

www.ispor.org



Assessing Real-Word Data Quality From Electronic Health Records for Health Technology Assessment

ISPOR 2023
Boston, MA, USA

**Presented by the ISPOR Using EHRs in
HTA Task Force**

Antitrust Compliance Statement

- ISPOR has a policy of strict compliance with both United States, and other applicable international antitrust laws and regulations.
- Antitrust laws prohibit competitors from engaging in actions that could result in an unreasonable restraint of trade.
- ISPOR members (and others attending ISPOR meetings and/or events) must avoid discussing certain topics when they are together including, prices, fees, rates, profit margins, or other terms or conditions of sale.
- Members (and others attending ISPOR meetings and/or events) have an obligation to terminate any discussion, seek legal counsel's advice, or, if necessary, terminate any meeting if the discussion might be construed to raise antitrust risks.
- The Antitrust policy is available on the ISPOR website.

Moderator:

Rachael Fleurence, PhD, Senior Advisor, National Institutes of Health, Bethesda, MD, USA

Speakers:

- **Scott Ramsey, PhD, MD**, Full Member & Co-Director, Fred Hutchinson Cancer Research Center, Lake Forest Park, WA, USA
- **Seamus Kent, PhD**, Senior Adviser, HTA and Market Access, Flatiron Health, Amsterdam, the Netherlands
- **Blythe Adamson, PhD, MPH**, Principal Scientist, Flatiron Health, New York, NY, USA

Agenda

1. Task Force overview and rationale (Dr Scott Ramsay)
2. The complexities of EHR data (Dr Blythe Adamson)
3. Understanding the suitability of real-world data to answer questions in HTA (Dr Seamus Kent)
4. Elements and format under consideration for the recommendations (Dr Blythe Adamson)
5. Q&A moderated by Dr Rachael Fleurence

Members of the task force - 1

- **Rachael Fleurence, PhD**, (Co-Chair), Senior Advisor, National Institutes of Health, USA
- **Scott Ramsey, PhD, MD**, (Co-Chair), Full Member & Co-Director, Fred Hutchinson Cancer Research Center, USA
- **Blythe Adamson, PhD, MPH**, Principal Scientist, Flatiron Health, USA
- **Ran Balicer, PhD, MD**, Chief Innovation Officer, Clalit Health Services, Tel Aviv, Israel
- **Elsa Bouee Benhamiche, PharmD, RPh**, Project manager, Institut National du Cancer, France
- **Jon Campbell, PhD**, Senior Vice President, Health Economics, Institute for Clinical and Economic Review (Seamus Kent, PhD, Senior Adviser, HTA and Market Access, Flatiron Health, the Netherlands)

Members of the task force - 2

- **Sebastian Garcia Marti, MSc, MD**, HTA and Health Economics Coordinator, IECS - Institute for Clinical Effectiveness, Argentina
- **Kevin Haynes, MS, PharmD**, Associate Director of Epidemiology, Janssen Research and Development, LLC, USA
- **Seamus Kent, PhD, MSc**, Senior Adviser, HTA and Market Access, Flatiron Health, Amsterdam, the Netherlands
- **Patrick Muller, MSc, PhD**, National Institute for Health and Care Excellence (NICE), UK
- **Joseph Ross, MD, PhD**, Professor of Medicine and Public Health, Yale University, USA
- **James Tcheng, MD**, Professor of Medicine & Professor of Family Medicine and Community Health (Informatics), Duke University School of Medicine, USA

SECTION

1

Motivation of ISPOR Task Force on EHR Data for HTA

Background and Objectives

- Background
 - Rapid growth in availability of EHR -derived RWD and organizations devoted to providing and analyzing EHR's for industry and regulators
 - Lack of frameworks for use of EHR derived RWD for Health Technology Assessment (HTA)
 - Need for users to understand the similarities and unique aspects of EHR relative to other types of RWD
- Overall Objective
 - Establish consensus on and provide emerging good practices for conducting, reporting and evaluating data quality of EHRs for health technology assessments
- Key elements:
 - Targeted literature review of standards for regulatory decision making
 - Recommendations and Data Quality Checklist
 - Limitations of EHR's and future directions

Feasibility assessment: What data is best fit for this question?

	Claims	Registries	Prior Clinical Trials Data	EHR-derived Data
Patient Population(s) Included	<p>+</p> <p>Typically, very large broad patient populations</p>	<p>+/-</p> <p>Population defined by particular disease, condition, or exposure</p>	<p>+ / -</p> <p>Disease and drug specific cohorts, unlikely representative</p>	<p>+ / -</p> <p>May be broad and/or disease-specific cohorts</p>
Clinical Depth	<p>-</p> <p>Limited granularity (e.g., biomarker test performed, but not result)</p>	<p>+ / -</p> <p>Collects uniform data on patients; range of clinical depth possible</p>	<p>+ / -</p> <p>Collects uniform data on patients; range of clinical depth possible</p>	<p>+</p> <p>Usually able to leverage all structured and unstructured data from patients' charts</p>
Completeness/Data Quality	<p>+</p> <p>Full visibility into events across full healthcare ecosystem (not limited to site of care). Need to consider open/closed claim tradeoffs</p>	<p>+ / -</p> <p>Serves more predetermined scientific, clinical or policy purpose → informs data collection & completeness</p>	<p>+ / -</p> <p>Serves more predetermined scientific, clinical or policy purpose → informs data collection & completeness</p>	<p>+ / -</p> <p>Visibility may be limited to specific sites of care and information transferred from other sites</p>

There are multiple consortia with organizations in the private and public sectors focused on evaluating EHR data...

REAL-WORLD EVIDENCE ALLIANCE

We are a coalition of real-world data and analytics organizations with a common interest in harnessing the power of real-world evidence to inform regulatory decision making to improve patients' lives.

AETION

Syapse™

 Concerta

Syneos[®]
Health

 flatiron.

