cdisc

2022
US
INTERCHANGE
26-27 OCTOBER | AUSTIN

Processing real-world data for regulatory decisions: minimizing data loss, maximizing information

Presented by Tasha Nagamine | Droice Labs

Meet the Speaker

Tasha Nagamine, MS

Title: Cofounder & CTO

Organization: Droice Labs

Tasha is an entrepreneur and seasoned technologist with 10+ years of experience in AI, research, and tech strategy. She brings deep expertise in RWD to build data-driven products that have processed over 100 million patient lives. Tasha received her BS in physics from Brown University and left a PhD in deep learning at Columbia University to start Droice.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *Tasha Nagamine is an employee of Droice Labs, a company that processes and prepares RWD for clinical and research applications.*

WHY RWD?

Studies used to demonstrate the safety and effectiveness of a drug should reflect the diversity of patients that the drug is intended for[1]

Clinical trial populations

- Small N
- Homogeneous
- Low noise

Real-world populations

- Large N
- Heterogeneous
- High noise

1 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8697198/

CHALLENGES WITH RWD

Because of the variation across patients, physicians, & sources...

...Diabetes will manifest differently in the data for different patients

CHALLENGES WITH RWD

Because of the variation across patients, physicians, & sources...

Disease variation

Type 2 diabetes with poor adherence & multiple complications (diabetic kidney disease, diabetic foot ulcers, neuropathy)

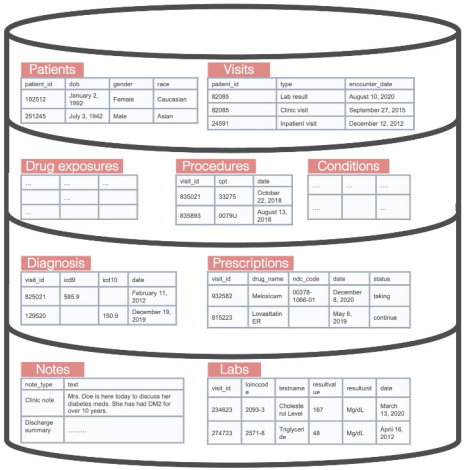
Patient 1

Patient 2

Well-controlled type 2 diabetes

...Diabetes will manifest differently in the data for different patients

Data source



CHALLENGES WITH RWD

Because of the variation across patients, physicians, & sources...

Disease variation

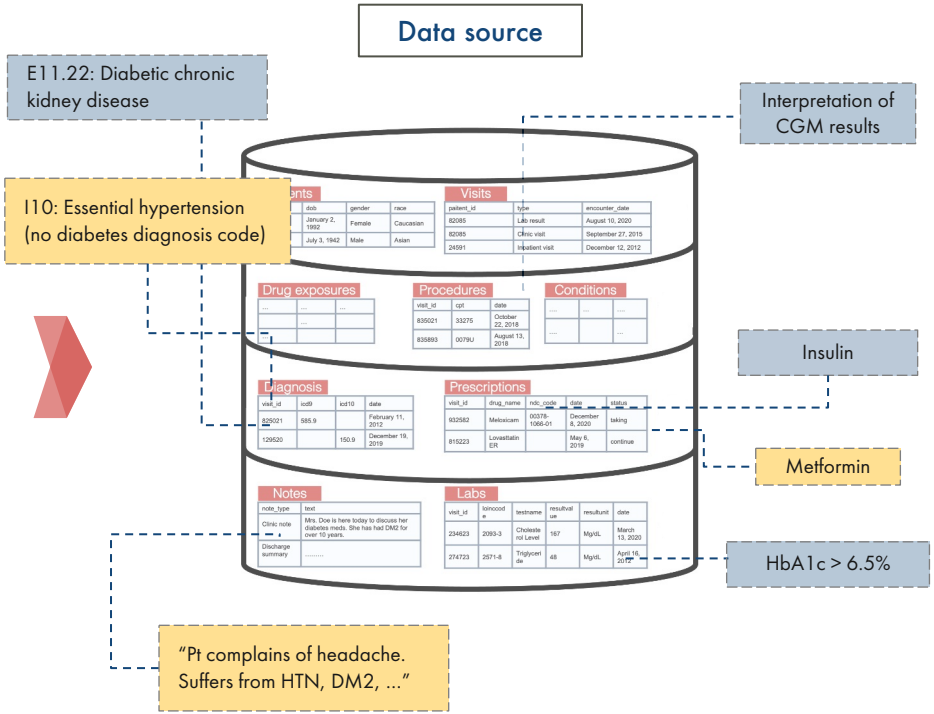
Type 2 diabetes with poor adherence & multiple complications (diabetic kidney disease, diabetic foot ulcers, neuropathy)

Patient 1

Patient 2

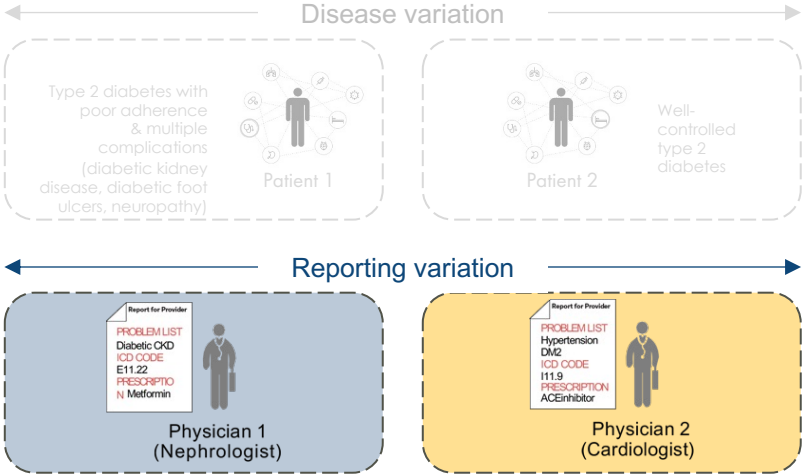
Well-controlled type 2 diabetes

...Diabetes will manifest differently in the data for different patients

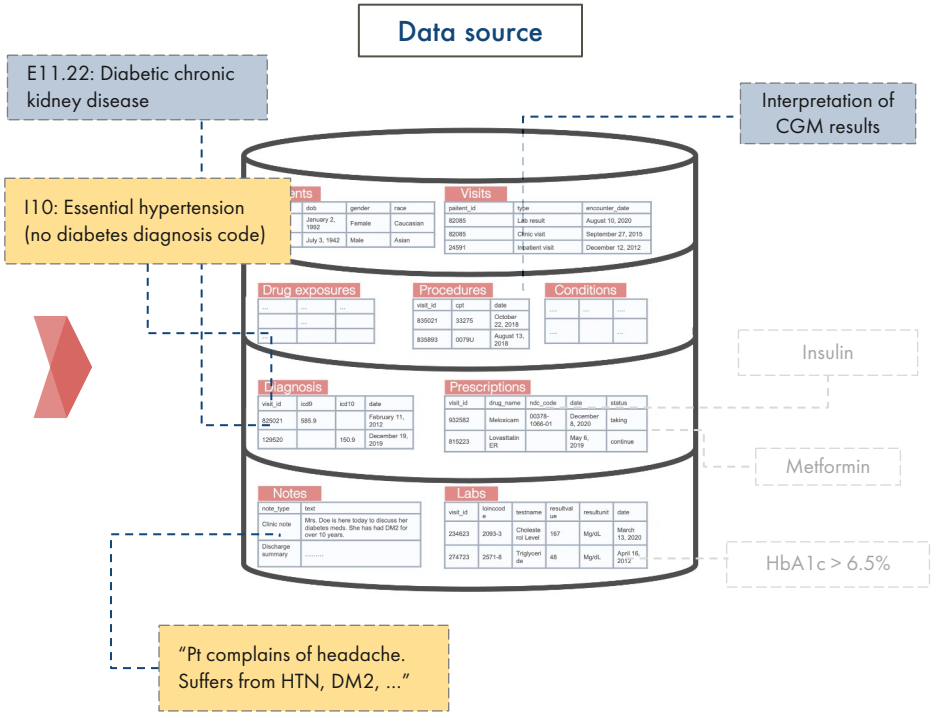


CHALLENGES WITH RWD

Because of the variation across patients, physicians, & sources...

