Prof. Pascal Coorevits

Certification and Labelling Adviser, EuroRec Representative

Assessing the data quality of hospital EHRs

CONGRESS '19
EAHM
BELGIUM - GHENT

Assessing the data quality in hospital EHRs

Pascal Coorevits

Professor of Medical Informatics and Statistics, Ghent University
Vice-President for Research, the EuroRec Institute
Data Quality Task Force Lead, i~HD

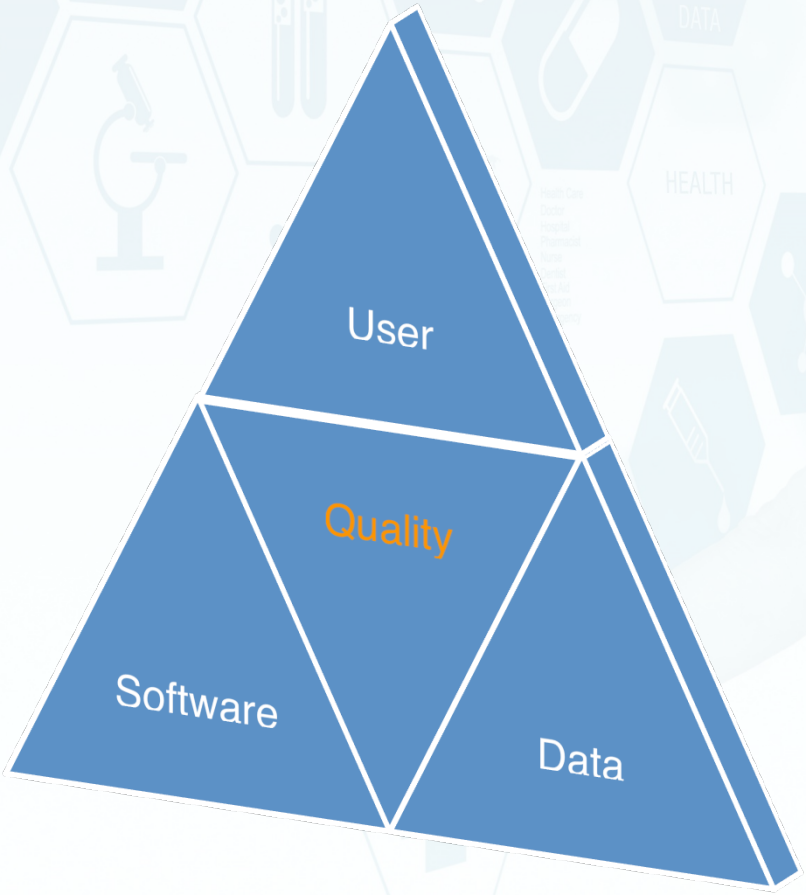
GHENT UNIVERSITY

The European Institute For Innovation Thr~ugh Health Data

CONGRESS '19
EAHM
BELGIUM - GHENT
11 > 14 SEPTEMBER 2019

Quality of health data is vital

- Patients and clinicians want health data to be safe, rapid and evidence based
- Healthcare managers also want to use resources efficiently and need insights for strategic planning
- Public health agencies need reliable data to guide healthcare and prevention programmes
- Healthcare funders need good quality data to reward high-quality and value-based care
- Pharma wants to re-use EHRs to accelerate clinical research
- Regulators and HTA agencies want to be able to trust Real World Evidence in decision making
- Everyone wants to achieve the best patient outcomes and they all know that good data is a critical success factor


User
Quality
Software
Data

Electronic Health Records

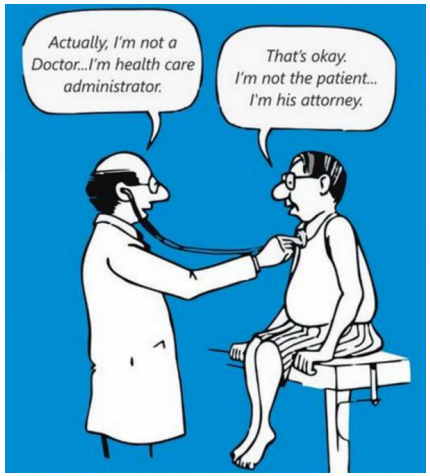
- Electronic Health Records offer enormous potential to improve the safety, quality and efficiency of healthcare
- Empirical evidence for the beneficial impacts of most eHealth technologies on the quality and safety of healthcare are often lacking, or at best, only modest
- Some of the reported issues: lacking of key EHR features, not using EHRs to their full potential, poor interoperability, low usability, low EHR (data) quality, etc…