"TEMPUS

 IQVIA

 Verana Health.®

OM1[®] 

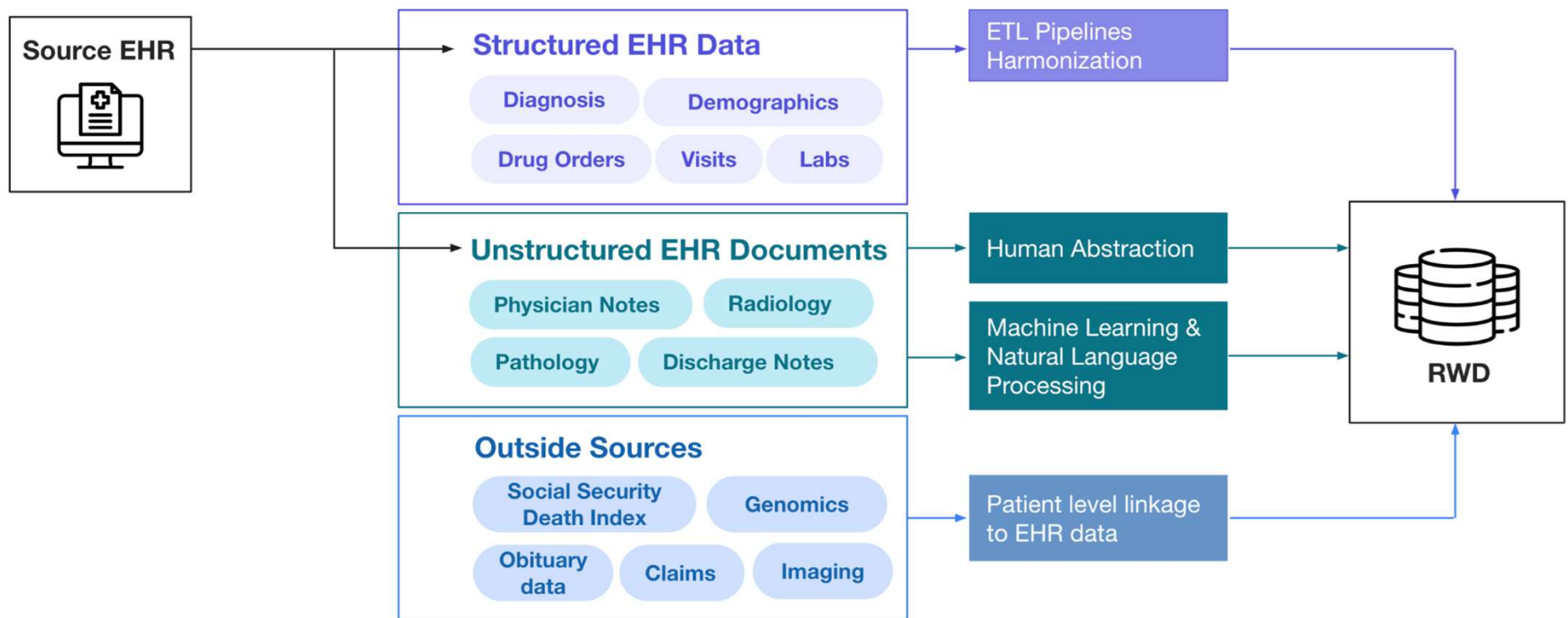
verily

SECTION

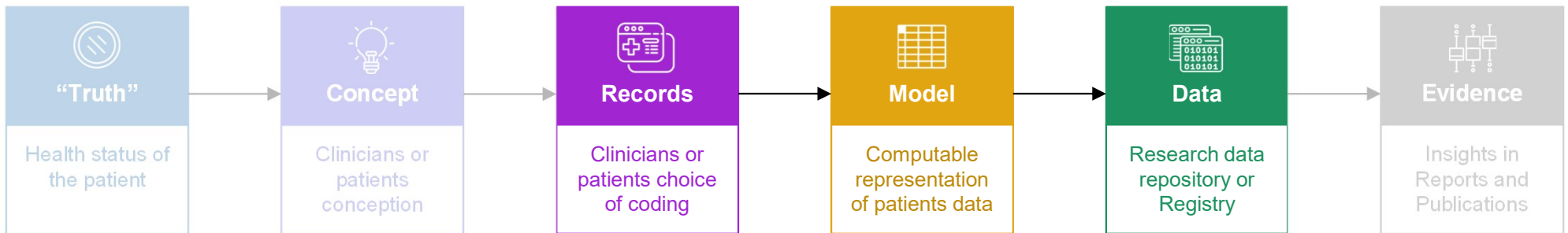
2

The Complexities of EHR Data

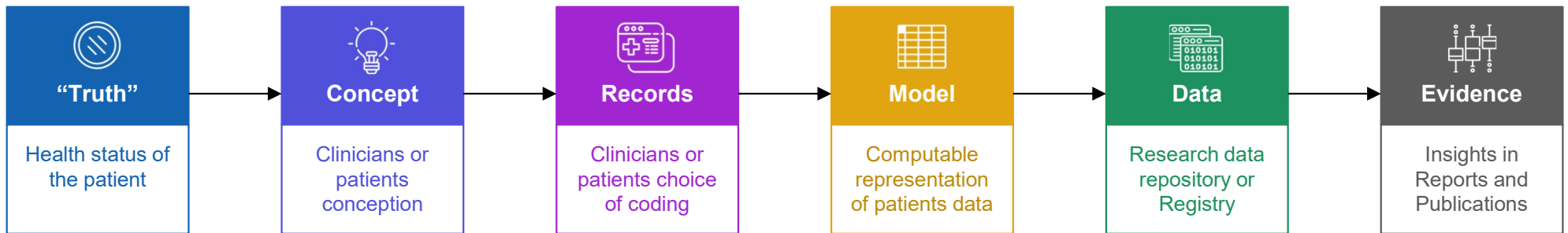
Electronic Health Record (EHR) source data requires curation of structured data and unstructured documents