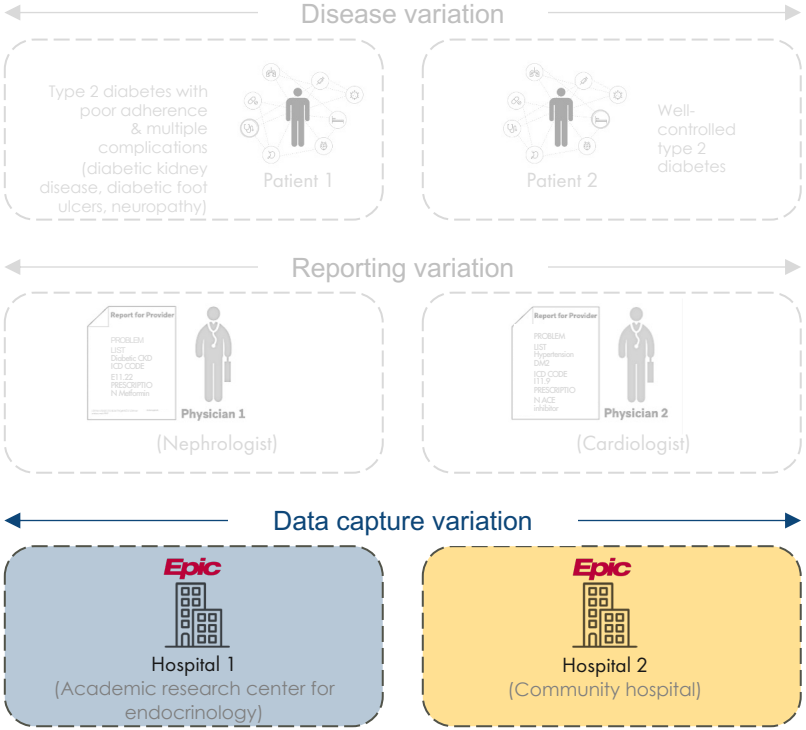


...Diabetes will manifest differently in the data for different patients

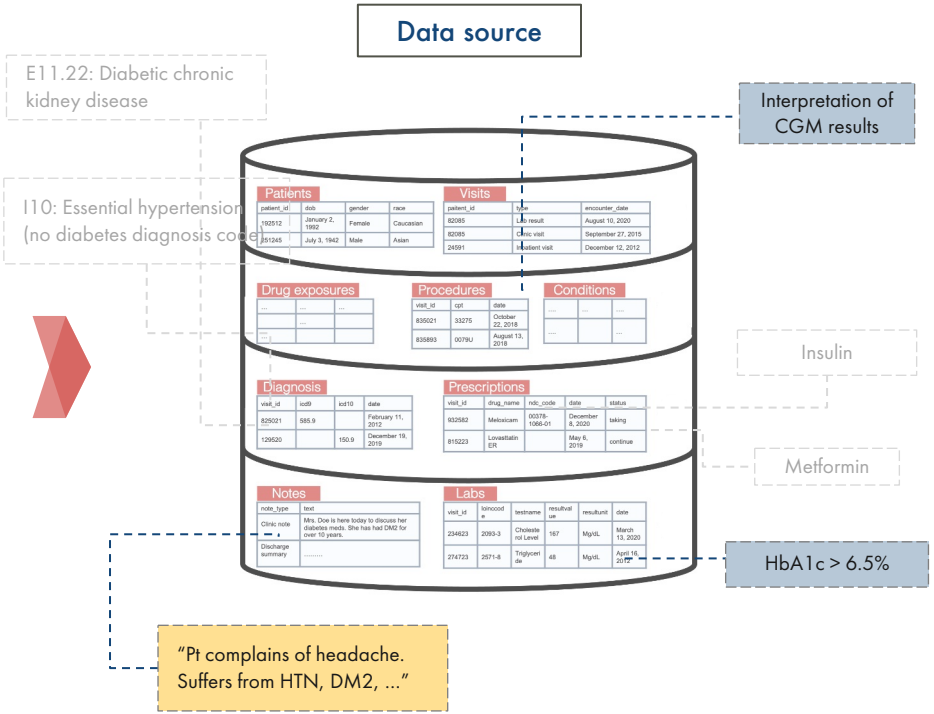


CHALLENGES WITH RWD

Because of the variation across patients, physicians, & sources...