Quality of EHRs and EHR data

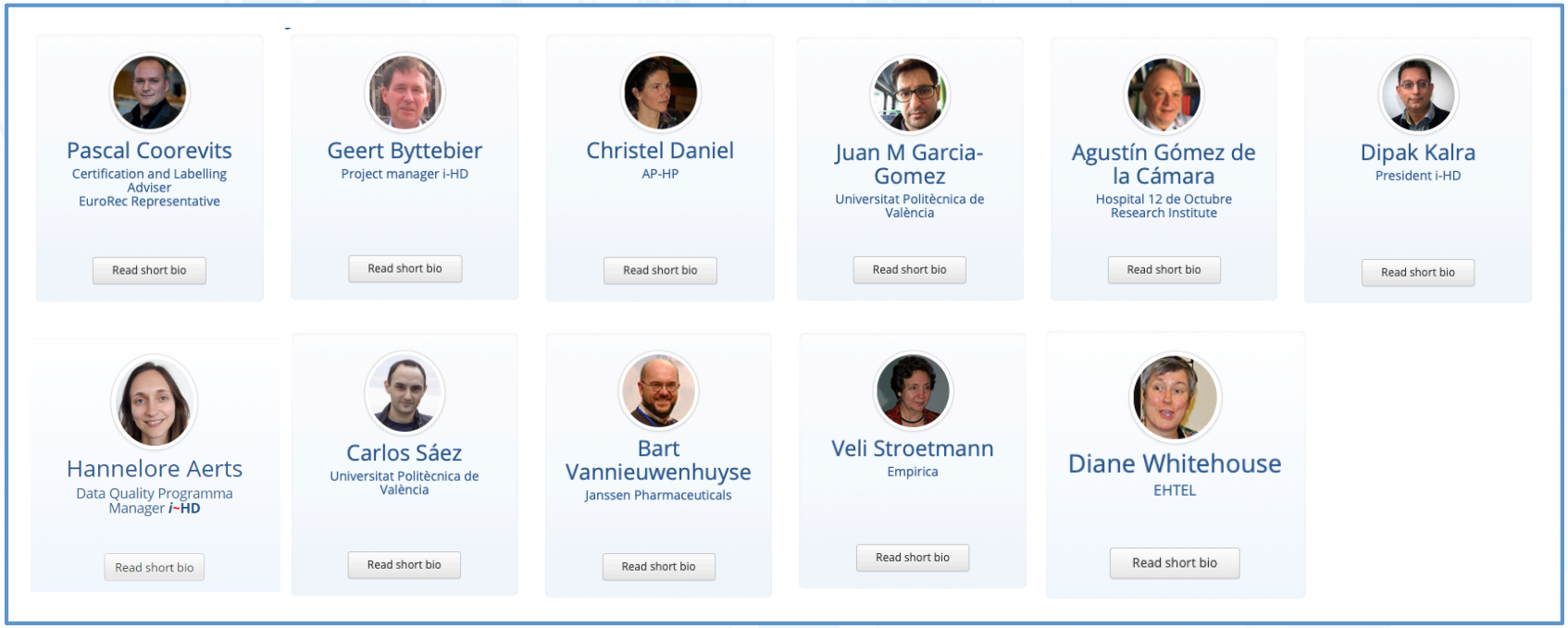
- To use EHRs efficiently for daily routine care, for clinical research, for big data analytics, … a number of functionalities are needed (e.g. security, confidentiality, trustworthiness, …)
- Mechanisms are required to ensure e.g. data correctness, completeness, consistency, …

Quality assurance is essential

Quality labelling & certification are needed



i~HD Data Quality Taskforce



i~HD Data Quality Taskforce aims

- Develop data quality assessment methods, tools and improvement strategies to maximise quality of health data
- Promote the importance of data quality
- Guidance in assessing and improving data quality
- Scale up a multi-stakeholder understanding and commitment to increase data quality

