



EHDEN

EUROPEAN HEALTH DATA & EVIDENCE NETWORK

EHDEN: Data Quality Dashboard

Clair Blacketer

Director, Epidemiology, Janssen R&D
PhD Student, Erasmus MC





What is EHDEN?

What is Data Quality Dashboard (DQD)?

How is the DQD applied across EHDEN?

What are the implications and plans for data quality in EHDEN?



What is EHDEN?

What is Data Quality Dashboard (DQD)?

How is the DQD applied across EHDEN?

What are the implications and plans for data quality in EHDEN?

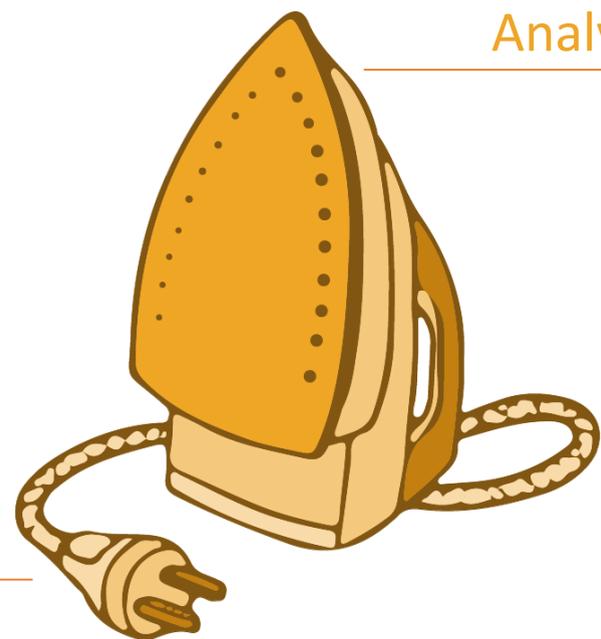


EUROPE: AN OCEAN OF DATA, A DESERT FOR ANALYSIS





WHY IS THIS THE CASE?



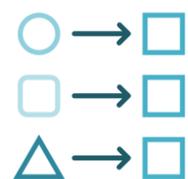
Analytical method

Link to data

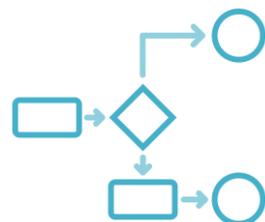
The data...



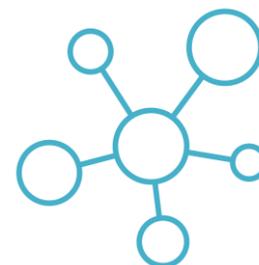
What will it require?



Data interoperability



Standardised analytics



Data network



Strong community



Vision

The European Health Data & Evidence Network (EHDEN) aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care

