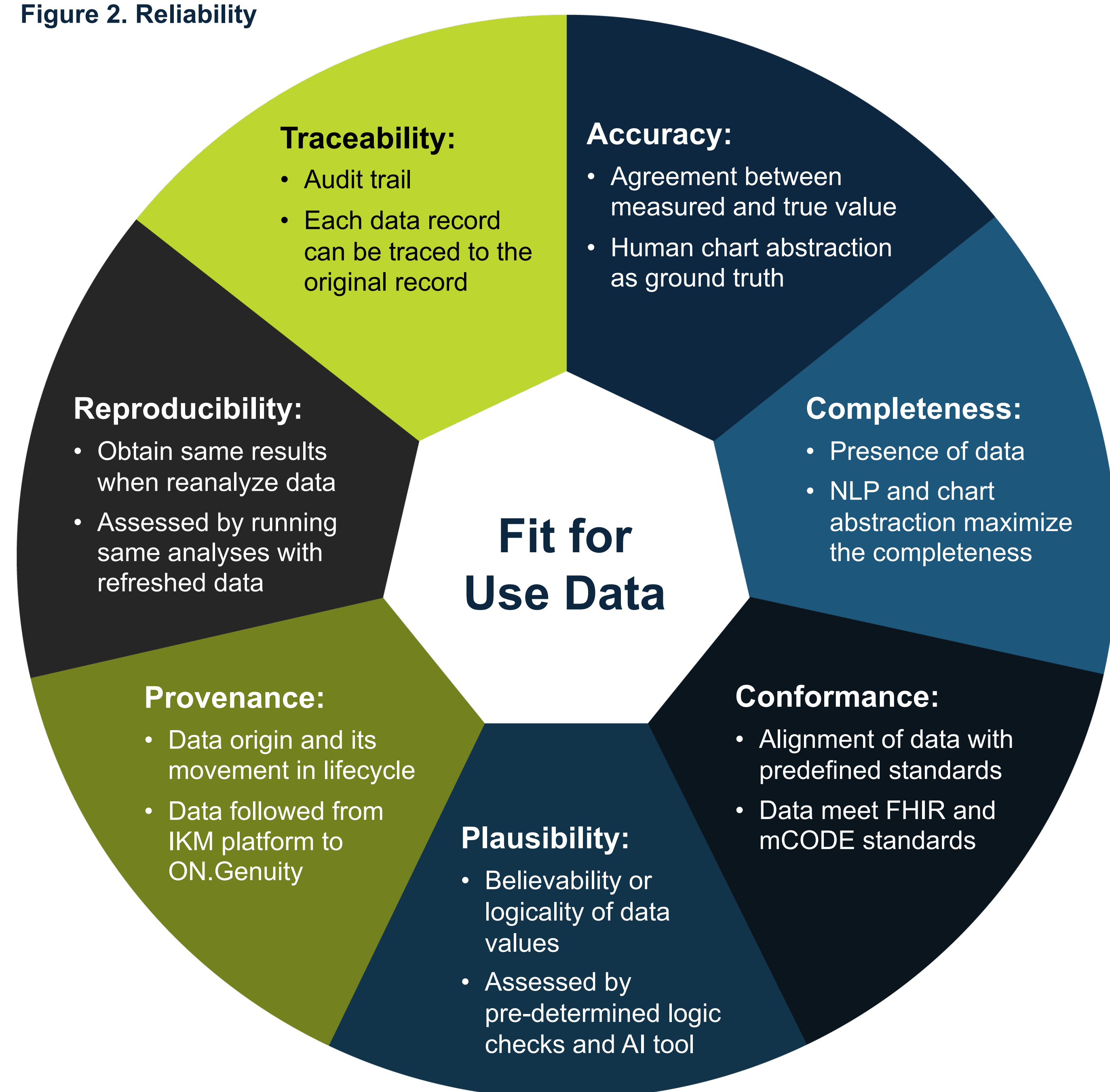


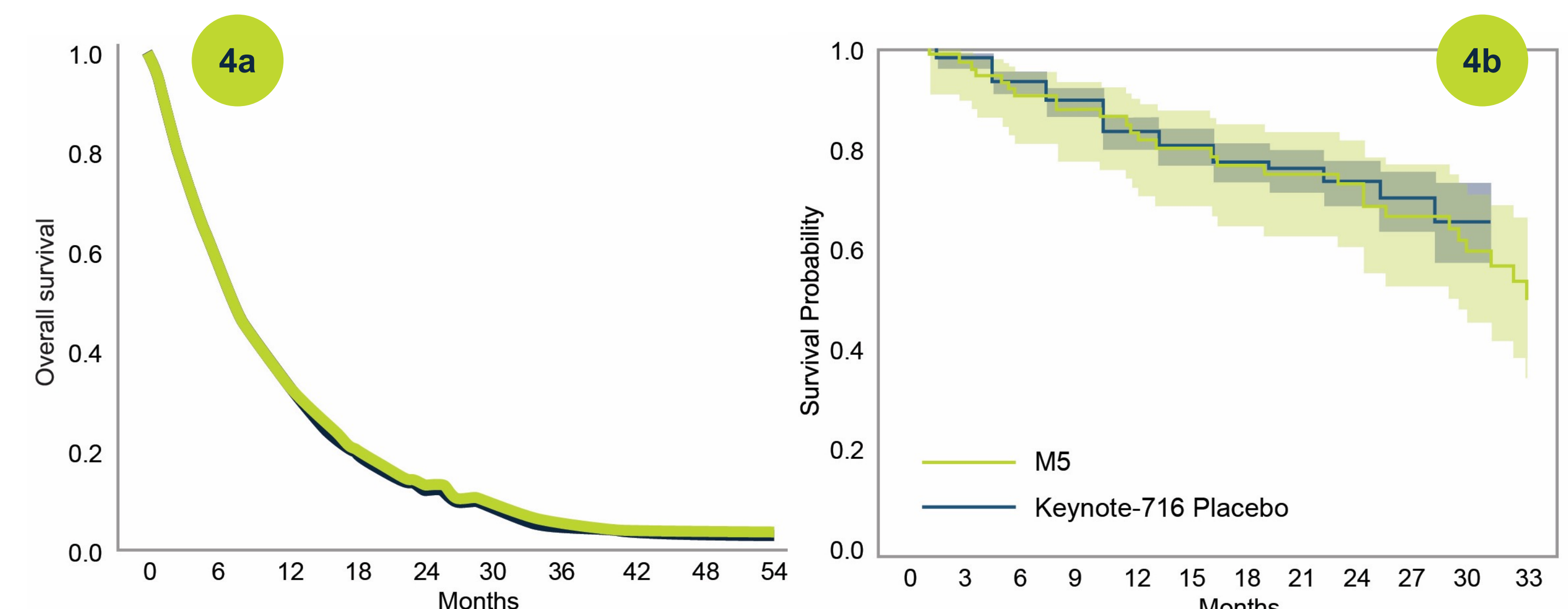
Table 1. Data Sources in Ontada's ON.Genuity RWD

Data source	Structured iKnowMed® EHR Data	Chart abstraction	Non-EHR Data
Example Variables	Gender, race, ethnicity, vital signs, labs, performance status, stage	Oral medication stop dates, surgery date, hospitalizations disease progression status	Claims data (amount paid, payer type), supplemental mortality data
Processing Method	Ingestion, standardization, reconciliation	Technology-enabled human abstraction, natural language processing (NLP)	Linked data at patient level
Highlights	<ul style="list-style-type: none">250+ variables were standardized following mCODE² initiative.EHR data were linked based on unique and accurate ID within iKnowMed®.		

Figure 3. External Validity

Generalizability (inference to broader population)	Replicability (confirm findings in different populations)	Transparency (clear communication and description)
Generalizable to community oncology setting in the US	Validation against National Death Index, published clinical trial data	Detailed documentation of derivations, design, methods, analyses

Figure 4. Implementation of external validity. (a) external validity of OS for metastatic pancreatic cancer showing identical curves between EHR and NDI, and (b) Kaplan-Meier for recurrence-free survival (RFS) in real-world melanoma patients vs KN-716 control arm



Discussion

Ontada's ON.Genuity EHR database was standardized using FHIR and mCODE², with increased relevance, reliability and external validity by linking to external structured data at a patient level, human chart abstraction, and leveraging NLP technology.

The quality and fit-for-use of RWD should be carefully evaluated for each study, including a detailed assessment of relevance, reliability and external validity in relation to the research objectives.

Conclusions

- The QCARD initiative represents a significant advancement in EHR-based oncology research, enhancing RWE from RWD and improving decision-making in clinical practice and policy.
- The quality and fit-for-use of RWD should be carefully evaluated in accordance with the research objectives.

References

- Rivera DR, Eckert JC, Rodriguez-Watson C, et al. The Oncology QCARD Initiative: Fostering efficient evaluation of initial real-world data proposals. *Pharmacoepidemiol Drug Saf.* 2024 Nov;33(11):e5818.
- Osterman TJ, Terry M, Miller RS. Improving Cancer Data Interoperability: The Promise of the Minimal Common Oncology Data Elements (mCODE) Initiative. *JCO Clin Cancer Inform.* 2020 Oct;4:993-1001.