Work on data quality dimensions and assessment methods

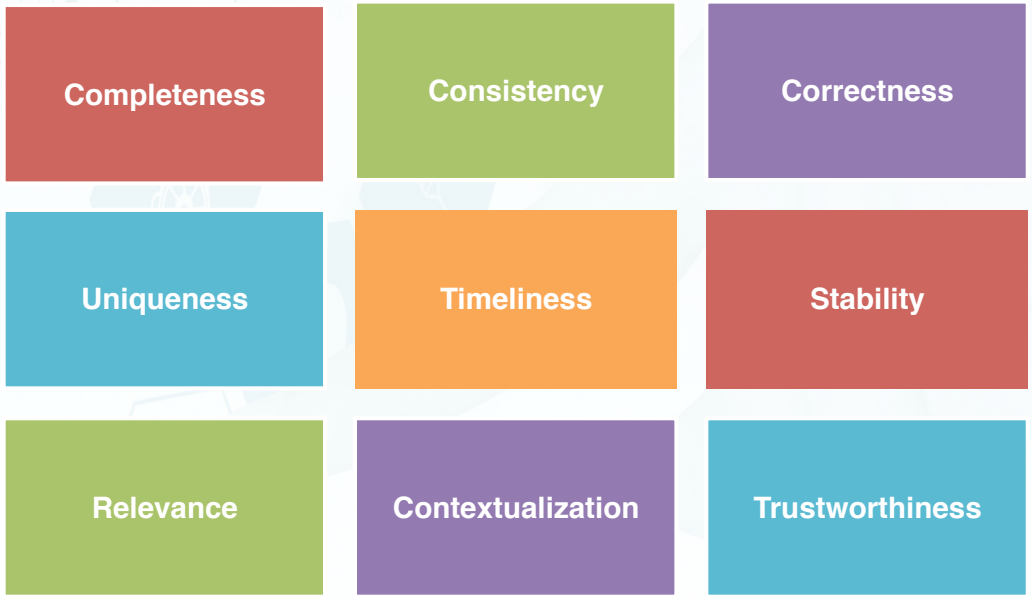
i~HD

Work on data quality dimensions and assessment methods

- ❑ Several data quality indicators used
- ❑ Several definitions for the same concept
- ❑ No universal approach

→ mapping exercise

→9 quality dimensions



Completeness – "data values are present"

Data Group	Data Item	Avg. usage	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 9	Site 9
Demographics	Gender	100%	100,00%	100,00%	100,00%	100,00%	100,00%	100%	100%	100,0%	100,00%
Demographics	Case Status	96%	99,87%	100,00%	60,00%	100,00%	100,00%	100,00%	100%	100,0%	100,00%
Demographics	Date of Birth	89%	100,00%	100,00%	99,00%	NA	100,00%	100%	100%	100,0%	100,00%
Demographics	Admission date	84%	100,00%	100,00%	100,00%	NA	100,00%	99,53%	58%	100,0%	100,00%
Diagnosis	Diagnosis Text	81%	50,46%	84,02%	100,00%	100,00%	98,05%	100,00%	14%	100,0%	80,98%
Diagnosis	Diagnosis Code	81%	50,46%	84,02%	100,00%	100,00%	98,05%	100,00%	14%	100,0%	80,98%
Demographics	Discharge date	75%	100,00%	100,00%	100,00%	NA	100,00%	100,00%	58%	100,0%	14,18%
Diagnosis	Diagnosis Date	70%	50,46%	84,02%	100,00%	100,00%	100,00%	NA	13%	100,0%	80,98%
Medication	Dosage	25%	20,36%	0,00%	NA	NA	94,43%	95%	NA	NA	12,21%
Findings	Weight	25%	29,56%	18,24%	NA	NA	89,17%	27,20%	36%	7,5%	13,82%
Laboratory Findings	Platelets Blood	48%	52,78%	33,14%	63,73%	NA	100,00%	100%	45%	NA	33,88%
Laboratory Findings	SGPT (ALT) in serum	47%	33,61%	22,29%	100,00%	NA	100,00%	100%	47%	NA	21,86%
Laboratory Findings	Total Protein in serum	46%	52,37%	14,96%	86,53%	NA	100,00%	100%	47%	NA	16,34%
Laboratory Findings	Total Bilirubin in serum	46%	33,03%	16,99%	100,00%	NA	100,00%	100%	47%	NA	19,58%

Variables such as Weight are quite frequently not present

Doods *et al. Trials* 2014, **15**:18
http://www.trialsjournal.com/content/15/1/18

TRIALS

RESEARCH Open Access

A European inventory of common electronic health record data elements for clinical trial feasibility

Justin Doods[1], Florence Botteri[2], Martin Dugas[1], Fleur Fritz[1*] and on behalf of EHR4CR WP7

i~HD

- Data quality issues found in a survival analysis of pancreatic cancer patients (Columbia University Medical Center, New York)

➢ Information inconsistency between different EHR data sources:

- In a few cases, pancreatitis was **diagnosed as being chronic in the pathology reports** but it was **reported as being only acute in the clinical notes**

➢ Information inconsistency within the same data sources :

- Some patients received **simultaneously two different ICD-9-CM codes** for their diagnoses of diabetes, both **250.01 and 250.02 for type-1 and type-2 respectively**

Consistency – "Data satisfies constraints"

Summit on Translat Bioinforma. 2010; 2010: 1–5. PMCID: PMC3041534
Published online 2010 Mar 1.