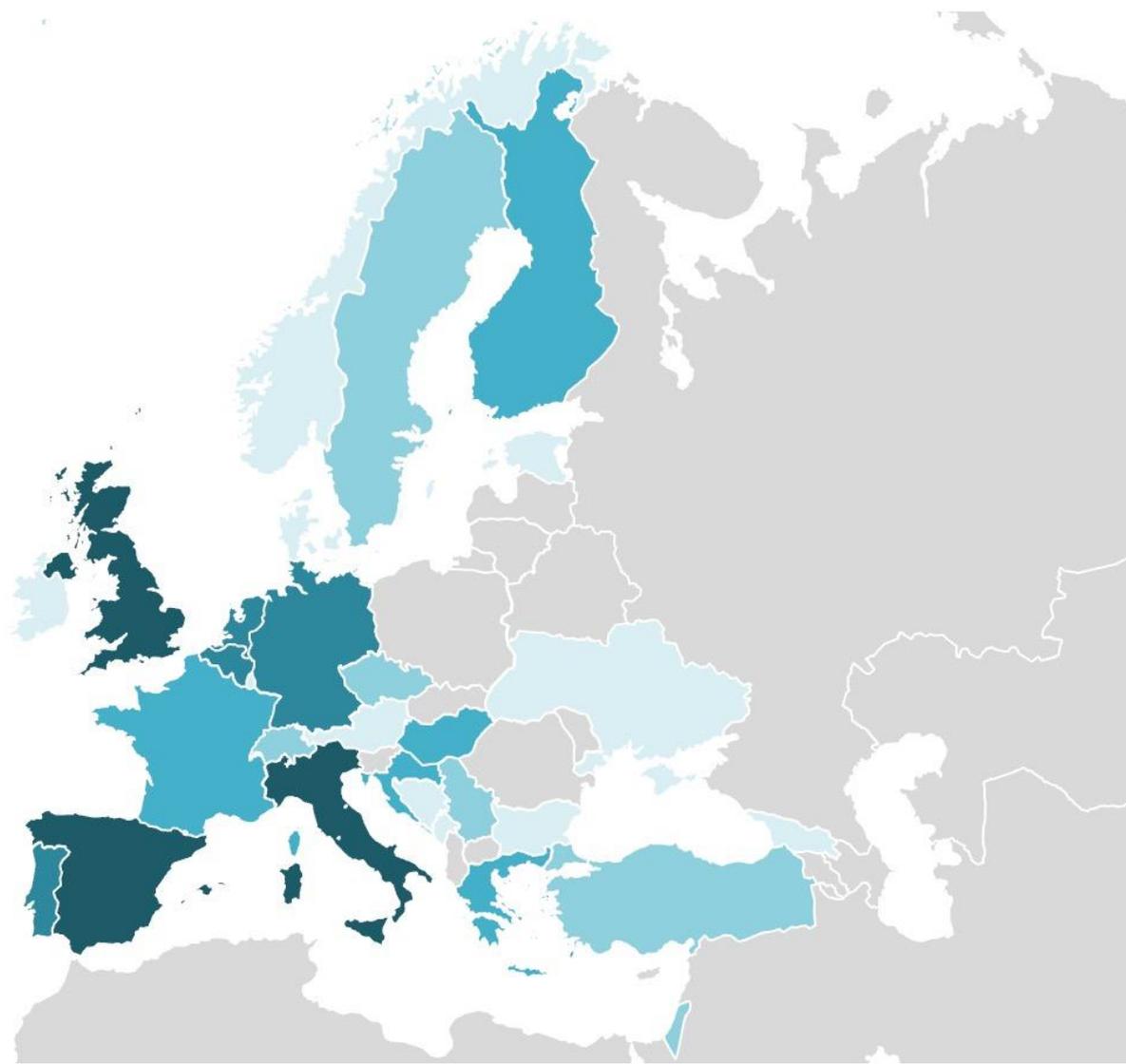
Mission

Our mission is to provide a new paradigm for the discovery and analysis of health data in Europe, by building a large-scale, federated network of data sources standardized to a common data model

