

KU LEUVEN

Using real world data for pragmatic clinical trials: from challenge to opportunity

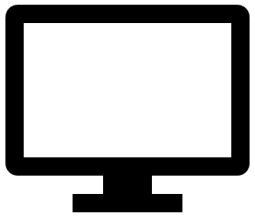
Nicolas Delvaux, MD

Real World Data

"data collected in a non-randomized controlled trial setting"

"data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources"

Real World Data



EHR



Patient generated data



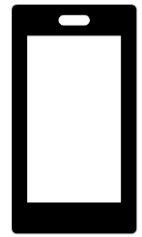
Registries



Billing



Wearables



Mobile devices

Table 1 Common limitations of randomized controlled trials and whether they can be amended by routinely collected health data

Limitations of RCTs [10]	What using RCD for RCTs can offer	Challenges	Potential of RCD to improve RCTs
Generalizability and real-world relevance	No specific data collection processes (follow-up visits, measurements) outside routine care, avoiding artificial situations	Random allocation of interventions may still require some deviation from routine care processes (e.g., obtaining informed consent).	Very high
Costs and resources	No costs to the trial for data collection processes and related activities (study site setup, study staff salary, monitoring and auditing activities, training costs)	Potential costs for obtaining the RCD (if the collecting entity does not provide it for free; e.g., data brokers); additional costs for data management, processing, merging, cleaning, and so forth	Very high
Specific conditions/ subgroup effects	Larger sample sizes that are less influenced by resource constraints and feasibility issues may provide sufficient power for evaluating subgroups.	More opportunities for exploratory analyses with spurious findings	High
Late outcomes	RCD can provide long-term outcome data without actively following patients and often reducing the number of patients lost to follow-up	Patients moving away from RCD infrastructure will be lost and may still require active contact, highly dependent on RCD infrastructure	High
Speed	No cumbersome outcome ascertainment (follow-up contacts, data recording and collection) and no need for setting up the data collection infrastructure, thus results can be obtained faster	Management, processing, merging, and "cleaning" of large datasets may be time-consuming. Reporting of specific adverse events may be delayed.	High to moderate
Conflicts of interest/ sponsorship bias	Collection of RCD is more objective and less easily manipulated to obtain a desired result.	Data may still be analyzed and reported nonobjectively to convey preferred conclusions.	Moderate
Understudied healthcare questions	Providing information on routine care allows researchers to address understudied healthcare questions because more resources are spared or different outcomes are collected.	Not all desired endpoints might be available; funding may not be the sole barrier	Moderate
Regulations	Obtaining approval for intervention imposes several bureaucratic loopholes; RCD are already available and might require different ethical clearance.	RCD still require approval in terms of data protection and confidentiality.	Moderate
Rare or uncommon conditions	Recruiting an appropriate sample size may be hard with rare diseases; larger samples with RCD and easier EHR or registry recruitment can reduce these difficulties.	Only possible if RCD resources are extensive, highly dependent on RCD infrastructure	Moderate

Abbreviations: EHR Electronic health record, *RCD* Routinely collected health data, *RCT* Randomized controlled trial

Opportunities for RWD to amend limitations of RCTs

Generalizability and real-world relevance	No specific data collection processes (follow-up visits, measurements) outside routine care, avoiding artificial situations	Random allocation of interventions may still require some deviation from routine care processes (e.g., obtaining informed consent).	Very high
Costs and resources	No costs to the trial for data collection processes and related activities (study site setup, study staff salary, monitoring and auditing activities, training costs)	Potential costs for obtaining the RCD (if the collecting entity does not provide it for free; e.g., data brokers); additional costs for data management, processing, merging, cleaning, and so forth	Very high
Understudied healthcare questions	Providing information on routine care allows researchers to address understudied healthcare questions because more resources are spared or different outcomes are collected.	Not all desired endpoints might be available; funding may not be the sole barrier	Moderate

Table 2 Barriers in the use of routinely collected health data for randomized controlled trials and options for improvement