Secondary Use of EHR: Data Quality Issues and Informatics Opportunities

Taxiarchis Botsis,[a,b] Gunnar Hartvigsen,[a,c] Fei Chen,[b] and Chunhua Weng[b]

Osteoporos Int
DOI 10.1007/s00198-016-3635-2

ORIGINAL ARTICLE

Clinical height measurements are unreliable: a call for improvement

A. L. Mikula[1] · S. J. Hetzel[2] · N. Binkley[3] · P. A. Anderson[4]

"Fifty percent of staff reported they on occassion enter patient reported height into the EHR rather than performing a measurement"

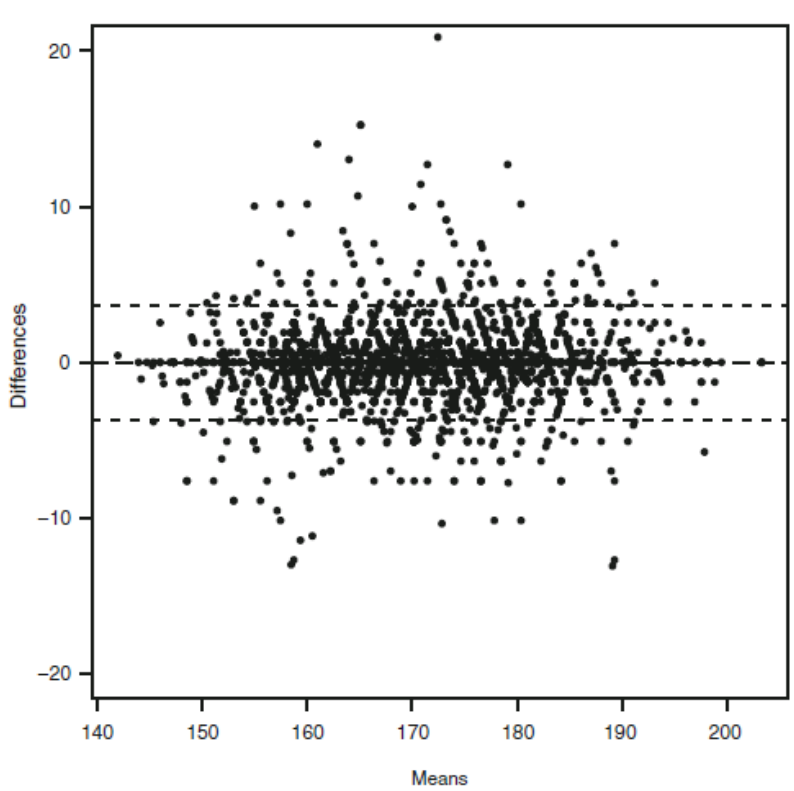


Fig. 4 Bland-Altman plot for height measurements of subjects. Each data point represents a single patient. *X axis* represent mean patient height in centimeters. *Y axis* represents difference between the first and the last height measurement for the individual patients in centimeters. The *dotted lines* represent 95 % CI

Slide courtesy of Carlos Saez and Juan M. García Gómez, Universitat Politècnica de València

i~HD Hospital Network of Excellence Data Quality Workshop

Towards better data quality in hospitals

Tuesday 23rd May 2017 - Wednesday 24th May 2017

i~HD

PARIS, 5 - 6 MARCH 2018

A two-day workshop run by the **i~HD**
Hospital Network of Excellence and Data Quality Task Force,
in collaboration with