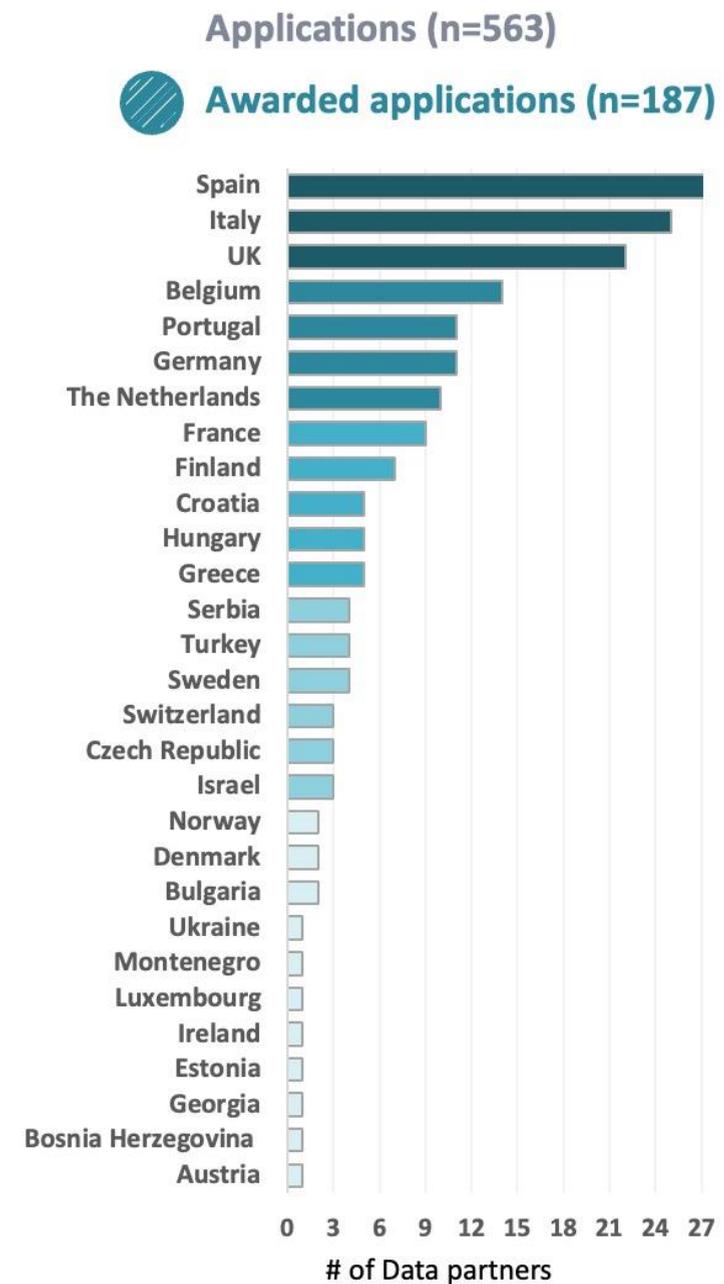


OVERVIEW DATA PARTNER CALLS [AFTER 7 CALLS]

<https://www.ehden.eu/datapartners/>



Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more)



~830 million
anonymous records
being mapped to
OMOP CDM



What is EHDEN?

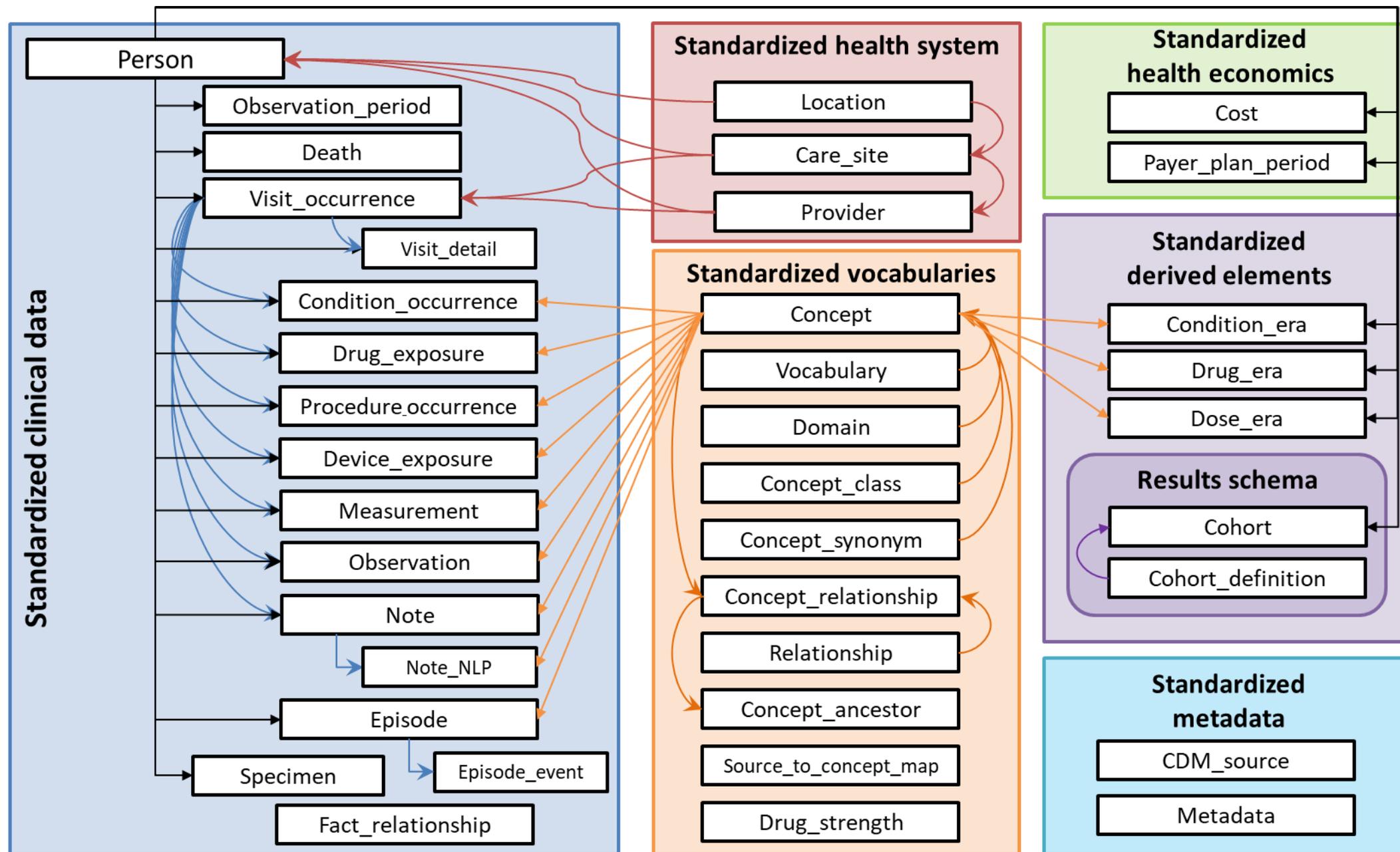
What is Data Quality Dashboard (DQD)?