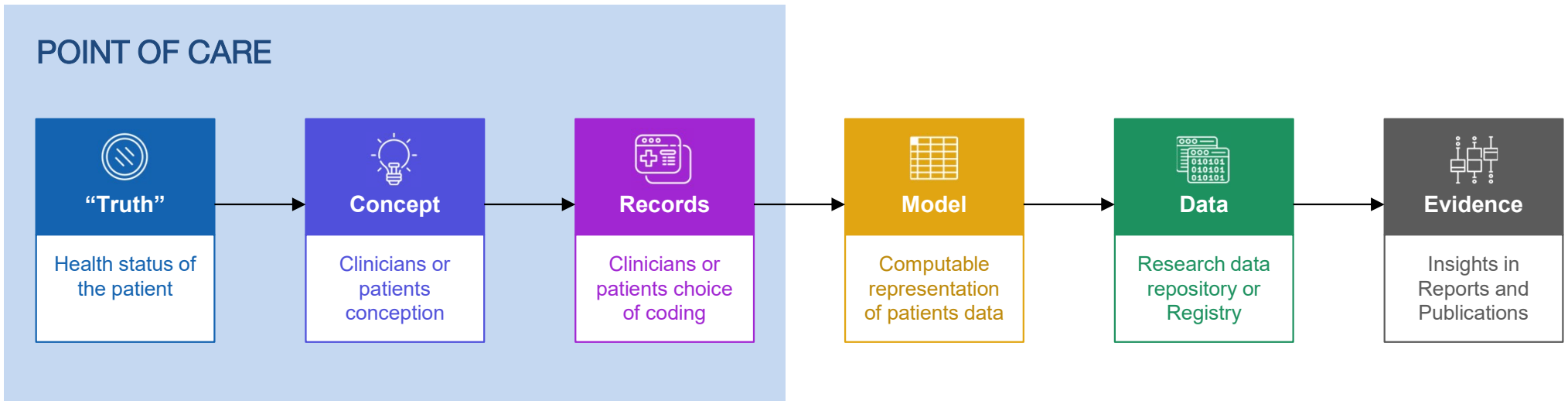
Process of Health Status to Clinical Data Documentation, Extraction, and Transformation for Research



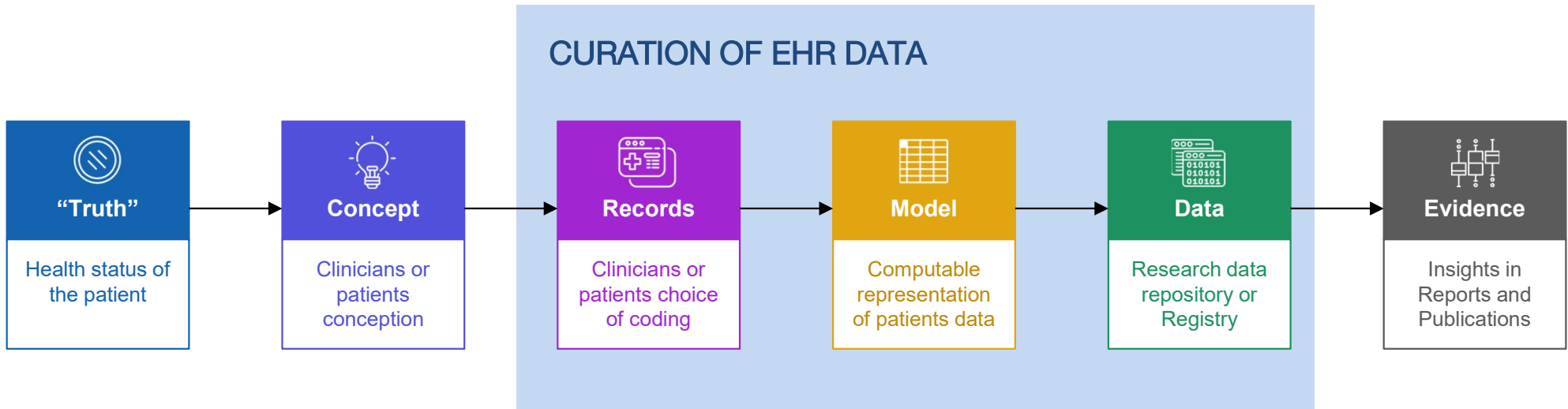
Process of Health Status to Clinical Data Documentation, Extraction, and Transformation for Research



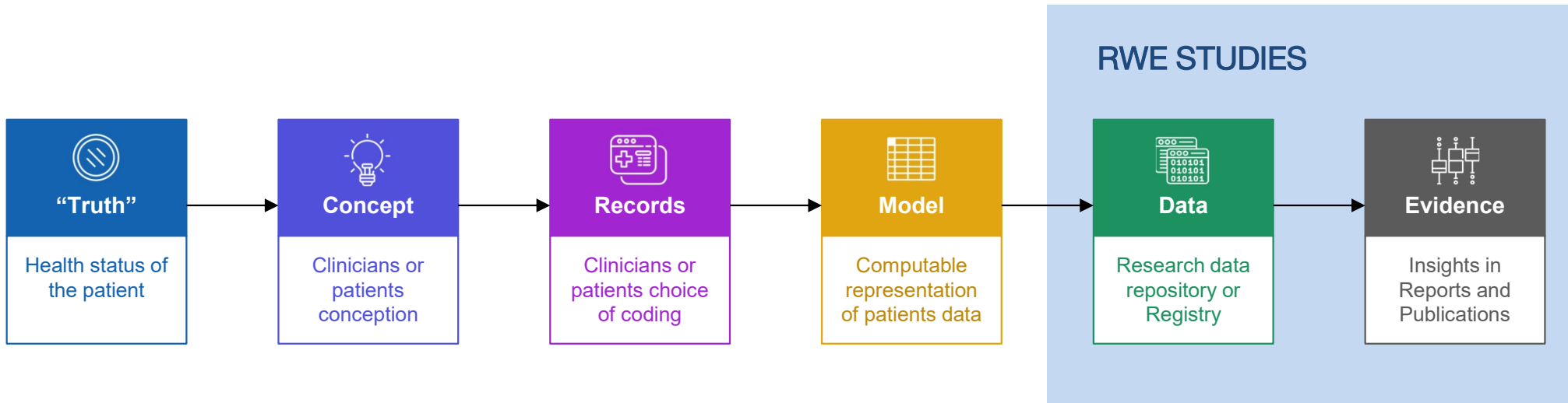
Process of Health Status to Clinical Data Documentation, Extraction, and Transformation for Research