Gaining the Benefits of Improved Health Data Quality

ICHOM International Consortium for Health Outcomes Measurement

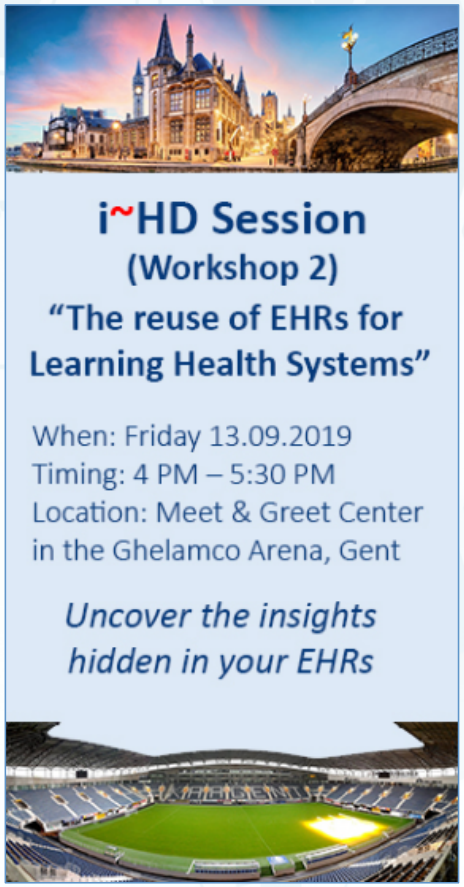
EPF European Patients Forum

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS
WIND - DSI AP-HP


i~HD@
LIFE SCIENCE LIVE
MAY 15 - 16, 2019
TURKU, FINLAND


i~HD @ VITALIS
21 - 23 May 2019
Gothenburg


i~HD
@
CONGRESS '19
EAHM
BELGIUM - GHENT
11-14 September, 2019
Ghent, Belgium


i~HD Session
(Workshop 2)
"The reuse of EHRs for Learning Health Systems"
When: Friday 13.09.2019
Timing: 4 PM – 5:30 PM
Location: Meet & Greet Center in the Ghelamco Arena, Gent
Uncover the insights hidden in your EHRs


Joining the Dots Conference - 27 & 28 November - Brussels
Trillium II
Digital Health Europe
i~HD
EHREDC
C3CLOUD
EHDEN
Exchanging use cases, insights, methods, successes, challenges
Quality and interoperability of high-value data sets
Promoting and sustaining federated learning health ecosystems
Patient empowerment in access and use of health data
Advancing solutions for multi morbidity
Calls to action for health ministries
Joint meeting with the eHealth Network on 28th November
Save the date - registration opens soon

i~HD

Assessing the Quality of Congestive Heart Failure outcomes data at the Hospital del Mar Barcelona

Pascal Coorevits
Geert Byttebier
Dipak Kalra
Geert Thienpont
Carlos Sáez
Juan M. García Gómez
Marta Durá-Hernández
Juan-Manuel Ramírez Anguita
Miguel-Angel Mayer

The European Institute For Innovation Thr~ugh Health Data

GRIB RESEARCH PROGRAMME ON BIOMEDICAL INFORMATICS

upf. Universitat Pompeu Fabra Barcelona

IMIM Parc de Salut MAR Barcelona
Institut Hospital del Mar d'Investigacions Mèdiques

UNIVERSITEIT GENT

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

BDSLab BIOMEDICAL DATA SCIENCE LAB

i~HD

Pilot DQA

- Scoping of the DQA
- Congestive Heart Failure
- 146.602 patient visits with diagnosis of CHF
- ICHOM “Heart Failure” outcome variables basis for selection of variables for DQA ➔ 22 variables (out of 72 ICHOM variables) were selected
- Data Quality Dimensions: Completeness, Correctness, Consistency, Uniqueness & Stability

Congestive Heart Failure (CHF) – ICD9 codes

428 Heart failure
Code, if applicable, heart failure due to hypertension first (402.0-402.9, with fifth-digit 1 or 404.0-404.9 with fifth-digit 1 or 3)
Excludes:
rheumatic (398.91)
that complicating:
abortion (634-638 with .7, 639.8)
ectopic or molar pregnancy (639.8)
labor or delivery (668.1, 669.4)

428.0 Congestive heart failure, unspecified
Congestive heart disease
Right heart failure (secondary to left heart failure)
Excludes:
fluid overload NOS (276.6)

428.1 Left heart failure
Acute edema of lung with heart disease NOS or heart failure
Acute pulmonary edema with heart disease NOS or heart failure
Cardiac asthma
Left ventricular failure

428.2 Systolic heart failure
Excludes:
combined systolic and diastolic heart failure (428.40-428.43)

428.3 Diastolic heart failure
Excludes:
combined systolic and diastolic heart failure (428.40-428.43)

428.4 Combined systolic and diastolic heart failure

428.9 Heart failure, unspecified
Cardiac failure NOS
Heart failure NOS
Myocardial failure NOS
Weak heart

?

402 Hypertensive heart disease
402.0 Malignant
402.00 Without heart failure
402.01 With heart failure

R package: automated pdf & html DQ reporting

Data and configuration files

main_config.csv

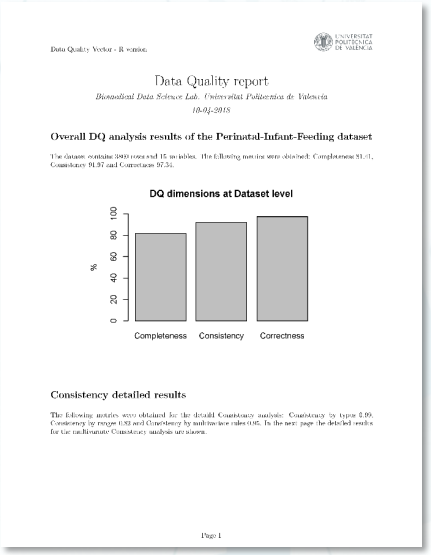
DQ analysis parameters, including paths to:

- dataset.csv

  The dataset to be analyzed

- dictionary_types.csv
- dictionary_ranges.csv
- dictionary_rules.R

  Data dictionary files with semantics about data types, ranges and multivariate rules


Data Quality Vector - R version

Data Quality report

Biomedical Data Science Lab, Universitat Politècnica de Valencia

10-04-2018

Overall DQ analysis results of the Perinatal-Infant-Feeding dataset

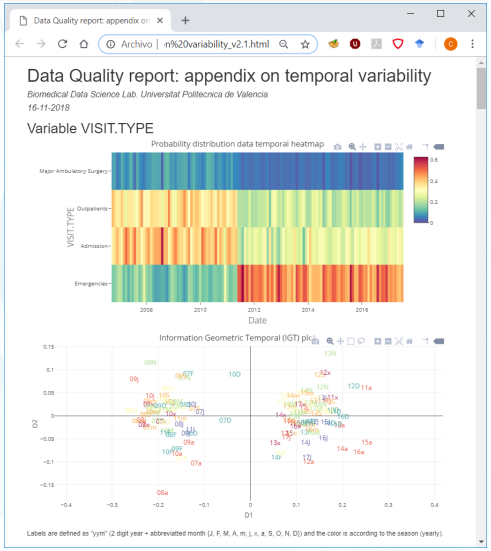
The dataset contains 3000 rows and 15 variables. The following metrics were obtained: Completeness 81.41, Consistency 94.97 and Correctness 97.54.

DQ dimensions at Dataset level

Consistency detailed results

The following metrics were obtained for the detailed Consistency analysis: Consistency by types 0.99, Consistency by ranges 0.93 and Consistency by multivariate rules 0.95. In the next page the detailed results for the multivariate Consistency analysis are shown.

Page 1



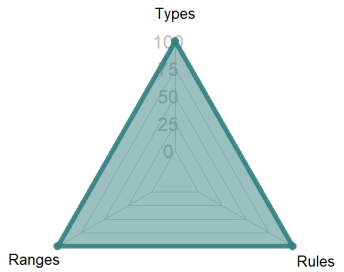
Types → dictionary_types.csv

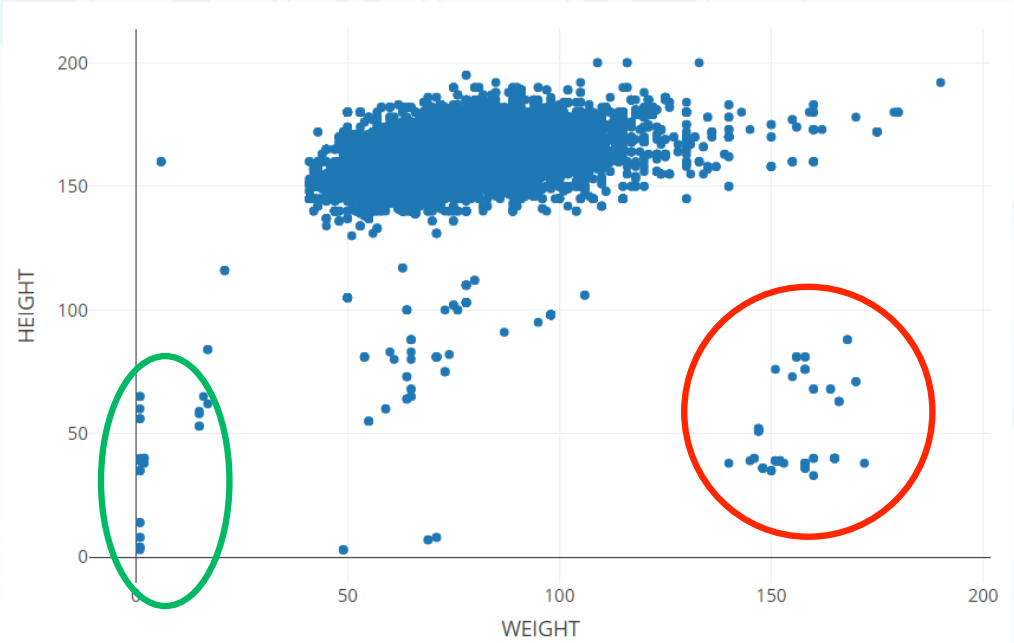
Ranges → dictionary_ranges.csv

Multivariate rule	Comment	Error.sum
ARIDATE <= DISDATE	Arrival date <= Discharge date	0
DATEofBIRTH <= DEATHDATE	Birth date <= Death date	0
DISDATE <= DEATHDATE	Discharge date <= Death date	0
DATEofBIRTH <= ARIDATE	Birth date <= Arrival date	0
(WEIGHT/(HEIGHT/100)^2) <= 70	BMI under 70	180
(WEIGHT/(HEIGHT/100)^2) >= 10	BMI over 10	15