How is the DQD applied across EHDEN?

What are the implications and plans for data quality in EHDEN?



THE OMOP COMMON DATA MODEL



Patient-centric
Tabular
Extendable
Built for analytics
Relational design

v 5.4



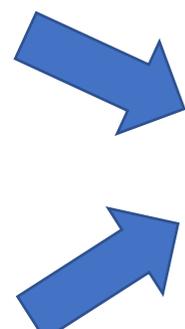
WHAT IS DATA QUALITY?

“...[is] the problem of ensuring the **validity** of the clinical record as a representation of the **true state of the patient.**”

Brennan et al., JAMIA 2000



DATA QUALITY DASHBOARD




IBM MARKETSCAN COMMERCIAL CLAIMS AND ENCOUNTERS DATABASE

- OVERVIEW
- METADATA
- RESULTS**
- ABOUT

RESULTS

IBM MARKETSCAN COMMERCIAL CLAIMS AND ENCOUNTERS DATABASE

Results generated at 2019-09-06 22:20:12 in 7 hours

Show entries

Search:

| | STATUS | CONTEXT | CATEGORY | SUBCATEGORY | LEVEL | DESCRIPTION | % RECORDS |
|--------------------------|--------|--------------|--------------|-------------|---------|--|-----------|
| <input type="checkbox"/> | PASS | Verification | Completeness | None | FIELD | The number and percent of records with a NULL value in the range_high of the MEASUREMENT. (Threshold=100%). | 82.14% |
| <input type="checkbox"/> | PASS | Verification | Completeness | None | FIELD | The number and percent of records with a NULL value in the visit_detail_id of the MEASUREMENT. (Threshold=100%). | 80.90% |
| <input type="checkbox"/> | PASS | Verification | Completeness | None | FIELD | The number and percent of records with a NULL value in the value_source_value of the MEASUREMENT. (Threshold=100%). | 79.89% |
| <input type="checkbox"/> | PASS | Validation | Completeness | None | TABLE | The number and percent of persons in the CDM that do not have at least one record in the DEVICE_EXPOSURE table (Threshold=100%). | 76.70% |
| <input type="checkbox"/> | FAIL | Verification | Plausibility | Atemporal | CONCEPT | For the combination of CONCEPT_ID 3016049 (Testosterone Free [Mass/volume] in Serum or Plasma) and UNIT_CONCEPT_ID 8845 (picogram per milliliter), the number and percent of records that have a value less than 5.00e+00. (Threshold=1%). | 72.43% |