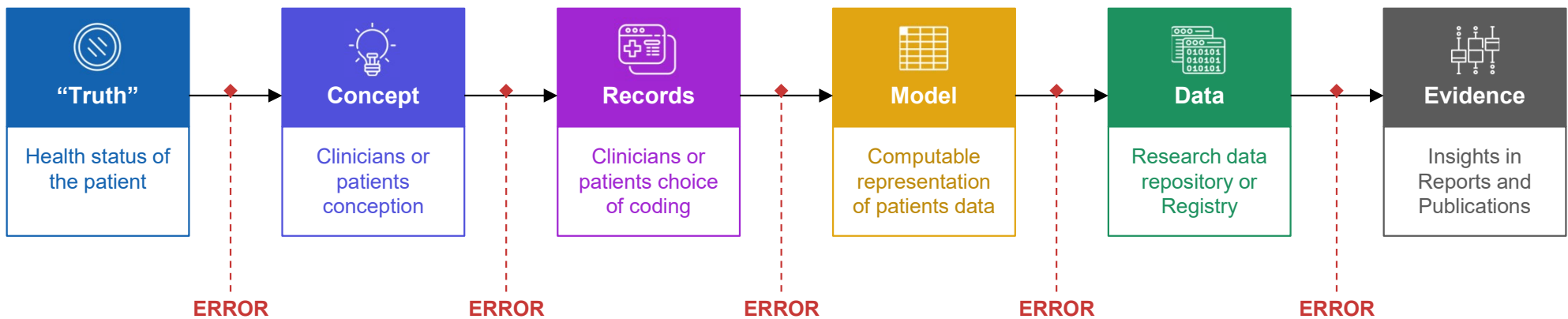
Process of Health Status to Clinical Data Documentation, Extraction, and Transformation for Research



Process of Health Status to Clinical Data Documentation, Extraction, and Transformation for Research



Process of Health Status to Clinical Data Documentation, Extraction, and Transformation for Research



SECTION

3

Understanding the suitability of real-world data to answer questions in HTA

Seamus Kent, PhD, Senior Adviser, HTA and
Market Access, Flatiron Health, the Netherlands

We are building on existing work

Foundational data quality

Automating Electronic Health Record Data Quality Assessment

[Obinwa Ozone](#), [Philip J. Scott](#) & [Adrian A. Hoggood](#) 

Journal of Medical Systems 47, Article number: 23 (2023) | [Cite this article](#)

EGEMS (Wash DC), 2016; 4(1): 1244.

PMCID: PMC5051581

Published online 2016 Sep 11. doi: [10.13063/2327-9214.1244](https://doi.org/10.13063/2327-9214.1244)

PMID: [27713905](https://pubmed.ncbi.nlm.nih.gov/27713905/)

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

[Michael G. Kahn](#), MD, PhD,ⁱ [Tiffany J. Callahan](#), MPH,ⁱ [Juliana Barnard](#), MA,ⁱ [Alan E. Bauck](#),ⁱⁱ [Jeff Brown](#), PhD,ⁱⁱⁱ [Bruce N. Davidson](#), PhD,^{iv} [Hossein Estiri](#), PhD,^v [Carsten Goerg](#), PhD,ⁱ [Erin Holve](#), PhD, MPH, MPP,^{vi} [Steven G. Johnson](#), MS,^{vii} [Siaw-Teng Liaw](#), MBBS, PhD, FRACGP, FACHI,^{viii} [Marianne Hamilton-Lopez](#), PhD, MPA,^{ix} [Daniella Meeker](#), PhD,^x [Toan C. Ong](#), PhD,^{xi} [Patrick Ryan](#), PhD,^{xii} [Ning Shang](#), PhD,^{xiii} [Nicole G. Weiskopf](#), PhD,^{xiv} [Chunhua Weng](#), PhD, FACMI,^{xiii} [Meredith N. Zozus](#), PhD,^{xv} and [Lisa Schilling](#), MD^{xi}

Fitness for purpose / Policy



We are building on existing work

Foundational data quality

Automating Electronic Health Record Data Quality Assessment

[Obinwa Ozone](#), [Philip J. Scott](#) & [Adrian A. Hoggood](#) 

Journal of Medical Systems 47, Article number: 23 (2023) | [Cite this article](#)

EGEMS (Wash DC), 2016; 4(1): 1244.

PMCID: PMC5051581

Published online 2016 Sep 11. doi: [10.13063/2327-9214.1244](https://doi.org/10.13063/2327-9214.1244)

PMID: [27713905](https://pubmed.ncbi.nlm.nih.gov/27713905/)

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

[Michael G. Kahn](#), MD, PhD,ⁱ [Tiffany J. Callahan](#), MPH,ⁱ [Juliana Barnard](#), MA,ⁱ [Alan E. Bauck](#),ⁱⁱ [Jeff Brown](#), PhD,ⁱⁱⁱ [Bruce N. Davidson](#), PhD,^{iv} [Hossein Estiri](#), PhD,^v [Carsten Goerg](#), PhD,ⁱ [Erin Holve](#), PhD, MPH, MPP,^{vi} [Steven G. Johnson](#), MS,^{vii} [Siaw-Teng Liaw](#), MBBS, PhD, FRACGP, FACHI,^{viii} [Marianne Hamilton-Lopez](#), PhD, MPA,^{ix} [Daniella Meeker](#), PhD,^x [Toan C. Ong](#), PhD,^{xi} [Patrick Ryan](#), PhD,^{xii} [Ning Shang](#), PhD,^{xiii} [Nicole G. Weiskopf](#), PhD,^{xiv} [Chunhua Weng](#), PhD, FACMI,^{xiii} [Meredith N. Zozus](#), PhD,^{xv} and [Lisa Schilling](#), MD^{xi}