General barriers or issues	Pressing questions	Possible solutions, actions and additional comments
Data		
• Availability • Management • Linkage • Accuracy • Validity	• Is the desired outcome variable or RCD source available? • Will it be possible to achieve the same data quality and accuracy with RCD as in traditional trials? • Is the data linkage and management feasible in institutions with limited IT infrastructure?	• A central register of databases available for clinical trial research would be helpful, ideally with details about data quality. • Establish core outcomes and structured outcome assessments in routine care • Create RCD trial guidelines and RCD source validation guidelines to help standardize their use and reduce sources of bias or uncertainty • Increase IT presence (particularly data analysts) to health research teams • The more RCD is sought out and used in research, the greater is its availability and differentiation.
Regulatory and ethics		
• Collecting and obtaining the data • Using and sharing the data	• What type of release must be given by the patients before their data can be collected or shared? • Is it ethical to use RCD without asking for their permission, even if their data are anonymized? • Can this data be considered of value and morally be sold? • How are concerns about privacy and informed consent approached (particularly in the context of population-wide trials or Zelen designs)? • Are data safety standards applied to RCD just as strictly as they are to traditional actively collected data? • Who is responsible for the safety of the data?	• Ethical guidelines specifically regarding the collection and dissemination of RCD should be developed. • Ethics and approval committees should deepen their knowledge of these novel ethical challenges. • Whereas personal data are collected daily from many sources (e.g., phone use), collection, storage, and dissemination of data related to health require more definite ethical oversight and greater transparency to the general public. • After safety issues are defined, researchers and stakeholders must ensure that data are safely handled, with full transparency of access.
Costs		
• Obtaining the data • Managing the data	• Will data collectors (e.g. health insurers) share their data? Freely or at a cost? • Is a constant increase in the generation of routine data really reducing the overall trial costs if the same institution collected the data in the first place? • When is the use of RCD cost-effective?	• The financial worth of health data is not defined or explored; empirical data are necessary to determine the cost of both producing and maintaining health data • Health data are already legally sold to many industries, and regulations/legislation must catch up with this aspect.
Novelty		
• Bureaucratic obstacles • Unawareness • Training to generate, collect, prepare, manage and analyze RCD for trials	• Will approval committees understand the implications of using RCD sources for clinical trials? • What are the challenges that can be expected bureaucratically because most submission templates do not assume the use of RCD and absence of patient contact? • Are data anonymization techniques clear? • What training is required to qualify individuals who generate, collect, prepare, and manage RCD for clinical trial research?	• Develop, in collaboration with approval committees, RCD-specific templates and submission forms, especially in such studies where no patient contact is foreseen and therefore speedy approval is desired. • Educate regarding data anonymization and confidentiality risks • Include the concept of using RCD for RCT in clinical research education and teaching • Create and use reporting guidance specifically for RCD-RCTs

IT Information technology, RCD Routinely collected health data, RCT Randomized controlled trial

Challenges to the use of RWD

Data

- Availability
- Management
- Linkage
- Accuracy
- Validity

- Is the desired outcome variable or RCD source available?
- Will it be possible to achieve the same data quality and accuracy with RCD as in traditional trials?
- Is the data linkage and management feasible in institutions with limited IT infrastructure?

- A central register of databases available for clinical trial research would be helpful, ideally with details about data quality.
- Establish core outcomes and structured outcome assessments in routine care
- Create RCD trial guidelines and RCD source validation guidelines to help standardize their use and reduce sources of bias or uncertainty
- Increase IT presence (particularly data analysts) to health research teams
- The more RCD is sought out and used in research, the greater is its availability and differentiation.

Challenges to the use of RWD

Regulatory and ethics

- Collecting and obtaining the data
- Using and sharing the data

- What type of release must be given by the patients before their data can be collected or shared?
- Is it ethical to use RCD without asking for their permission, even if their data are anonymized?
- Can this data be considered of value and morally be sold?
- How are concerns about privacy and informed consent approached (particularly in the context of population-wide trials or Zelen designs)?
- Are data safety standards applied to RCD just as strictly as they are to traditional actively collected data?
- Who is responsible for the safety of the data?