Showing 126 to 130 of 3,351 entries

Previous 1 ... 25 27 ... 671 Next



IBM® MARKETSCAN® MULTI-STATE MEDICAID DATABASE

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

IBM® MARKETSCAN® MULTI-STATE MEDICAID DATABASE

Results generated at 2020-08-24 15:44:34 in 3 hours

| | Verification | | | | Validation | | | | Total | | | |
|--------------|--------------|-----------|-------|--------|------------|----------|-------|--------|-------|-----------|-------|------------|
| | Pass | Fail | Total | % Pass | Pass | Fail | Total | % Pass | Pass | Fail | Total | % Pass |
| Plausibility | 1849 | 6 | 1855 | 100% | 281 | 6 | 287 | 98% | 2130 | 12 | 2142 | 99% |
| Conformance | 550 | 13 | 563 | 98% | 80 | 0 | 80 | 100% | 630 | 13 | 643 | 98% |
| Completeness | 322 | 5 | 327 | 98% | 12 | 0 | 12 | 100% | 334 | 5 | 339 | 99% |
| Total | 2721 | 24 | 2745 | 99% | 373 | 6 | 379 | 98% | 3094 | 30 | 3124 | 99% |



> [J Am Med Inform Assoc. 2021 Sep 18;28\(10\):2251-2257. doi: 10.1093/jamia/ocab132.](#)

Increasing trust in real-world evidence through evaluation of observational data quality

Clair Blacketer ^{1 2}, Frank J Defalco ¹, Patrick B Ryan ^{1 3}, Peter R Rijnbeek ²

Affiliations + expand

PMID: 34313749 PMCID: [PMC8449628](#) DOI: [10.1093/jamia/ocab132](#)

[Free PMC article](#)

Abstract

Objective: Advances in standardization of observational healthcare data have enabled methodological breakthroughs, rapid global collaboration, and generation of real-world evidence to improve patient outcomes. Standardizations in data structure, such as use of common data models, need to be coupled with standardized approaches for data quality assessment. To ensure confidence in real-world evidence generated from the analysis of real-world data, one must first have confidence in the data itself.

Materials and methods: We describe the implementation of check types across a data quality framework of conformance, completeness, plausibility, with both verification and validation. We illustrate how data quality checks, paired with decision thresholds, can be configured to customize data quality reporting across a range of observational health data sources. We discuss how data



What is EHDEN?

What is Data Quality Dashboard (DQD)?

How is the DQD applied across EHDEN?

What are the implications and plans for data quality in EHDEN?