Summary

- Tools are focused on understanding inherent quality of the data and data processing/curation
- Agnostic to research question
- Large number of distinct frameworks with substantial variation in definitions and nomenclature
- Frameworks often specific to CDMs
- **Key concepts: completeness, conformance, plausibility, consistency, accuracy, timeliness**
- Can be assessed internally or using external data

We are building on existing work

Policy frameworks

NICE
National Institute for Health and Care Excellence

HAS
HAUTE AUTORITÉ DE SANTÉ



IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Institute for Quality and Efficiency in Health Care

Duke
MARGOLIS CENTER
for Health Policy



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Use of Real-World Data and Real-World Evidence to Support Drug Reimbursement Decision-Making in Asia.

A non-binding guidance document prepared by the REAL World Data in Asia for Health Technology Assessment in Reimbursement (REALISE) working group

Clinical Pharmacology & Therapeutics / Volume 111, Issue 1 / p. 122-134

Review | [Open Access](#) | 

The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework

Nicolle M. Gatto, Ulka B. Campbell, Emily Rubinstein, Ashley Jaksa, Pattra Mattox, Jingping Mo, Robert F. Reynolds

Summary

- Differentiate between data provenance and data fitness for purpose
- Focus on fitness for purpose in relation to a specific research question
- Distinguish between data reliability and data relevance when assessing fitness for purpose
- Some frameworks agnostic to data source (e.g., NICE); others focused on specific sources of data (e.g., FDA, IQWiG)

How are we building on this work?

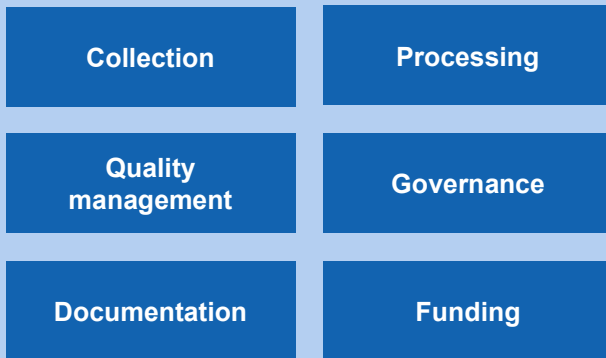
- Identify common strands from existing frameworks – strive for international alignment
- Ensure relevance to HTA needs and HTA evaluation processes
- Focus on user & application – developers of evidence to meet HTA needs
- Be specific to challenges and opportunities of using EHR data
- Develop useable and introductory guide for those less familiar with EHR data

Proposed framework

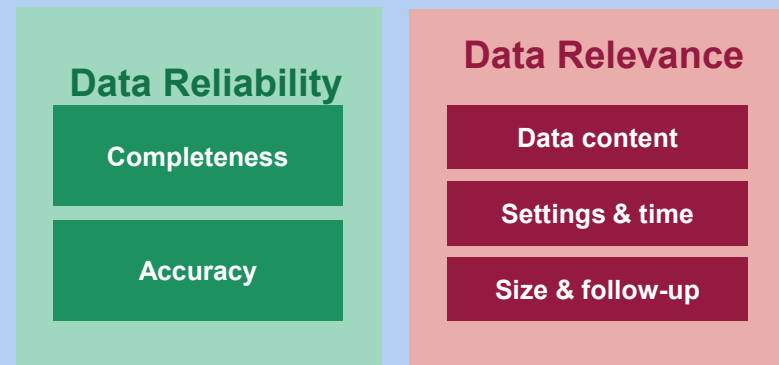
Data must be trustworthy, enable internally valid estimates, and be relevant to the decision context.

Proportionate risk-based approach depending on use case and decision context.

Data Provenance



Data Fitness for Purpose



Data provenance

Is it reasonable to use the data to inform HTA decisions?

Challenges to using EHR data

- Data processing is a complex process, potentially involving multiple parties
- Learning from unstructured data contained within EHR records
- Ensuring appropriate permissions for secondary use of data, which varies across countries
- Data networks may need to integrate data from different systems

Questions to understand data provenance?

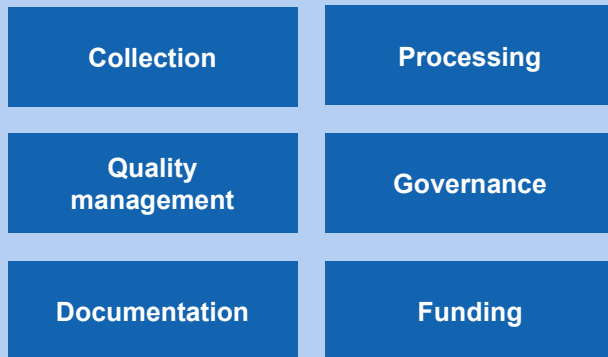
- What was the initial purpose of data collection?
- How has data been processed (incl. linkage, transformations, etc.)?
- Is documentation sufficient for reviewers to fully understand the data and its processing? Are steps in place to share proprietary software where necessary?
- What quality management processes are in place to ensure the integrity of the data? How is quality assessed over the data lifecycle?
- What is the legal basis for the secondary use of the EHR data?
- What data governance processes are in place?