- Ethical guidelines specifically regarding the collection and dissemination of RCD should be developed.
- Ethics and approval committees should deepen their knowledge of these novel ethical challenges.
- Whereas personal data are collected daily from many sources (e.g., phone use), collection, storage, and dissemination of data related to health require more definite ethical oversight and greater transparency to the general public.
- After safety issues are defined, researchers and stakeholders must ensure that data are safely handled, with full transparency of access.

KU LEUVEN

Bert Aertgeerts
Nicolas Delvaux
Tine De Burghgraeve

UNIVERSITEIT GENT

An De Sutter
Veerle Piessens

KCE
Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

France Vrijens
Elisabeth Marynen
Frank Hulstaert

sciensano

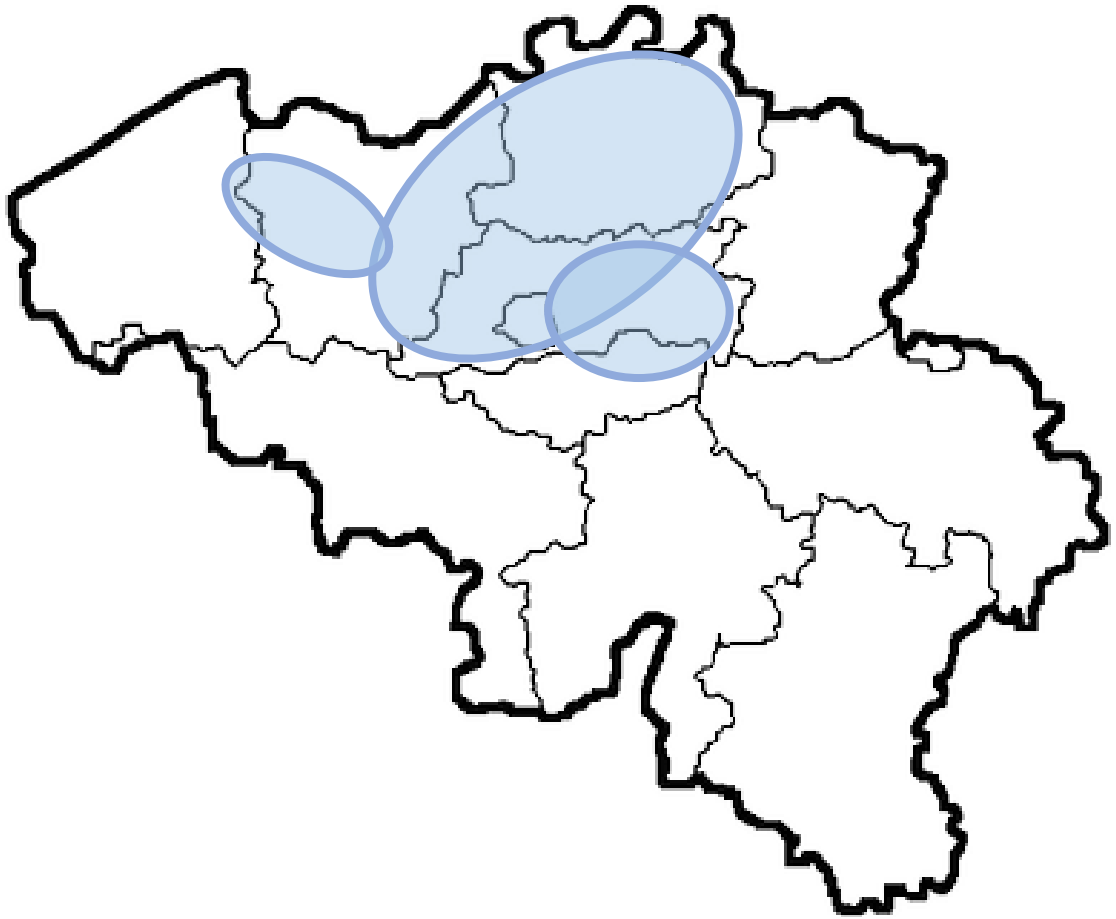
Roel Heylen
Johan Van Bussel

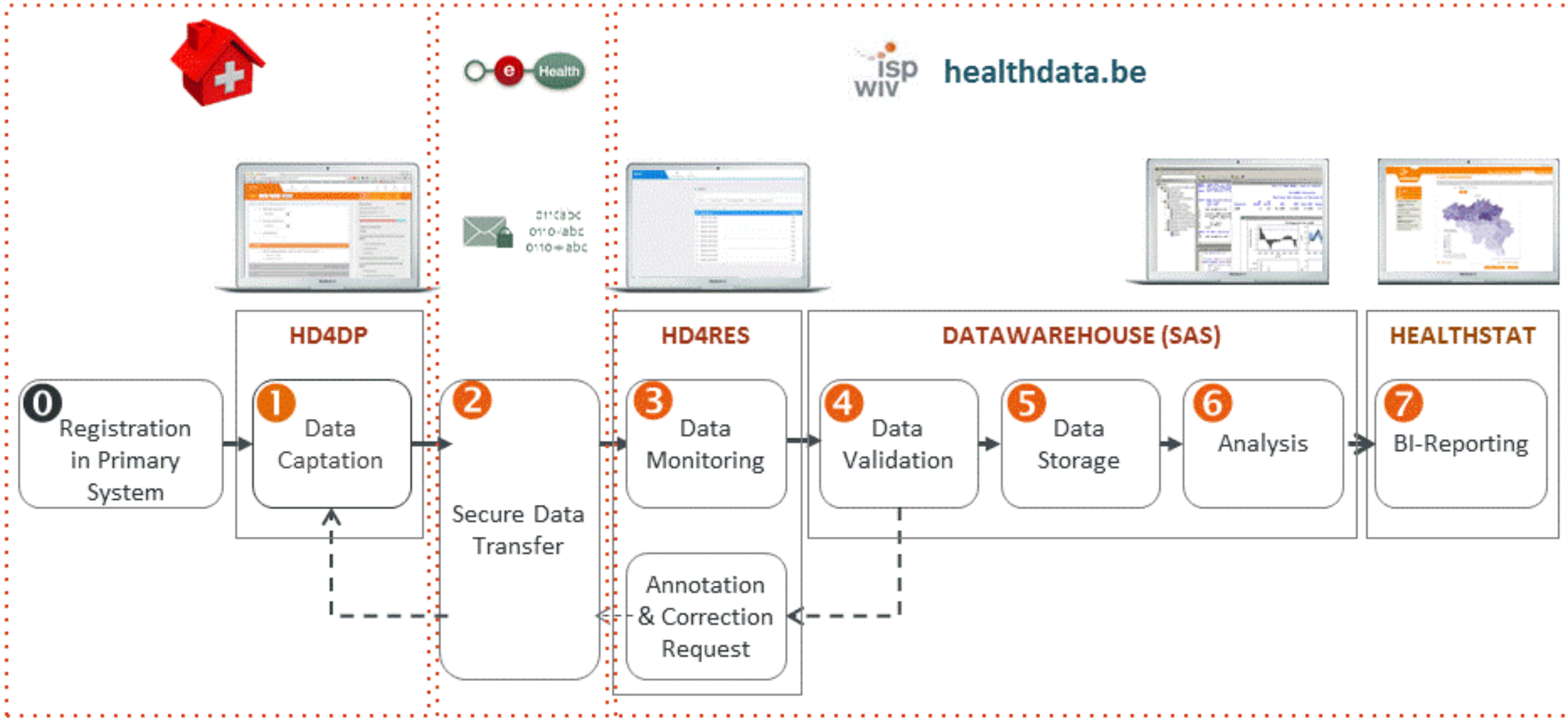


280 GPs
72 PCPs

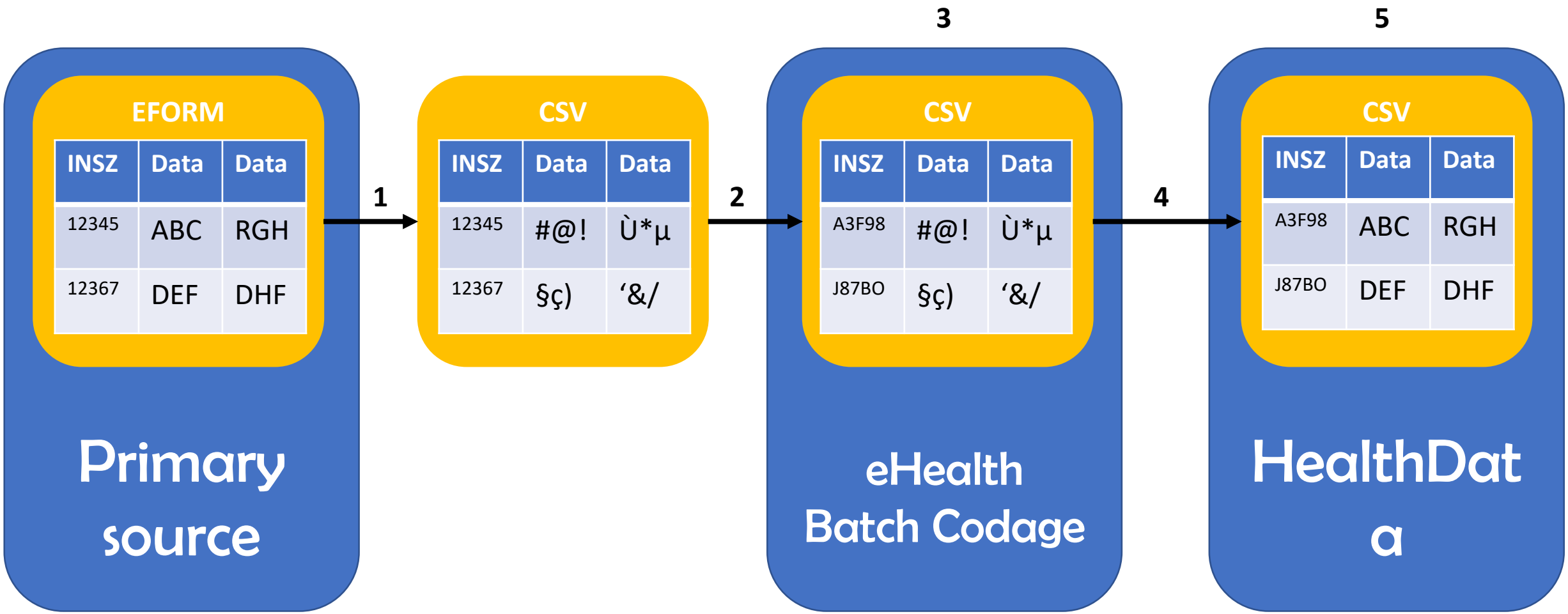
3 private laboratories

10 665 individuals




Health
isp wiv
healthdata.be
HD4DP
HD4RES
DATAWAREHOUSE (SAS)
HEALTHSTAT
0 Registration in Primary System
1 Data Captation
2 Secure Data Transfer
3 Data Monitoring
4 Data Validation
5 Data Storage
6 Analysis
7 BI-Reporting
Annotation & Correction Request

DATA FLOW



1: export & coding; 2: encryption; 3: coding; 4; encryption; 5: decoding

Opportunities ELMO study

From signed research contract to start of study – 3.5 months!