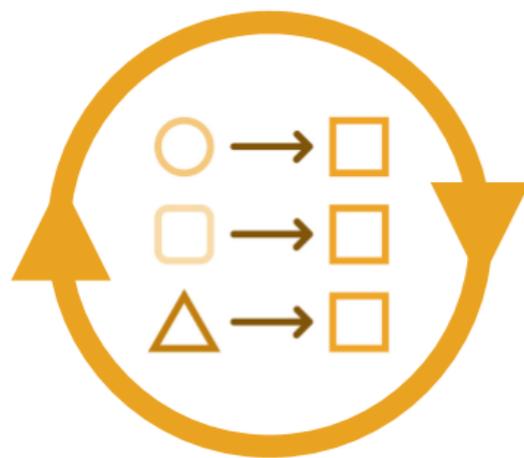


COVID-19 RAPID COLLABORATION CALL



EHDEN COVID-19 Taskforce

Intense collaboration



ETL development &
iterative mapping review



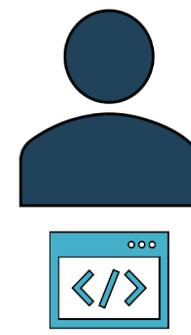
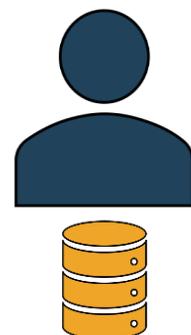
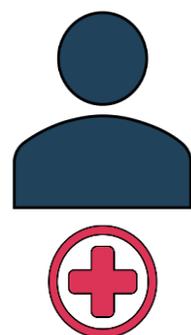
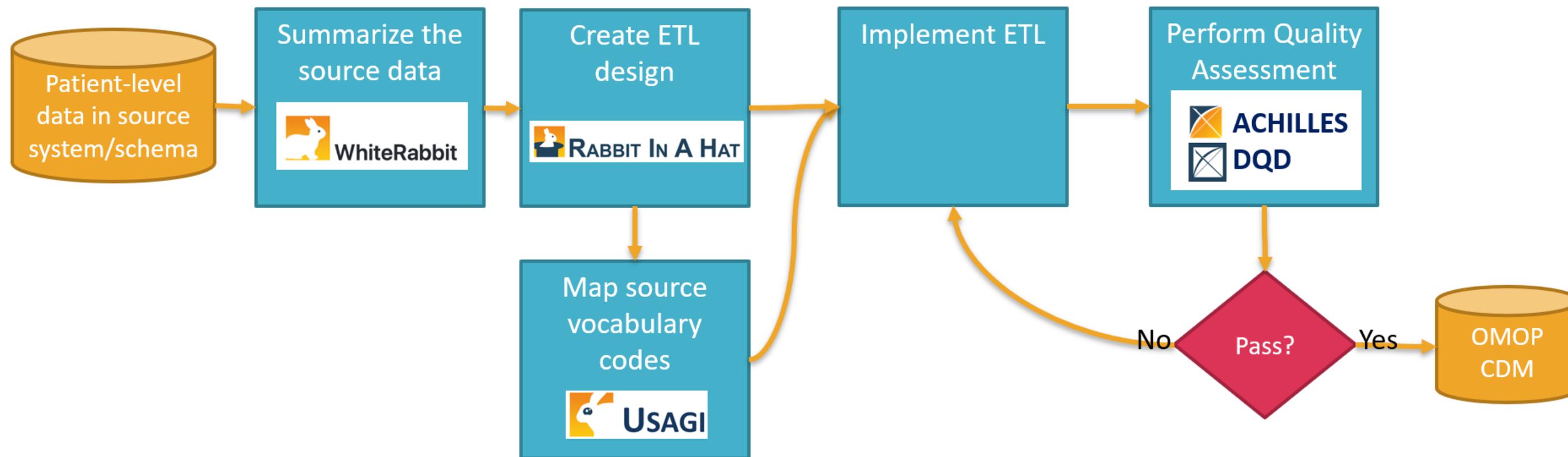
**Engagement in the
EHDEN research network**



Data Partner



EXTRACT, TRANSFORM, & LOAD (ETL)



Medical Knowledge

Data Expertise

Technical ETL Expertise

CDM Experience



applied sciences

Submit to this Journal

Review for this Journal

Edit a Special Issue

Article Menu

Academic Editors



Elena Cardillo



Robert Vander Stichele



Dipak Kalra

Subscribe SciFeed

Recommended Articles

Related Info Link

More by Authors Links

IK

Order Article Reprints



Open Access Article

Using the Data Quality Dashboard to Improve the EHDEN Network

by Clair Blacketer ^{1,2,3,*†} , Erica A. Voss ^{1,2,3,*†} , Frank DeFalco ^{1,3} ,
 Nigel Hughes ^{1,3} , Martijn J. Schuemie ^{1,3} , Maxim Moinat ^{2,3,4} and
 Peter R. Rijnbeek ^{2,3}

¹ Janssen Pharmaceutical Research and Development LLC, Titusville, NJ 08560, USA

² Department of Medical Informatics, Erasmus University Medical Center, 3015 GD Rotterdam, The Netherlands

³ OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York 10027, NY, USA

⁴ The Hyve, 3511 MJ Utrecht, The Netherlands

* Authors to whom correspondence should be addressed.

† Co-first authors.

Appl. Sci. **2021**, *11*(24), 11920; <https://doi.org/10.3390/app112411920>

Received: 31 October 2021 / Revised: 7 December 2021 / Accepted: 8 December 2021 /

Published: 15 December 2021

(This article belongs to the Special Issue **Semantic Interoperability and Applications in Healthcare**)

Download

Browse Figures

Review Reports

Versions Notes

Abstract

Federated networks of observational health databases have the potential to be a rich resource to inform clinical practice and regulatory decision making. However, the lack of standard data quality processes makes it difficult to know if these data are research ready. The EHDEN COVID-19 Rapid Collaboration Call presented the opportunity to



What is EHDEN?