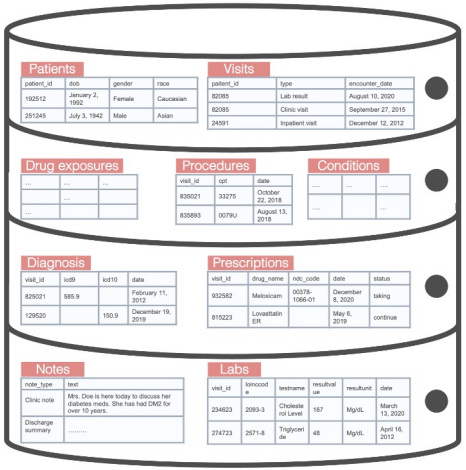


...Diabetes will manifest differently in the data for different patients

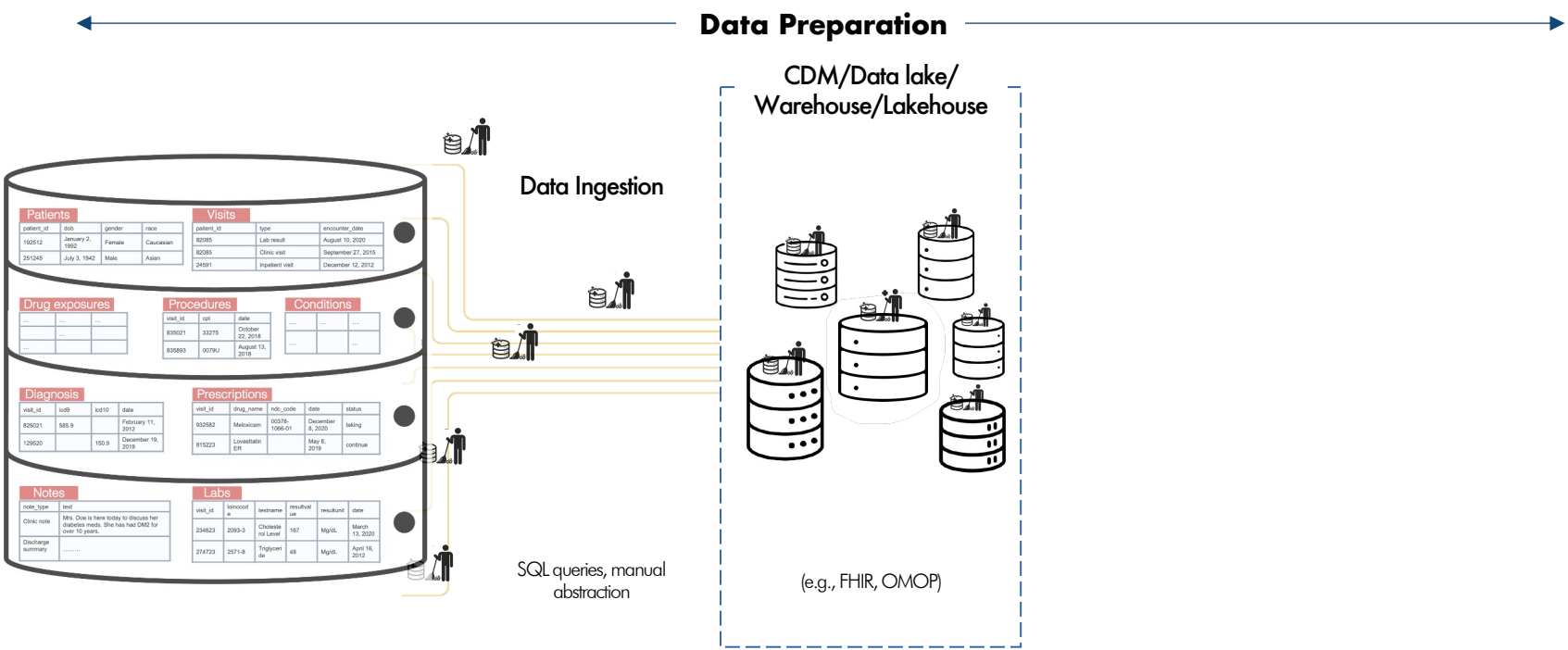


CURRENT STATE OF HEALTHCARE DATA: INFORMATION LOSS

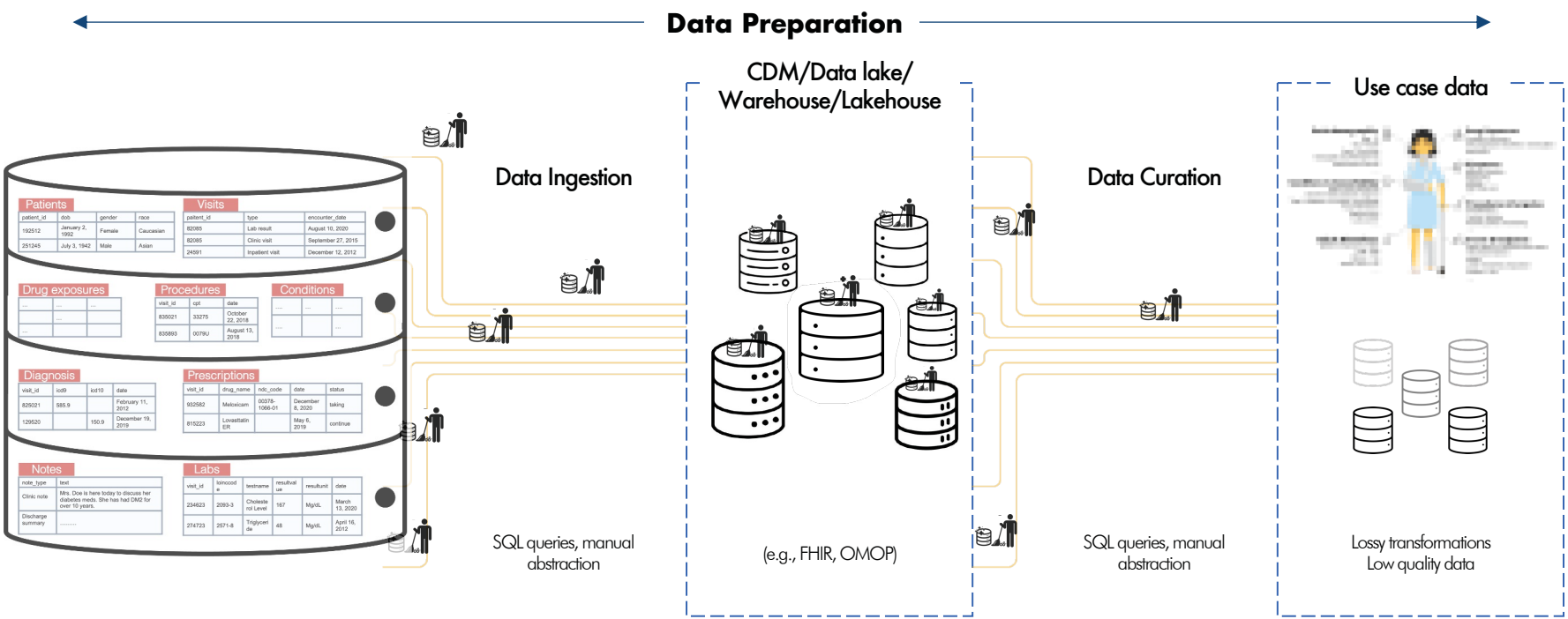
Data Preparation



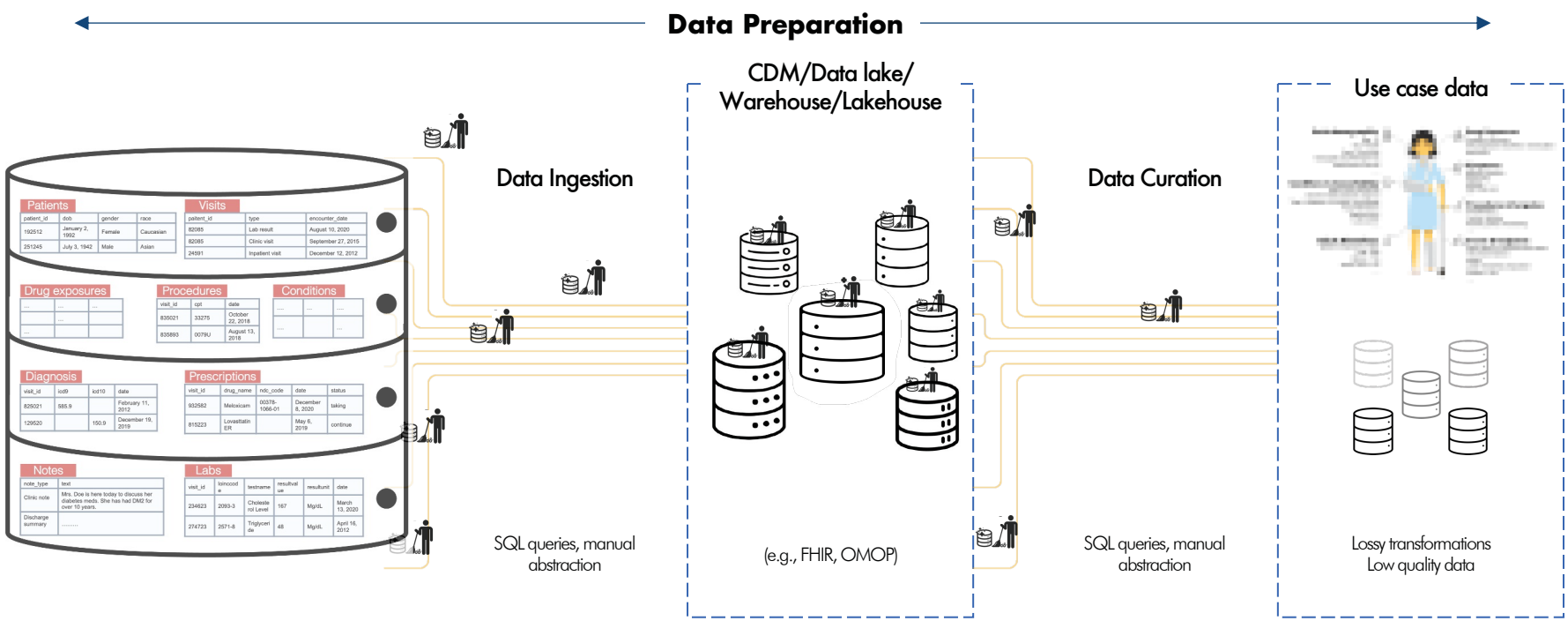
CURRENT STATE OF HEALTHCARE DATA: INFORMATION LOSS