Physician recruitment – limited extra work

Patient recruitment

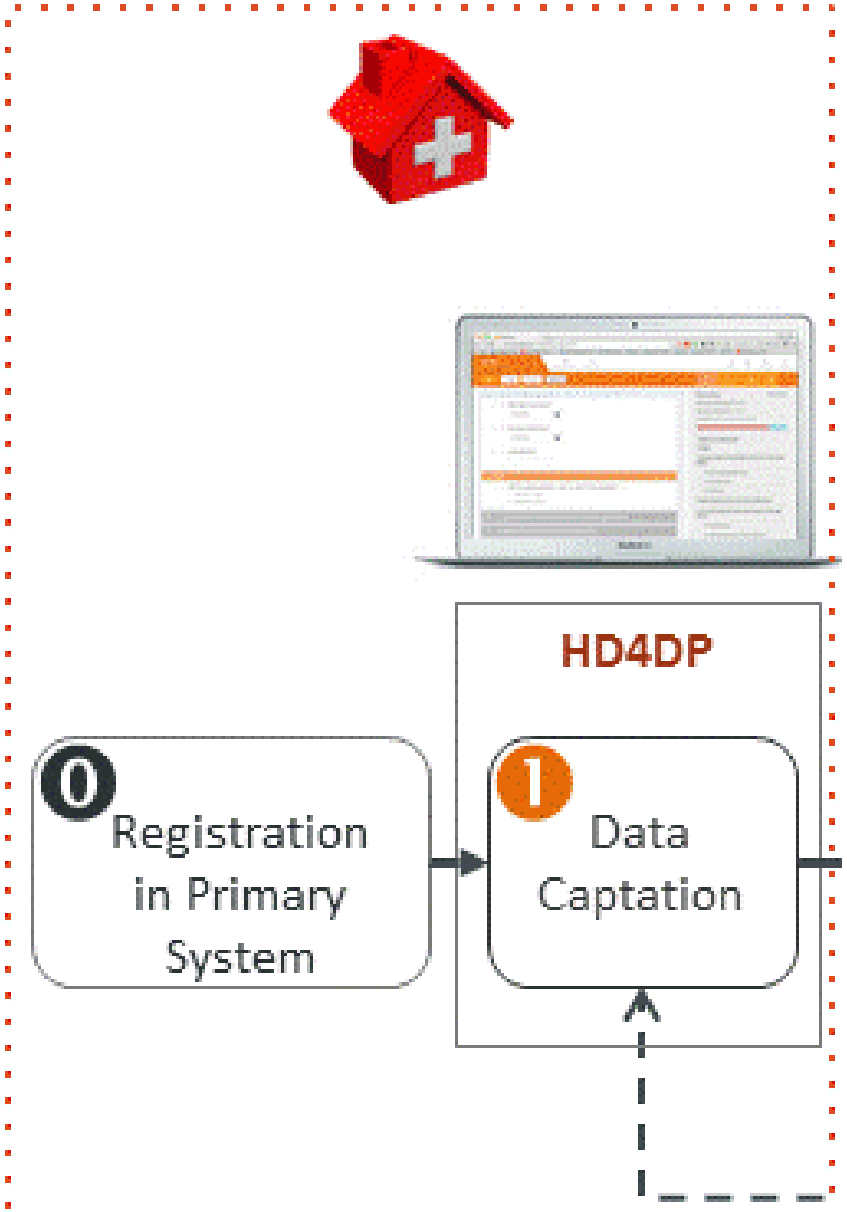
Pseudonymization allows for linkage of patient specific data across multiple systems and registers

Safe and secure data transfers

Use of existing encrypted data channels

Randomization limits issues of data quality

Challenges in Phase 0 and 1 – data capture at primary source



- Technical issues with 1 lab
- Variety of laboratory coding systems

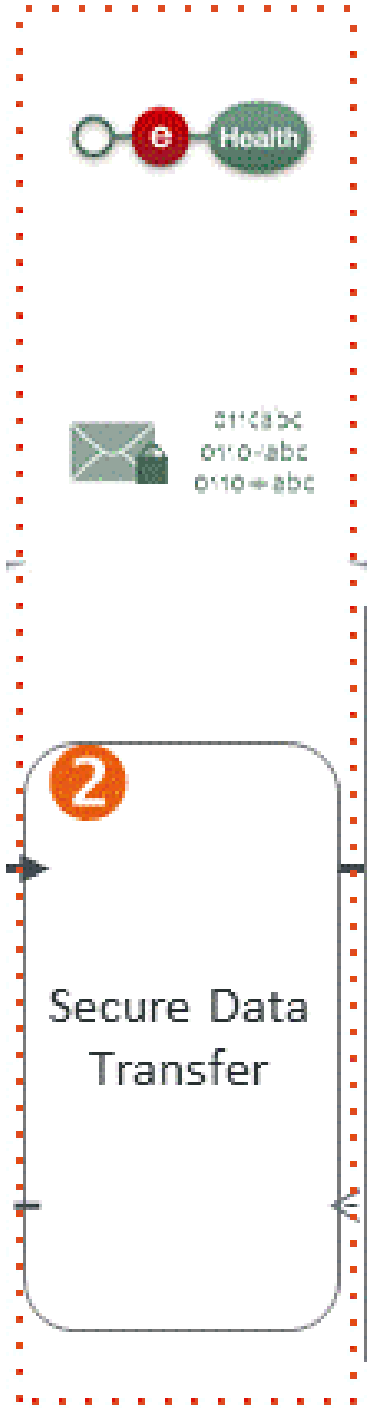
- Data capture quality
- Data capture through own personnel