What is Data Quality Dashboard (DQD)?

How is the DQD applied across EHDEN?

What are the implications and plans for data quality in EHDEN?



ACADEMIC/SCIENTIFIC IMPACT: ~50 METHODOLOGICAL & SCIENTIFIC PUBLICATIONS

academic.oup.com/jcem

COVID-19 – Support for our author and subscriber community

Body Mass Index and Risk of COVID-19 Diagnosis, Hospitalization, and Death: A Cohort Study of 2 524 926 Catalans **FREE**

Martina Recalde, et al.

A comprehensive understanding of the association between body mass index (BMI) and coronavirus disease 2019 (COVID-19) is still lacking. The objective of this study was to investigate associations between BMI and risk of COVID-19 diagnosis, hospitalization with COVID-19, and death after a COVID-19 diagnosis or hospitalization (subsequent death), accounting for potential effect modification by age and sex.

Latest Issue
Volume 106, Issue 12
December 2021

ijc33846.pdf

RESEARCH

OPEN ACCESS

For numbered affiliations see end of article.
Correspondence to: P Ryan
ryan@ohdsi.org
<http://orcid.org/0000-0002-9727-2138>
Additional material is published online only. To view please visit the journal online
Cite this as: *BMJ* 2021;373:n1038
<http://dx.doi.org/10.1136/bmj.n1038>
Accepted: 16 April 2021

Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study

Albert Prats-Urbe,¹ Anthony G Sena,^{2,3} Lana Yin Hui Lai,⁴ Waheed-Ul-Rahman Ahmed,^{5,6} Heba Alghoul,⁷ Osaid Alser,⁸ Thamir M Alshammari,⁹ Carlos Areia,¹⁰ William Carter,¹¹ Paula Casajust,¹² Dalia Dawoud,¹³ Asieh Golzar,^{15,16} Jitendra Jonnagaddala,¹⁷ Paras P Mehta,¹⁸ Mengchun Gong,¹⁹ Daniel R Morales,²⁰ Fredrik Nyberg,²² Jose D Posada,²³ Martina Recalde,^{24,25} Elena Roel,^{24,25} Karishma Shah,⁵ Nigam H Shah,²³ Lisa M Schilling,¹¹ Vignesh Subbian,²⁶ David Vizcaya,²⁷ Lin Zhang,^{28,29} Ying Zhang,¹⁹ Hong Zhu,³⁰ Li Liu,³⁰ Jaehyeong Cho,³¹ Kristine E Lynch,³² Michael E Matheny,^{33,34} Seng Chan You,³⁵ Peter R Rijnbeek,³ George Hripcsak,³⁶ Jennifer CE Lane,⁵ Edward Burn,^{1,24} Christian Reich,³⁷ Marc A Suchard,³⁸ Talita Duarte-Salles,²⁴ Kristin Kostka,^{37,39} Patrick B Ryan,^{2,40} Daniel Prieto-Alhambra¹

BMI in COVID-19: JCEM pick of the week!!