CURRENT STATE OF HEALTHCARE DATA: INFORMATION LOSS



CURRENT STATE OF HEALTHCARE DATA: INFORMATION LOSS



- **Speed** — Slow & manual
- **Reproducibility** — Difficult to port & reproduce
- **Traceability** — Lacks transparency
- **Quality** — Incomplete, lossy & error-prone
- **Cost** — Expensive & inefficient

TRADEOFF: SCALE VS. QUALITY

Quality

Complete
Reproducible
Lineage & provenance
Minimal data loss

Incomplete
Not reproducible
Low traceability
Lossy transformations

Small scale

manual query writing, abstraction, & curation

Low data volume, high quality

- TAs: oncology, rare disease
- Use cases: RCT (SDTM, ADaM), disease registries, …
- Quality assurance depends on manual curation/low patient volumes

Large scale

FHIR / OMOP

manual query writing & ETL-based data pipelines

High data volume, low quality

- TAs: all (e.g., cardiovascular, endocrine, etc.)
- Use cases: interoperability, HEOR, …
- Difficult to ensure data quality

Scale

N = 10-100 | N = 100-1,000 | N = 1,000-10,000 | N = 10,000-100,000 | N = 100,000-1,000,000 | N > 1,000,000

CHRONIC KIDNEY DISEASE

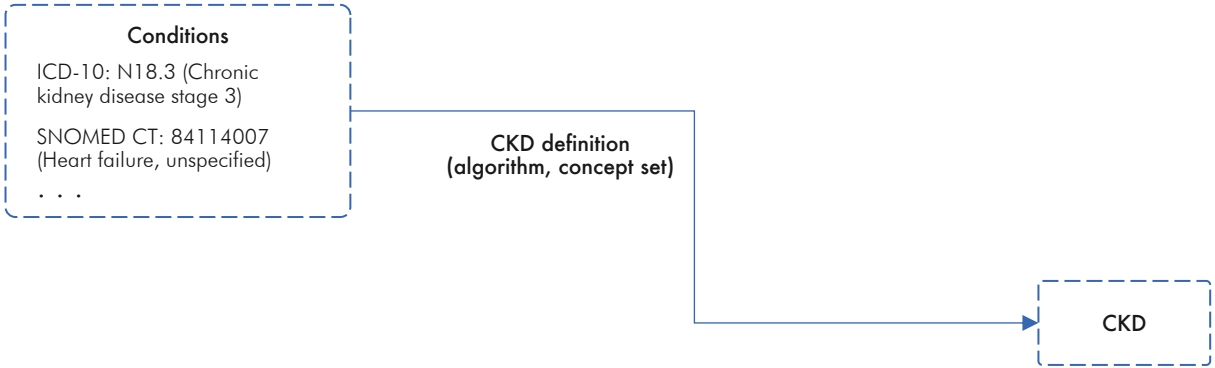
CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.

CKD

RAW DATA | FHIR/OMOP DATA | PHENOTYPED DATA | ANALYSIS-READY DATA

CHRONIC KIDNEY DISEASE

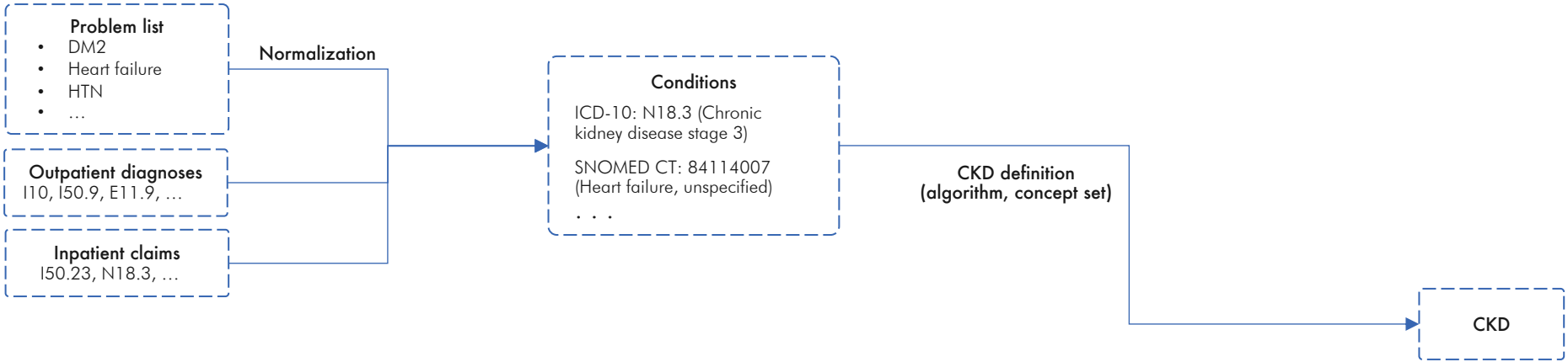
CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.



RAW DATA | FHIR/OMOP DATA | PHENOTYPED DATA | ANALYSIS-READY DATA

CHRONIC KIDNEY DISEASE

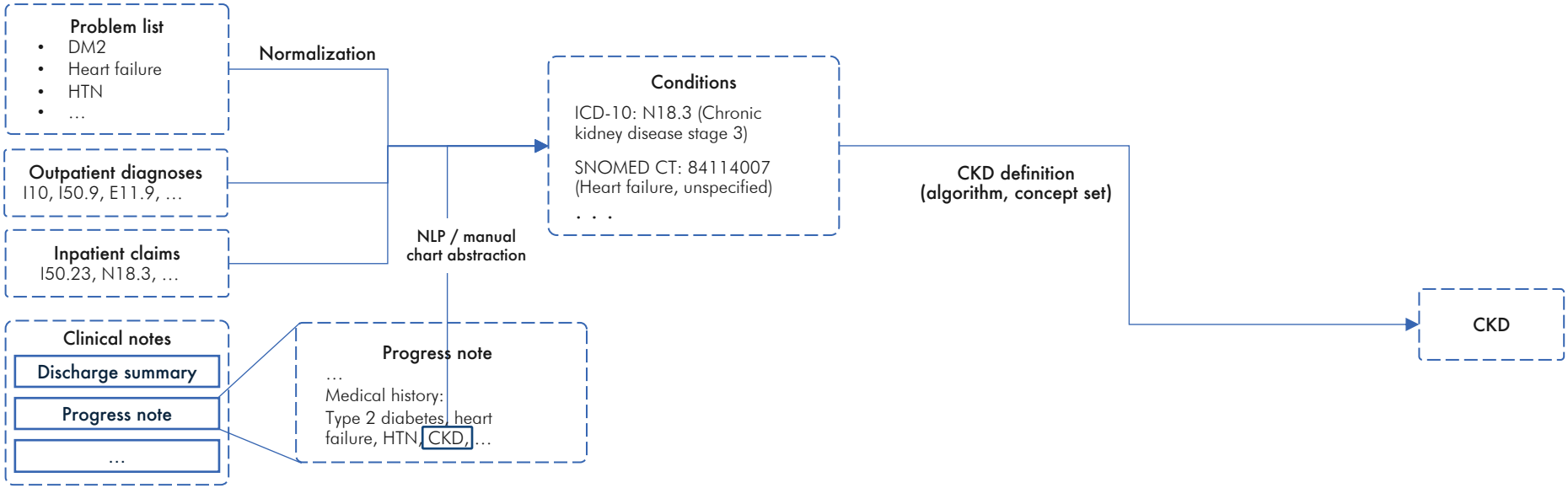
CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.