Proposed framework

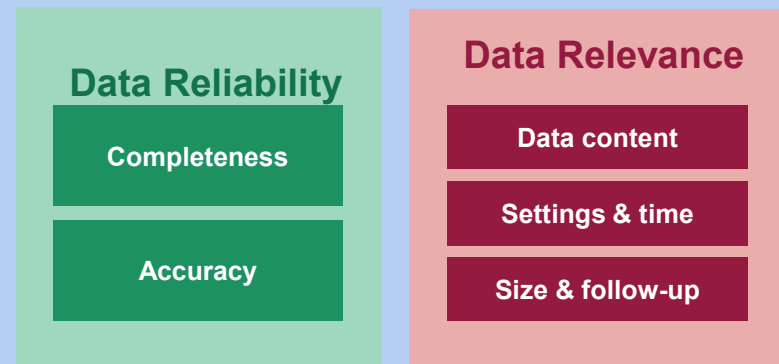
Data must be trustworthy, enable internally valid estimates, and be relevant to the decision context.

Proportionate risk-based approach depending on use case and decision context.

Data Provenance



Data Fitness for Purpose



Data reliability

Is the data reliable enough to produce internally valid estimates for a given research question?

Challenges to using EHR data

- Substantial missing data from:
 - Partial reporting of full health journey
 - Missing clinical/other events and results
- Data may not be recorded accurately
- Delays in integrating data into EHR systems
- Complex data processing

How can we understand data quality?

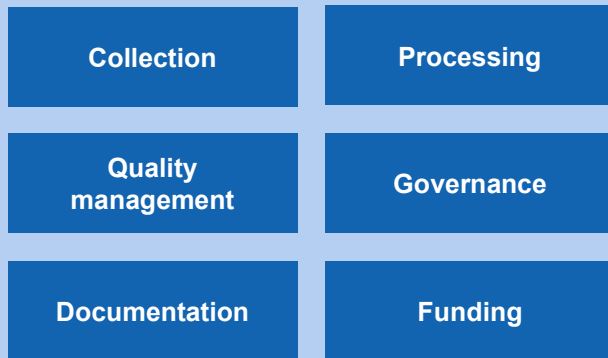
- Understand amount, patterns, and reasons for missing data
- Present quantitative data on data accuracy for all key study variables
- Metric depends on variable – e.g., categorical or continuous
- For accuracy of endpoints validation should ideally compare data against a known gold standard
- Where not feasible, alternative approaches include assessing plausibility, consistency, and conformance
- Address data limitations using appropriate study design and statistical methods
- Use sensitivity/bias analysis to assess the potential impact on results or adjust

Proposed framework

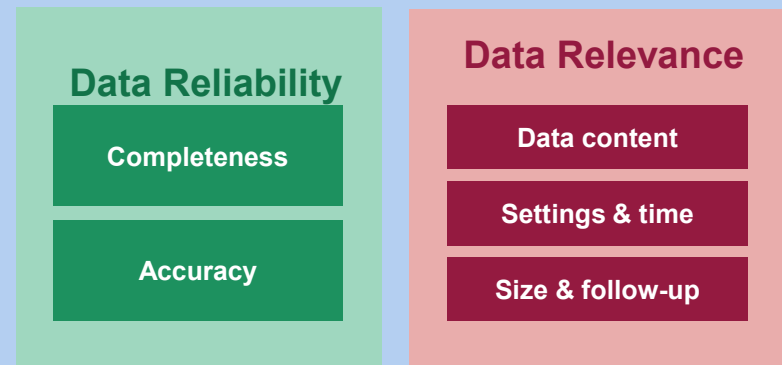
Data must be trustworthy, enable internally valid estimates, and be relevant to the decision context.

Proportionate risk-based approach depending on use case and decision context.

Data Provenance



Data Fitness for Purpose



Data relevance

Does the data allow the research question to be answered?

Challenges to using EHR data

- Data items reported reflect what is needed for clinical and administrative purpose rather than for research
- Data may come from particular care providers, regions, or countries and may not be relevant to target population
- Delays in access may limit relevance of the data to current treatment and limit follow-up of data.

How can we understand data relevance?

- Are all required data elements collected and at the right level of granularity?
- Is data collected at relevant time points?
- Is the population similar to the intended target population?
- Are care settings and treatment patterns relevant to the target country?
- Does the study period reflect current clinical practice and outcomes?
- Are sample size and length of follow-up sufficient to answer the research question?

SECTION

4

Elements and format under consideration for the recommendations

Blythe Adamson, PhD, MPH,
Senior Principal Scientist
Flatiron Health, USA

What format and style of checklist would be most useful?

CHEERS section/item	Item No	References									
		[25]	[26]	[27]	[28]	[29]	[30]	[24]	[31]	[32]	
Title and abstract											
Title	1	Y	Y	Y	Y	Y	Y	N	Y	Y	
Abstract	2	Y	Y	Y	Y	Y	Y	P	P	P	
Introduction											
Background and objectives	3	Y	Y	Y	Y	Y	Y	P	Y	Y	
Methods											
Target population and subgroups	4	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Setting and location	5	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Study perspective	6	N	N	N	Y	Y	Y	Y	Y	N	
Comparators	7	Y	Y	Y	Y	Y	Y	P	NA	Y	
Time horizon	8	Y	Y	Y	Y	Y	Y	P	Y	N	
Discount rate	9	N	P	N	N	Y	Y	NA	NA	N	
Choice of health outcomes	10	Y	Y	Y	Y	Y	Y	P	Y	P	
Measurement of effectiveness (single study-based estimates)	11a	NA	NA	NA	NA	NA	NA	NA	N	P	
Measurement of effectiveness (synthesis-based estimates)	11b	P	P	P	P	Y	P	P	P	NA	
Measurement and valuation of preference based outcomes	12	Y	Y	Y	Y	Y	NA	N	NA	P	
Estimating resources and costs (single study-based economic evaluation)	13a	NA	NA	NA	NA	NA	NA	NA	NA	P	
Estimating resources and costs (model-based economic evaluation)	13b	P	P	P	P	Y	P	P	P	NA	
Currency, price date, and conversion	14	Y	Y	Y	Y	Y	Y	Y	N	Y	
Choice of model	15	Y	Y	Y	Y	Y	P	P	N	NA	
Assumptions	16	Y	Y	Y	Y	Y	Y	Y	N	N	
Analytical methods	17	P	P	P	P	P	P	P	P	P	
Results											
Study parameters	18	Y	Y	Y	Y	Y	Y	Y	P	Y	
Incremental costs and outcomes	19	Y	Y	Y	Y	Y	P	Y	P	Y	