- Self-reporting technically not feasible
- Interviews through own personnel



- Data capture quality

1. Health care system bias, emanating from:
   - Reimbursement system, pay for performance parameters
   - Role of general practitioner in the health care system; gatekeeping/nongatekeeping
   - Professional clinical guidelines
   - Ease of access by patients to their records
   - Data sharing between health care providers
2. Practice workload
3. Variations between electronic health record (EHR) system functionalities and lay-out
4. Coding systems and thesauruses
5. Knowledge and education regarding the use of EHR systems

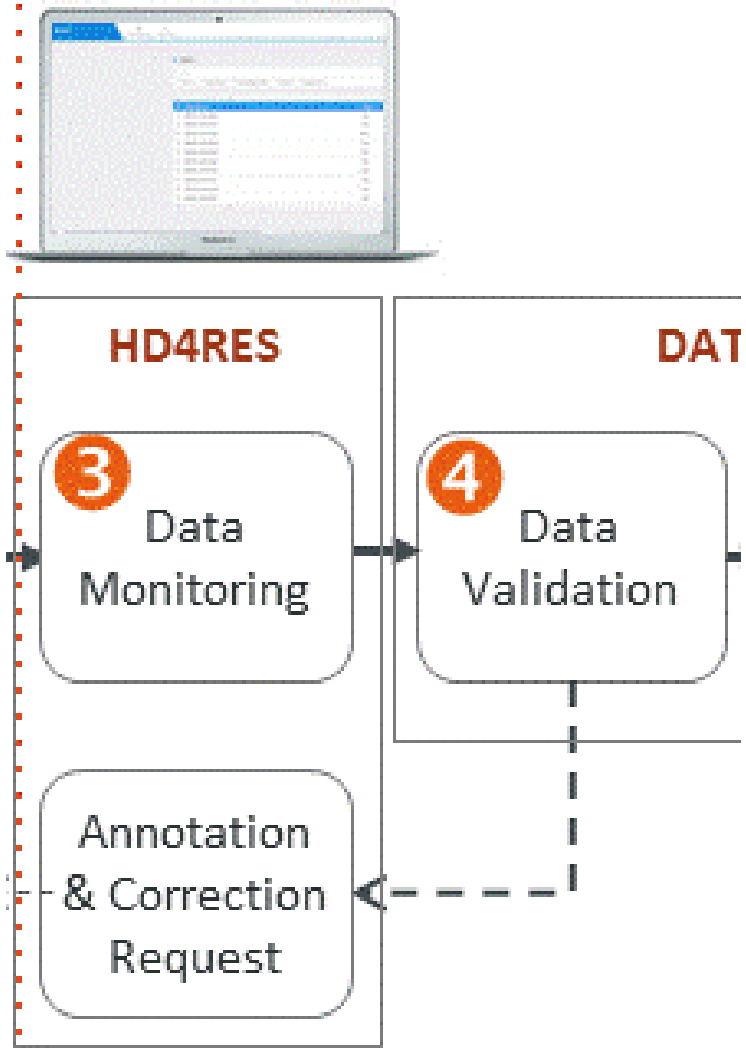


Challenges in Phase 2 – pseudonymization, TTP



- Technical issues related to various EHR systems
- Issues with coding of sensitive data
- Issues with data loss
- Data channels not designed for large data volumes

Challenges in Phase 2 – pseudonymization, TTP



- Technical issues due to encryption key used by primary source
- Platform for queries not applicable to GP EHR

- Query resolvement was slow and unresponsive

Potential solutions for using RWD in Belgium primary care

- Smaller number of EHR systems
- Less variation is methods for documenting routine care
- Facilitation of informed consent at primary source

- Cloud storage of data enables centralized data transfer (like INTEGO)
- Accessory data channels for transferring large datasets

- Oversight over full chain of data transfer by research team (monitoring)
- Registry based trials