RAW DATA | FHIR/OMOP DATA | PHENOTYPED DATA | ANALYSIS-READY DATA

CHRONIC KIDNEY DISEASE

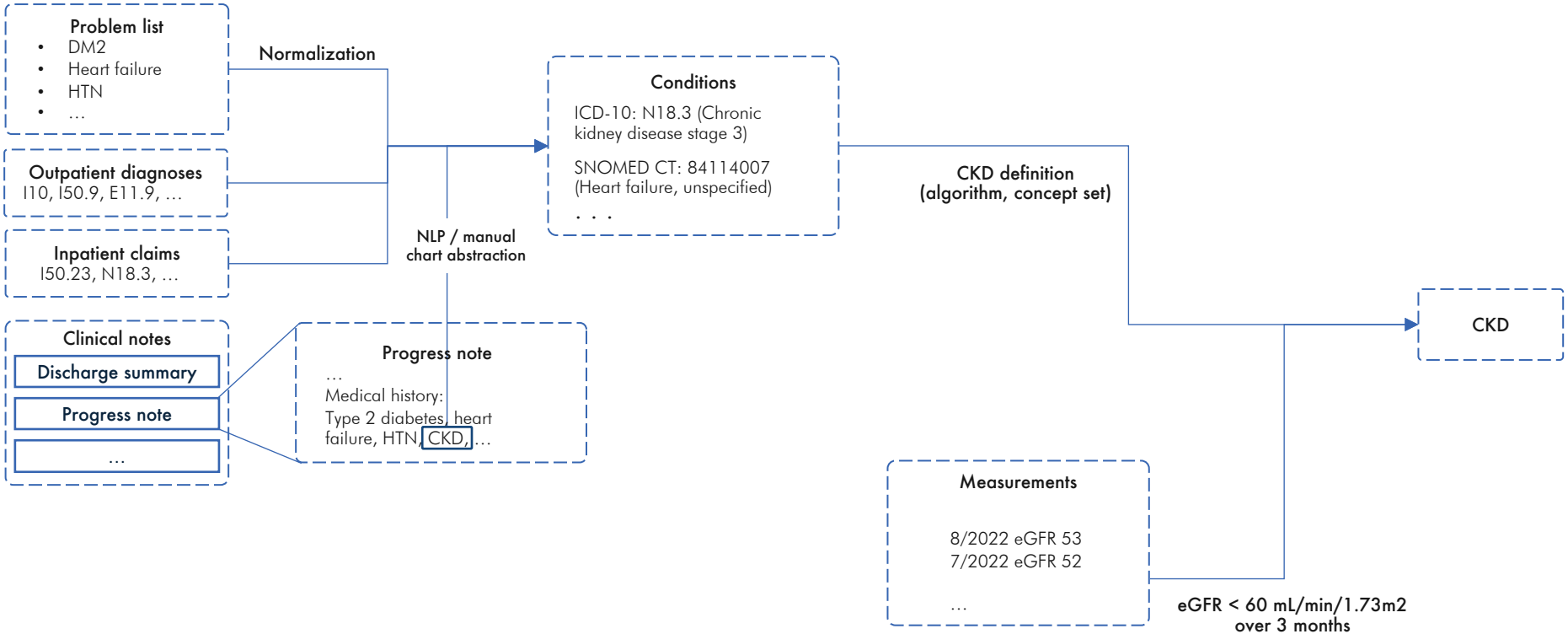
CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.



RAW DATA | FHIR/OMOP DATA | PHENOTYPED DATA | ANALYSIS-READY DATA

CHRONIC KIDNEY DISEASE

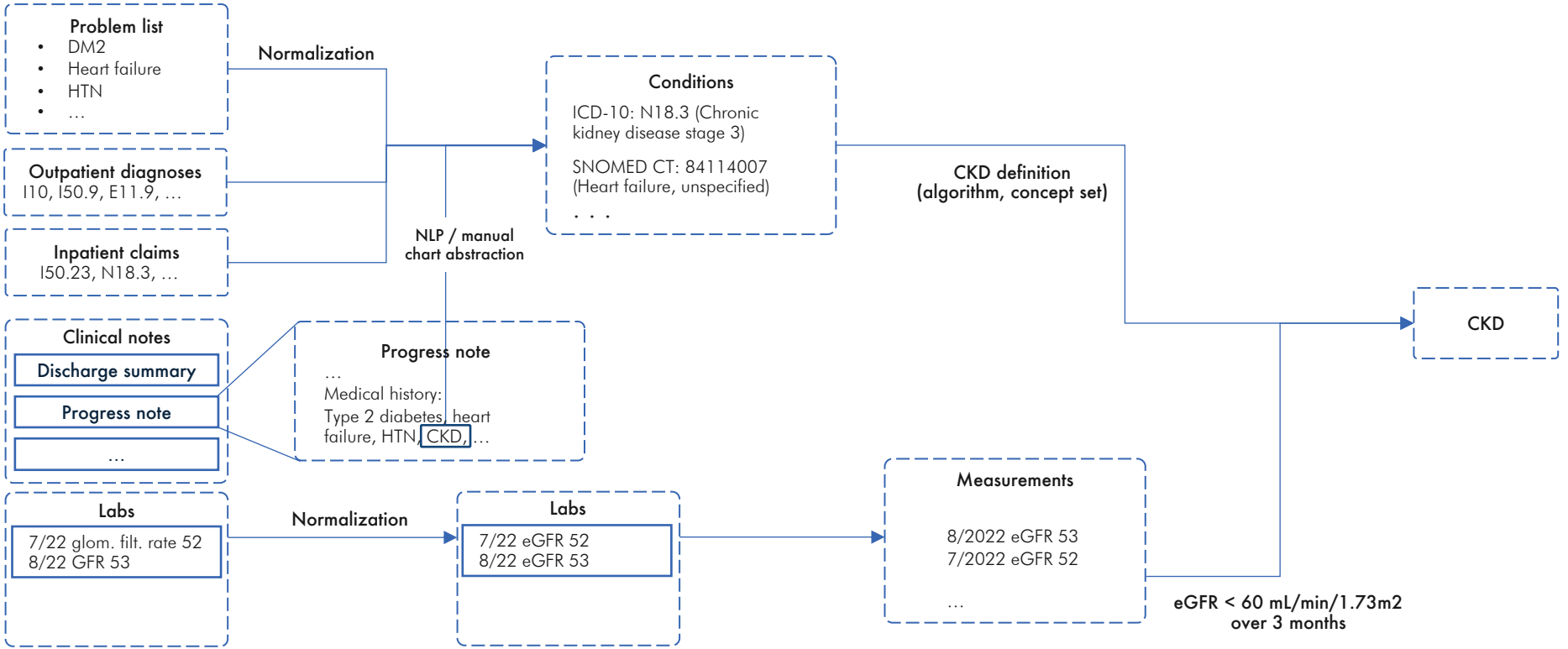
CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.