(Continued)

CHEERS

Sector	Type of Impact (Categories impacted within each sector with unit of measure if relevant)	Included in Analysis from the following perspective?			Notes on Sources of Evidence
		Patient	Healthcare Sector	Societal	
FORMAL HEALTHCARE SECTOR					
HEALTH	Health Outcomes (Effects):				Appendix Table 15.8 See Footnote A
	➤ Longevity effects, Years	✓	✓	✓	
	➤ Health-related quality of life (HRQoL), QALYs	✓	✓	✓	
	➤ Years in Alcohol Use Disorders (AUD), Years	✓	✓	✓	
	➤ Disutility due to adverse events from treatment, QALYs	✓	✓	✓	
	➤ Disutility of being incarcerated, QALYs	✓	✓	✓	
	➤ Spillover HRQoL, caregiver		×	×	
	Medical Costs:				
	➤ Paid for by third-party payers, \$		✓	✓	
	➤ Paid for by patients out-of-pocket, \$	✓	✓	✓	
➤ Future related medical costs, \$		✓	✓		
➤ Future unrelated medical costs, \$		✓	✓		
INFORMAL HEALTHCARE SECTOR					
HEALTH	➤ Patient time costs, Earnings \$	✓		✓	No data available
	➤ Unpaid caregiver time costs			×	
	➤ Transportation costs				
NON-HEALTHCARE SECTOR					
PRODUCTIVITY	➤ Uncompensated household production, patient	×		×	No data available
	➤ Productivity effects in formal market, Earnings \$			✓	
	➤ Years in employment, Years			✓	
CONSUMPTION	➤ Future consumption unrelated to health, \$			✓	
SOCIAL SERVICES					
	➤ None			-	
LEGAL / CRIMINAL JUSTICE					
	➤ Costs of AUD-related crimes – Tangible, \$			✓	
	➤ Costs of AUD-related crimes – QoL, \$			✓	
	➤ Costs related to criminal justice system, \$			✓	
	➤ # of AUD-related crimes, # of crimes			✓	
	➤ Years in incarceration, Years			✓	
	➤ # of AUD-related motor vehicle accident (MVA), # of MVA			✓	
EDUCATION					
	➤ None			-	
HOUSING					
	➤ None			-	
ENVIRONMENT					
	➤ None			-	

A Caregivers for individuals with current or previous alcohol problem reported significant higher caregiver distress, compared to those for individuals with no previous alcohol problem. The caregiver burden was measured by the Neuropsychiatric Inventory Caregiver Distress Scale (NPI-C) and the Family Burden Scale (FBS) in the original study, and we were unable to convert the estimate of the caregiver burden into the health utility weight.

Cost-Effectiveness Impact Inventory

Element	Definition
Purpose	Is the purpose of the algorithm clearly stated at the outset? Is the implementation of the algorithm in a healthcare setting fair and ethical?
Appropriateness	Is there a clear justification that the algorithm is acceptable in the context within which it is being applied?
Limitations	Have the strengths and limitations, in the context of the purpose, been identified? This should cover both the algorithm and any data used.
Implementation	Consideration of access, implementation, and resource issues when implemented in healthcare settings.
Sensitivity and specificity	For classification algorithms, has the model performance and accuracy (specificity and sensitivity) been appropriately evaluated?
Algorithm characteristics	Has the ML mechanism been clearly characterized and described? Is there sufficient transparency for the results to be reproducible?
Data characteristics	Is the selection of data sets justified and are the key characteristics known? This should extend to training sets, test sets and validation sets.
Explainability	Are the outputs of the algorithm clearly understandable by both the healthcare professional and the patient?

ML indicates machine learning.

Machine Learning PALISADE Checklist

Proposed Elements for Consideration: **Data Relevancy**

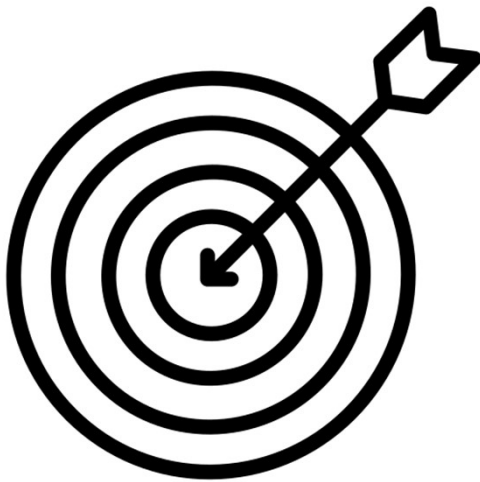
The availability of critical variables and sufficient number of representative patients within the appropriate time period to address a given use case.



- Availability of key study elements
- Representativeness of population
- Timeliness and study time period
- Care setting and treatment pathways
- Timing and frequency of measurements
- Sufficiency of sample size and follow-up time

Proposed Elements for Consideration: **Data Reliability**

The degree to which data represent the clinical concept intended



- Accuracy
 - Validation
 - Verification
- Completeness
- Target concept and operational definition
- Provenance