RESEARCH: SPECIAL PAPER

OPEN ACCESS

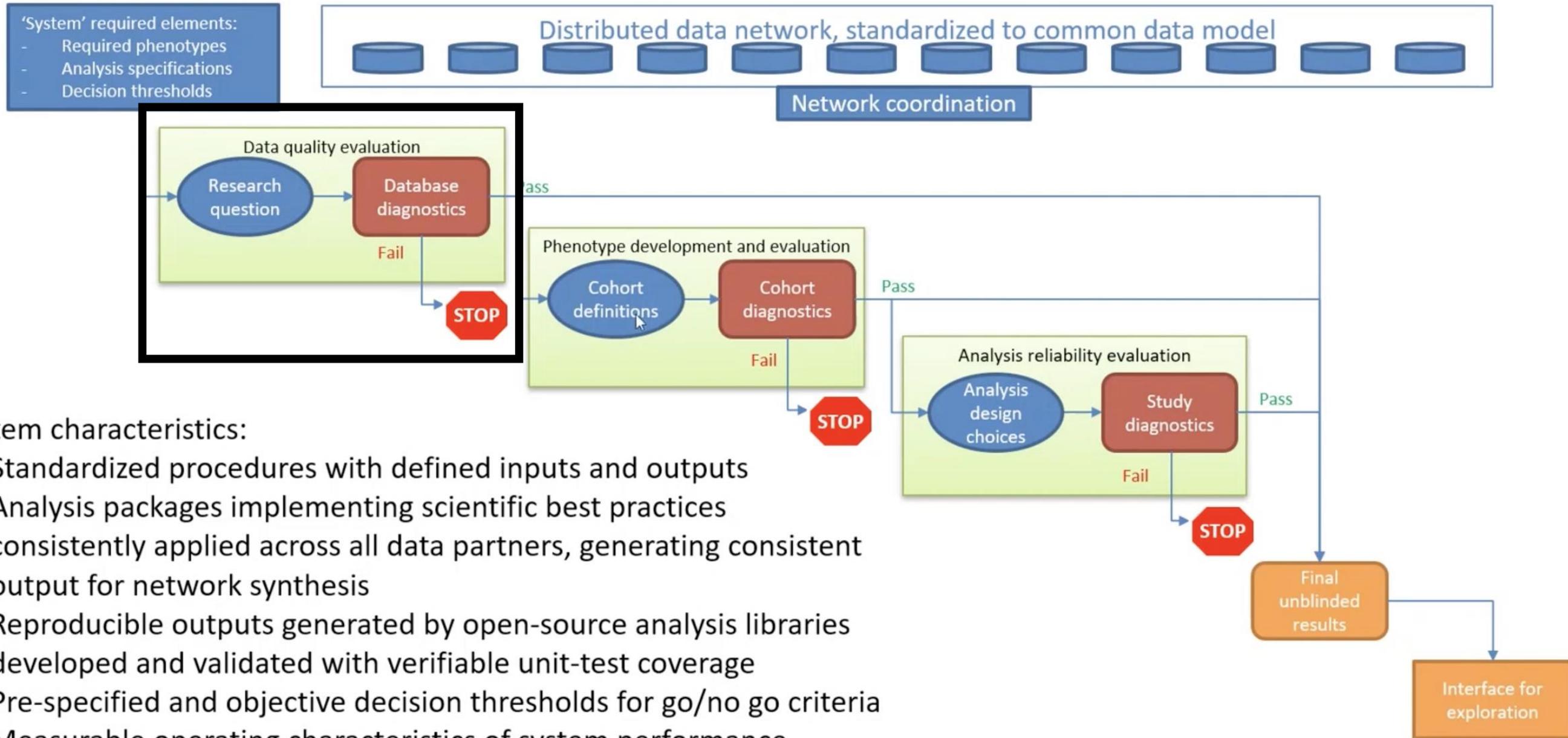
FAST TRACK

Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study

Xintong Li,¹ Anna Ostropolets,² Rupa Makadia,³ Azza Shoaibi,³ Gowtham Rao,³ Anthony G Sena,^{3,6} Eugenia Martinez-Hernandez,⁴ Antonella Delmestri,¹ Katia Verhamme,^{6,7} Peter R Rijnbeek,⁶ Talita Duarte-Salles,⁵ Marc A Suchard,^{8,9} Patrick B Ryan,^{2,3} George Hripcsak,² Daniel Prieto-Alhambra^{1,6}



Engineering Open Science Systems that Build Trust into the Evidence Generation and Dissemination Process



System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance



FINAL THOUGHTS

- We are building one of the largest international collaborative research networks on the planet within EHDEN and OHDSI
- Each member of the network is required to adhere to certain expectations regarding data quality, as defined and evaluated by the community through the Data Quality Dashboard
- These data quality procedures give us confidence in the data in the network, and, in turn, the evidence we are generating
- Ongoing methodological work continues to contribute to our understanding of the quality of the network, from the source data to the terminology and standard mapping conventions
- Many of the EHDEN partners will also contribute to the DARWIN initiative, this consistency EHDEN is enforcing will only benefit DARWIN



www.ehden.eu



[@IMI_EHDEN](https://twitter.com/IMI_EHDEN)



[IMI_EHDEN](https://www.linkedin.com/company/IMI_EHDEN)



github.com/EHDEN



This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.