RAW DATA | FHIR/OMOP DATA | PHENOTYPED DATA | ANALYSIS-READY DATA

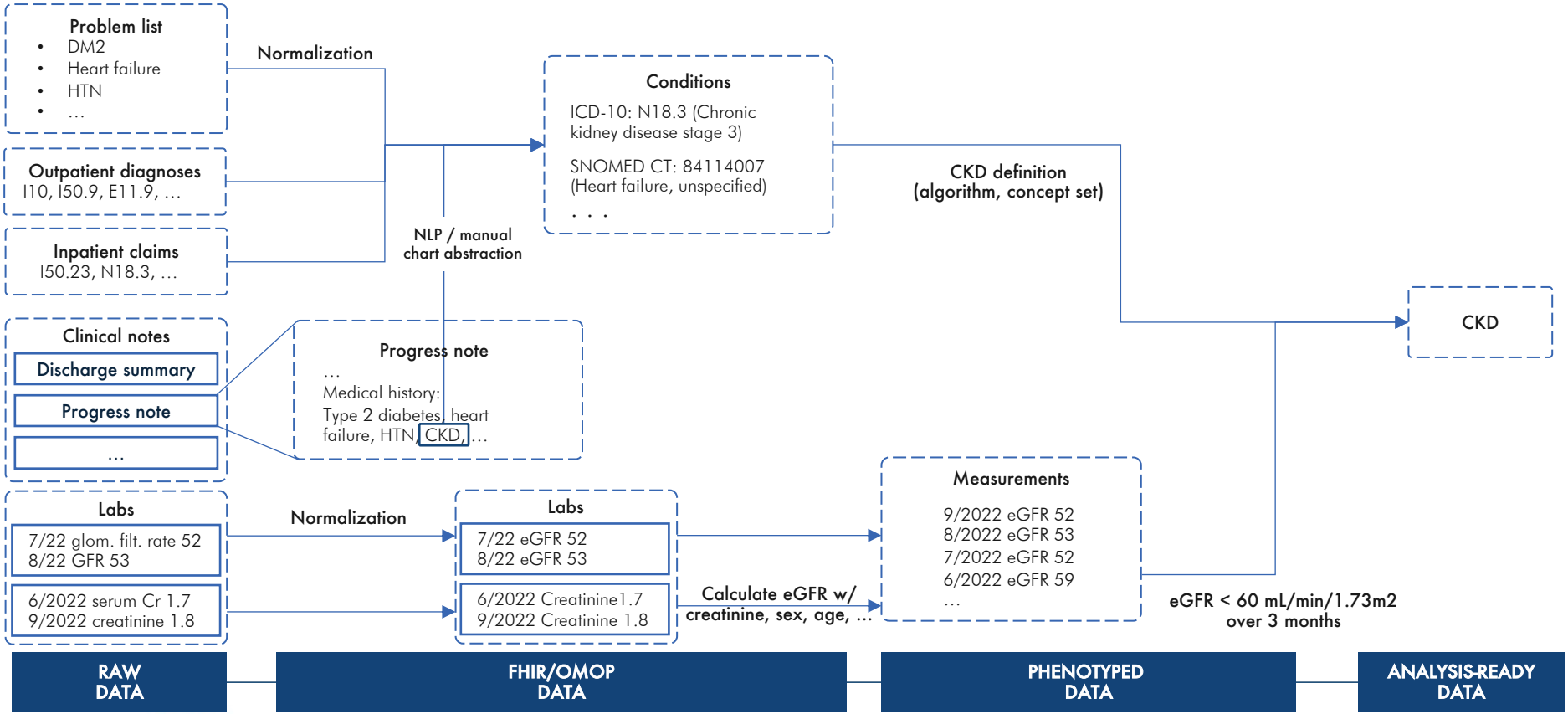
CHRONIC KIDNEY DISEASE

CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.



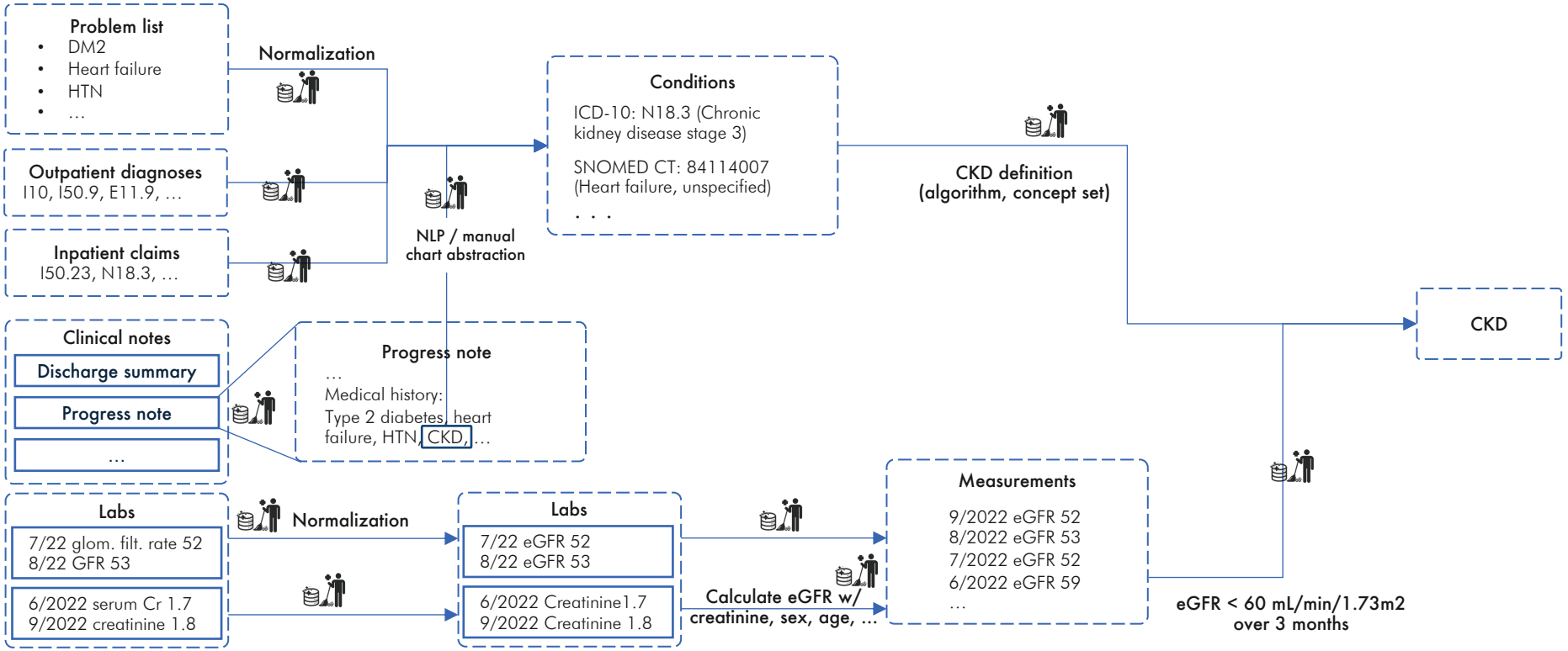
CHRONIC KIDNEY DISEASE

CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.