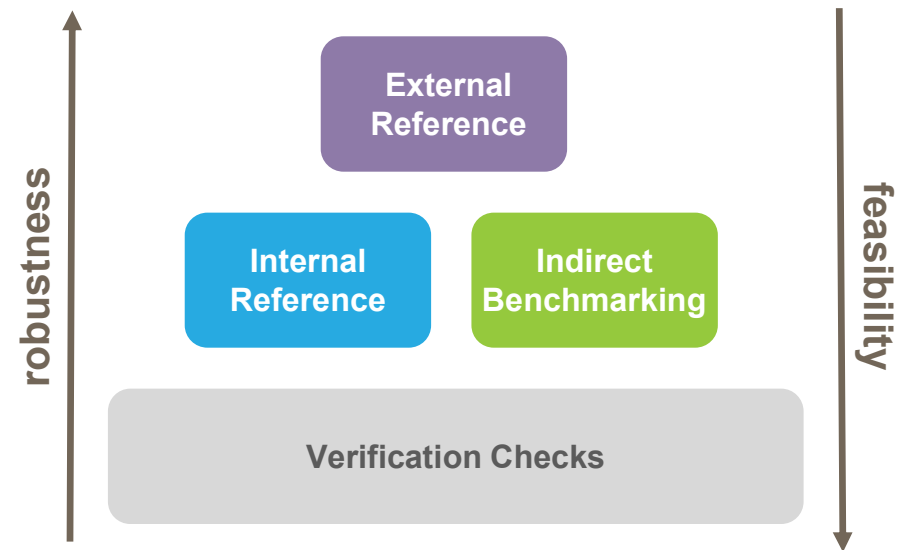
Approaches for validation and verification of EHR data under consideration as good practices

Points of validation:

- Field Level
- Patient Level
- Site Level
- Sub-Cohort of Cohort Level

Types of validation output:

- Sensitivity, Specificity
- Positive and Negative Predictive Values
- Descriptive Statistics
- Agreement Metrics
- Completeness Rates
- Error Rates



Additional recommendations under consideration

- Meaningful EHR use-cases for HTA include natural history, modeling inputs, extrapolation, real world comparative-effectiveness analysis, and more. Access to recent relevant data may enable “living HTA” with more dynamic value assessment over the lifetime of a product.
- Curation of variables for known confounders of the research question (eg, genomic testing results, functional status, vitals, endpoints) using unstructured documents is a key advantage over other types of RWD sources.
- Documentation of protocols and statistical analysis plans may require more detail and amendments (compared to analyses using other RWD data types) as often more decisions need to be made in the analysis of more complex health data.

Seeking Your Feedback

These are elements the ISPOR Task Force on EHR for HTA believes are important to assess EHR data quality for HTA and are considering including in our recommendations.

*Is this appropriate and useful?
What are we missing?*





4

Q&A

To contact the presenters:
taskforce@ispor.org

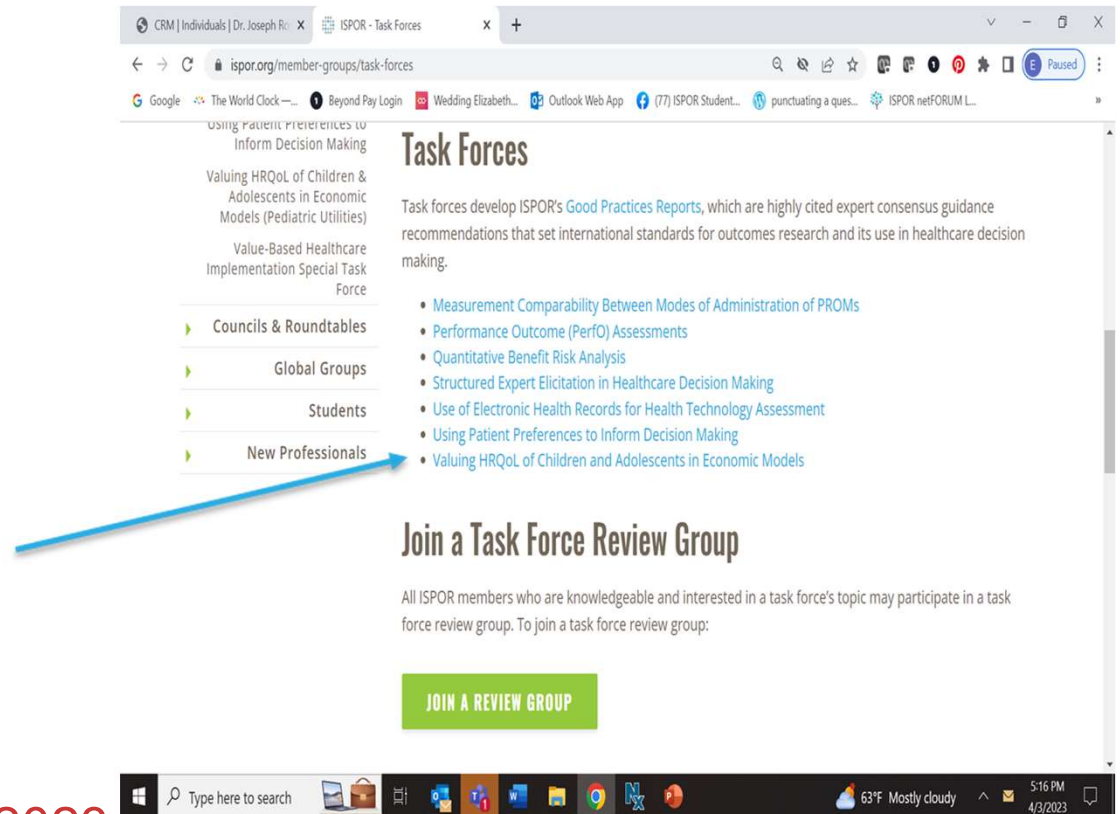
Join Our Task Force Review Group!

1. Visit ISPOR home page www.ispor.org
2. Select “Member Groups”
3. Select “Task Forces”
4. Scroll down to Join a Task Force Review Group
5. Click button to “Join a Review Group”

Use of Electronic Health Records for HTA

****You must be an ISPOR member to join a Task Force Review Group.****

Likely timeline: Fall 2023



Open questions for discussion

- To what extent should we focus on innate data quality versus data suitability or relevance?
- Do we identify the key challenges of EHR data?
- Is our proposed framework structure appropriate? What items should be removed or added?
- Do we identify relevant challenges across countries?
- Can the framework support users of EHR data when generating evidence for HTA evaluations?
- What are useful formats for this Task Force ?