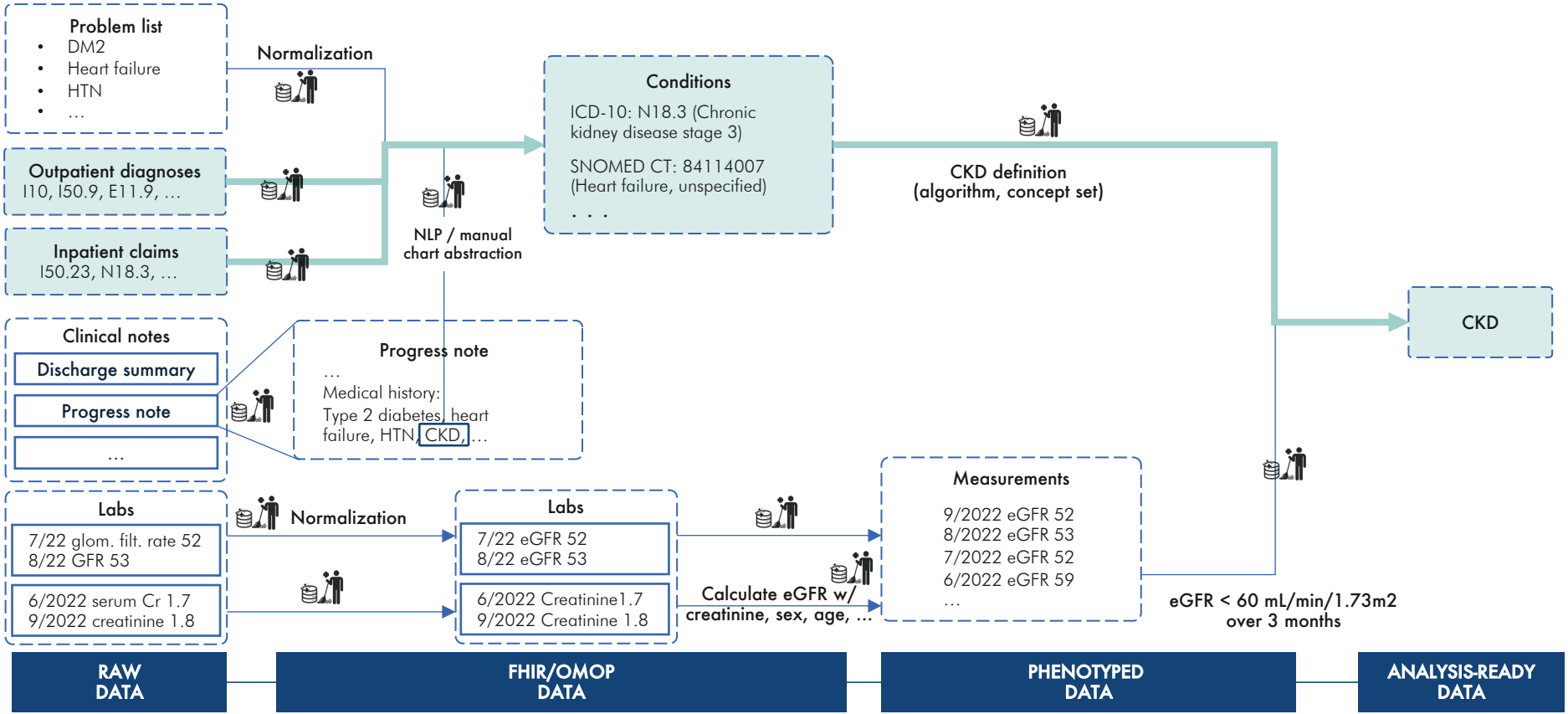
CHRONIC KIDNEY DISEASE

CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.

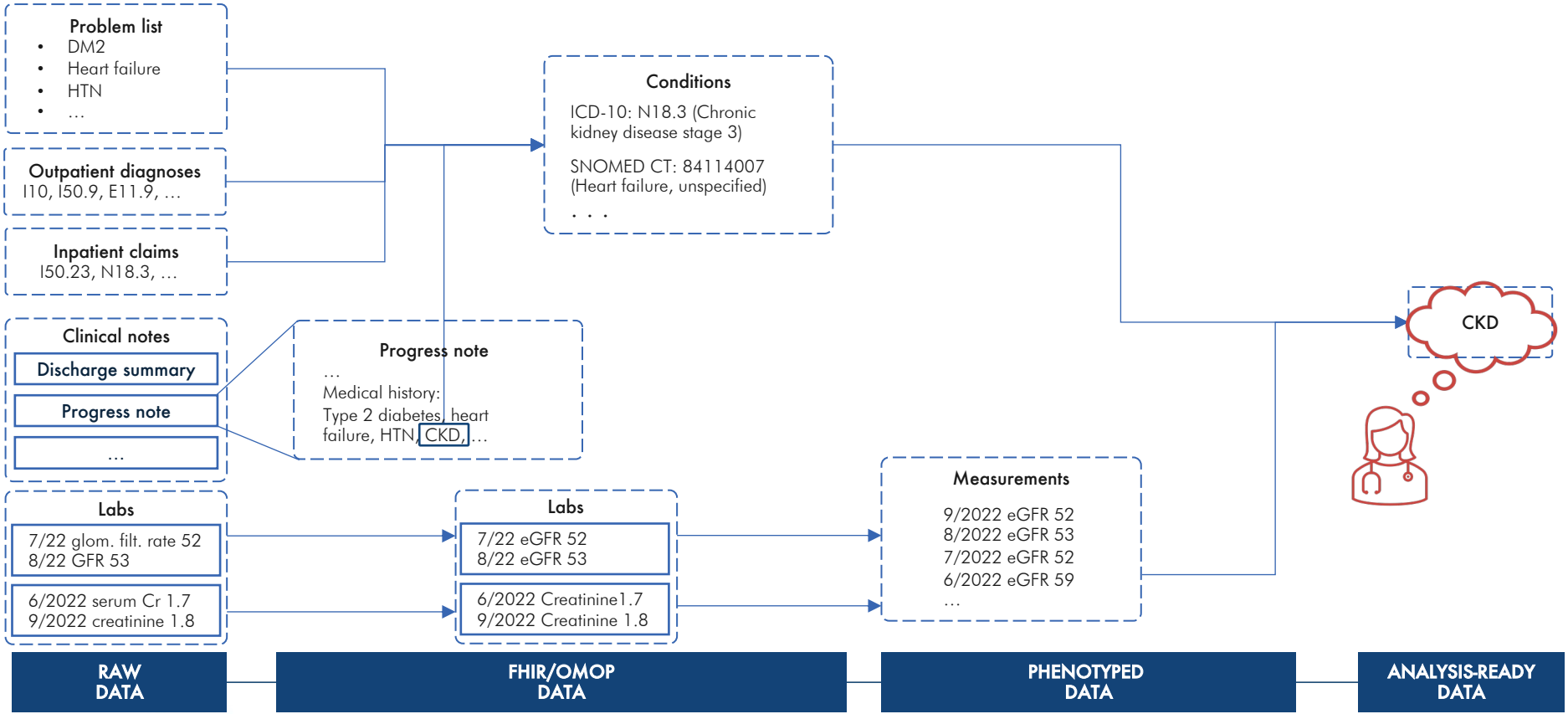


CHRONIC KIDNEY DISEASE

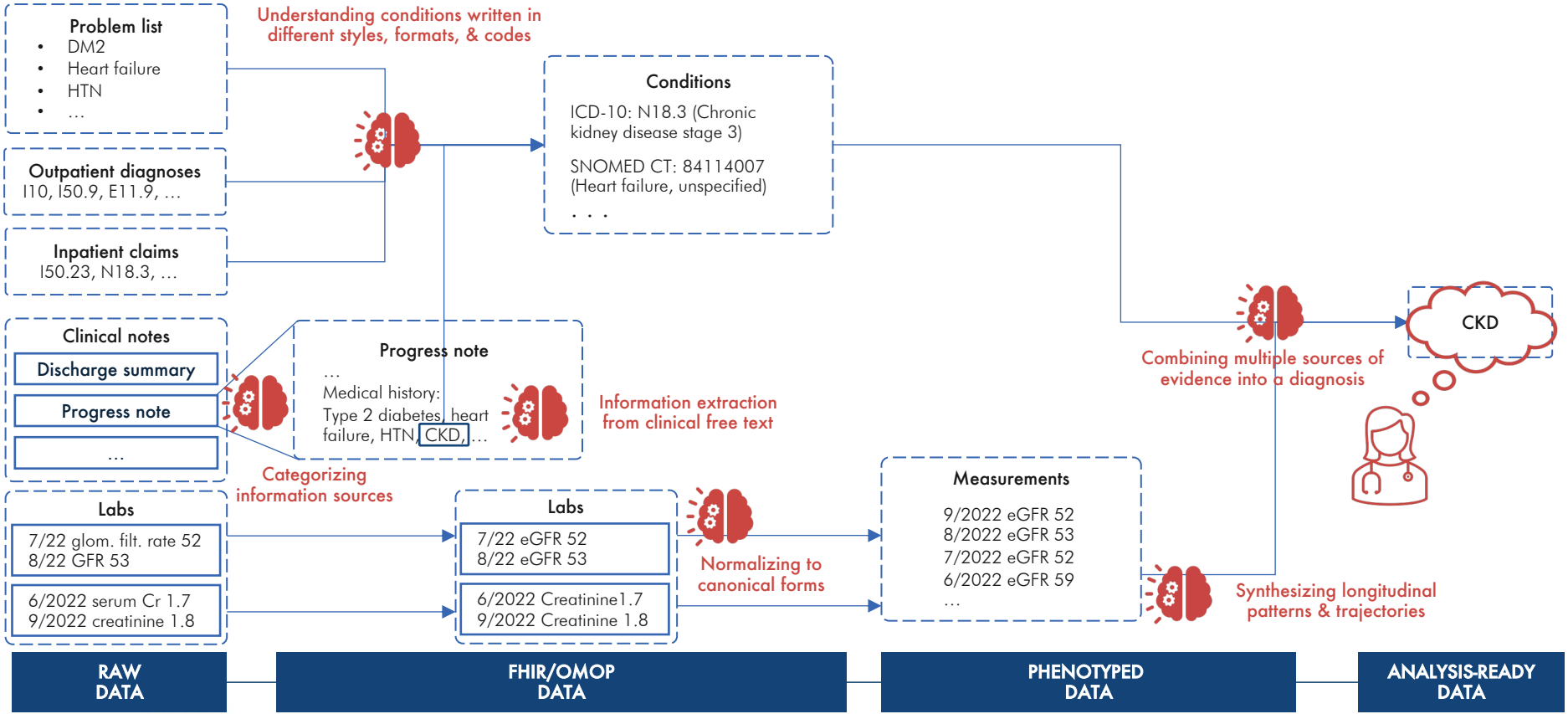
CKD is often under-coded or underreported, depending on practice variation/billing incentives in RWD.



AN INTELLIGENCE-BASED APPROACH



AN INTELLIGENCE-BASED APPROACH



AI MIDDLEWARE APPROACH

Required for general purpose RWD (i.e., data with large N & high noise) to maximize information and minimize data loss.

Data Preparation

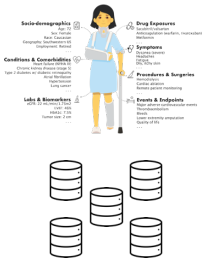
Data sources

- EMR
- Claims
- HIE
- HL7 FHIR Feeds
- Labs & diagnostics
- Trial & research data
- Apps

AI middleware

Droice Hawk

Use case data



No data loss
Regulatory-grade data

- ☑ **Speed** — Fast & scalable
- ☑ **Reproducibility** — Portable & reproducible
- ☑ **Traceability** — Transparency
- ☑ **Quality** — Comprehensive& lossless
- ☑ **Cost** — Cost efficient

AI MIDDLEWARE APPROACH

Required for general purpose RWD (i.e., data with large N & high noise) to maximize information and minimize data loss.

Data Preparation

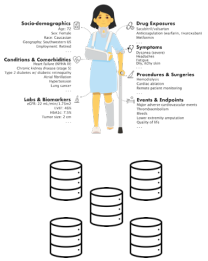
Data sources

- EMR
- Claims
- HIE
- HL7 FHIR Feeds
- Labs & diagnostics
- Trial & research data
- Apps

AI middleware

Droice Hawk

Use case data



No data loss
Regulatory-grade data

- ☑ **Speed** — Fast & scalable
- ☑ **Reproducibility** — Portable & reproducible
- ☑ **Traceability** — Transparency
- ☑ **Quality** — Comprehensive& lossless
- ☑ **Cost** — Cost efficient

Characteristics of an AI middleware:

- Content vs. structure
- Scalability (N, variability)
- Maximizes information, minimizes data loss

Ingredients for the AI middleware:

- Software backbone supporting a cascade of AI algorithms & models
- Data (high noise, high variability)

TRADEOFF: SCALE VS. QUALITY

Quality

Complete
Reproducible
Lineage & provenance
Minimal data loss

Incomplete
Not reproducible
Low traceability
Lossy transformations

Small scale

manual query writing, abstraction, & curation

Low data volume, high quality

- TAs: oncology, rare disease
- Use cases: RCT (SDTM, ADaM), disease registries, …
- Quality assurance depends on manual curation/low patient volumes

Large scale

FHIR / OMOP

manual query writing & ETL-based data pipelines

High data volume, low quality

- TAs: all (e.g., cardiovascular, endocrine, etc.)
- Use cases: interoperability, HEOR, …
- Difficult to ensure data quality

Scale

N = 10-100 | N = 100-1,000 | N = 1,000-10,000 | N = 10,000-100,000 | N = 100,000-1,000,000 | N > 1,000,000

SOLUTION: MAXIMIZE INFORMATION USING AN AI MIDDLEWARE APPROACH

Quality

Trustable, reproducible, with lineage, complete; no information loss

Small scale

manual query writing, abstraction, & curation

Low data volume, high quality

- TAs: oncology, rare disease
- Use cases: RCT (SDTM, ADaM), disease registries, …
- Quality assurance depends on manual curation/low patient volumes

General solution

High data volume, high quality

- TAs: all (e.g., cardiovascular, endocrine, etc.)
- Use cases: Research, regulatory R&D
- Quality assurance via AI middleware/software approach

Large scale

FHIR / OMOP

manual query writing & ETL-based data pipelines

High data volume, low quality

- TAs: all (e.g., cardiovascular, endocrine, etc.)
- Use cases: interoperability, HEOR, …
- Difficult to ensure data quality

Scale

Available at scale

WHY RWD? WHY NOW?

Regulatory push for the use of RWD

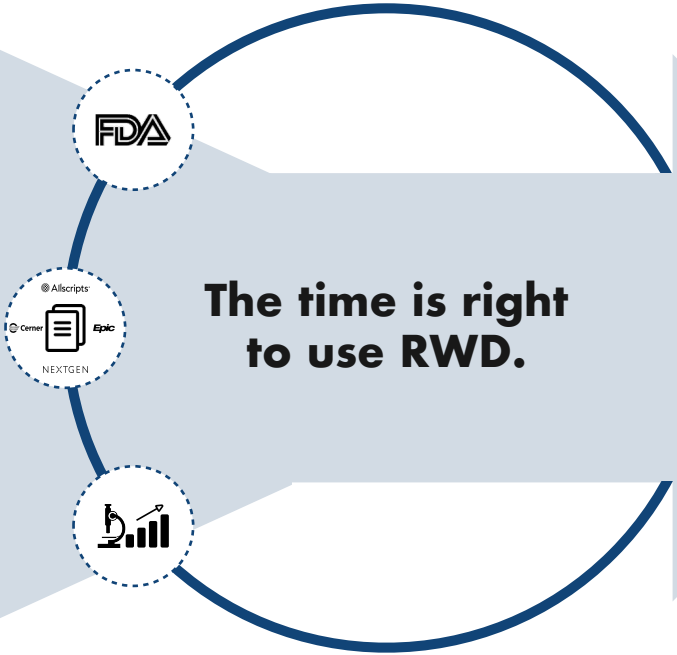
21st Century Cures Act (2016)

Longitudinal patient data availability

96% of US hospitals have adopted EHR

Increasing RWE adoption by the life sciences industry

46% CAGR of published RWE research over last 5 years



The time is right to use RWD.

VALUE

Scientific value

Larger, more representative & diverse patient populations reflect disease in the real-world

Business value

RCT
- Trial design & feasibility
- External control arms
- Primary & secondary evidence for drug approvals

RWE
- Longitudinal outcomes studies
- Quantifying unmet needs
- Label expansion
- Disease registries
- Patient journeys
- …

Thank You!