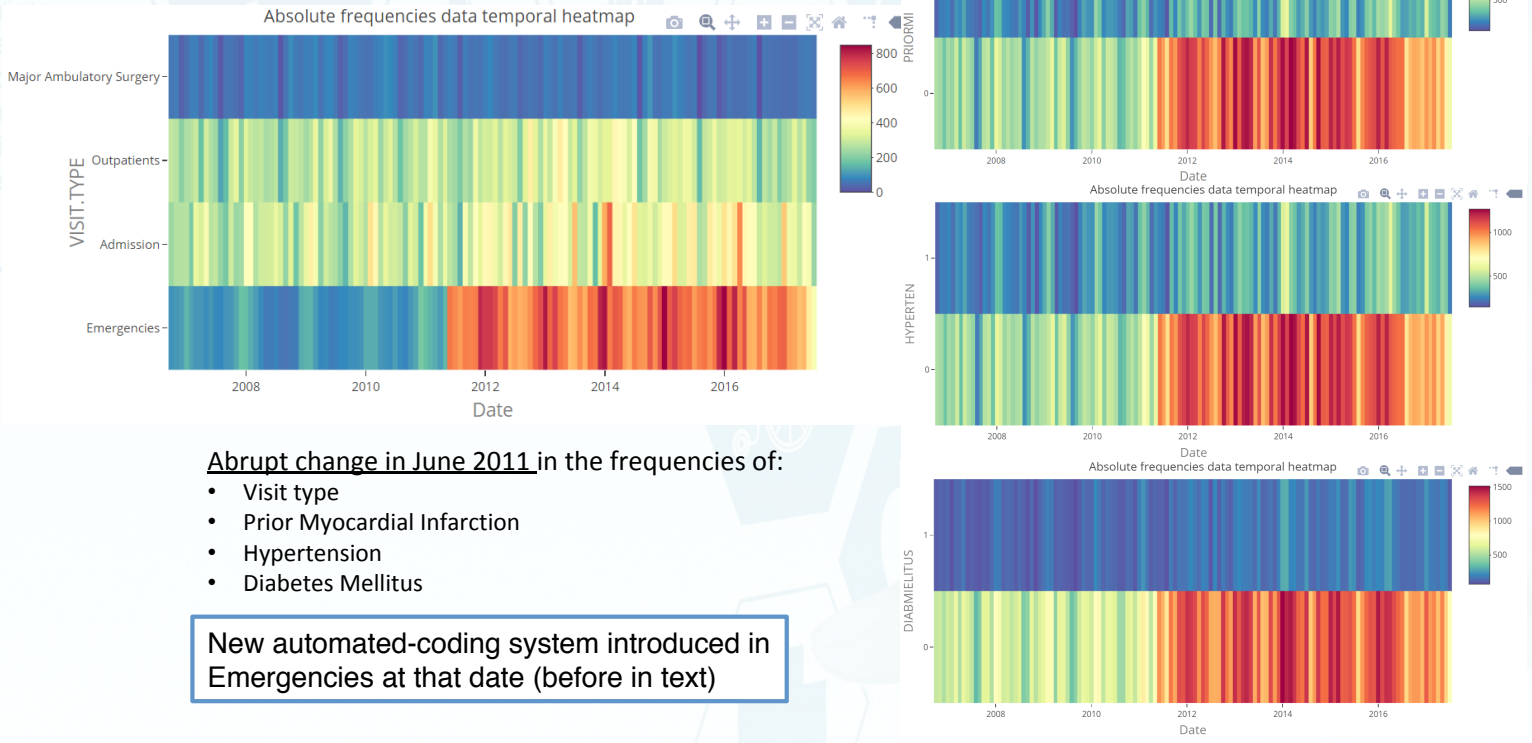
By Types: 99,99 %
By Ranges: 100 %
By Multivariate rules: 99,98 %

Consistency results by types, ranges and rules





Slide courtesy of Carlos Saez and Juan M. García Gómez, Universitat Politècnica de València



Abrupt change in June 2011 in the frequencies of:

- Visit type
- Prior Myocardial Infarction
- Hypertension
- Diabetes Mellitus

New automated-coding system introduced in Emergencies at that date (before in text)

The European Institute For Innovation Thr~ugh Health Data

Report - Pilot Project - Hospital Parc de Salut Mar Barcelona, Catalonia, Spain

DQS4H

Data Quality Service for Hospitals

i~HD

Table of content

- It is essential to have **considerable knowledge of the EHR** (types of data available, how the data were collected or who collected it)
- The **assessment** of the data is the very first step to improve the quality of your data
- Once you know about the quality of your data, it is important to **monitor it regularly**
- **Multidisciplinary** approach is highly recommended
- Thinking of using EHR for different purposes such as research, EHR data models would need to be **expanded and redesigned** and data quality assessment can assist in doing these tasks
- It is of value that an **external assessment** of the data quality is performed by an independent organization
- High-quality data **enhance the validity and reliability** of study findings
- It is critical to ensure that **the metrics** are feasible, valid, and meaningful for a specific EHR and purpose and its quality improvement

Do you want to make better use of your electronic health record?

Do you want to improve the quality of your health data?

DQS4H

Data Quality Service for Hospitals

www.i-hd.eu

i~HD

Preparation and planning

- Webinar on data quality
- Needs and objectives
- Define the scope and domain(s)
- Scope the data sets
- Partnership and confidentiality agreements, GDPR
- Assign roles, timelines
- Contracts

Dataset generation

- Onsite visit
- Prioritise the DQ dimensions
- Select the EHR variables
- Prioritise and localise the i~HD DQ rule library
- Pre-assessment of the data set extract
- Fine tuning the data set
- Final data set for assessment

Assessment

- Validation of the data set
- Tools based analysis of the variables for agreed dimension
- Graphical outputs + descriptive interpretation by i~HD medical and statistical experts
- Preliminary findings discussed with the hospital, to exchange insights on causative factors

Outcome

- Final written report
- Presentation to the team
- Discussion of recommendations
- Improvement strategy planning
- Workshops, online tutorials
- Issue successful hospitals with a **Data Quality Seal**

i~HD

Wat makes the i~HD DQS4H unique?

Pragmatic: minimally invasive to hospital operations

Evidence based: well researched, published, assessment methodology

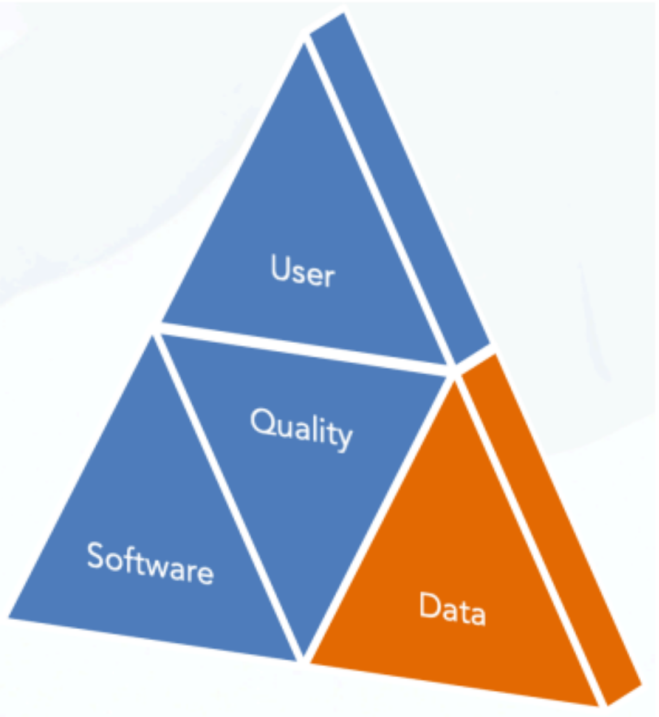
Flexible: can be tailored to your data quality drivers

Focussed: we can help you choose the most suitable dimensions and EHR data variables

Staged: clear sequence of steps with regular interactions and feedback loops

Holistic: we consider quality in the context of your user workflows and your EHR system

Extendable: data sets can be added incrementally, to chart out a data quality improvement journey



Frank Staelens
OLV Hospital Aalst

i~HD@iHD_HealthData . Oct 9

i~HD member,@fstaelens, OLV Hospital Aalst: *My vision is that medical data is an important vehicle to measure and improve quality of care in our hospital. i~HD supports us in analysing and visualising our health data and its quality. This helps in providing high quality care."*

- i~HD@iHD_HealthData . Oct 15
- @Joanxcomella, Director of @VHIR_: *Given #iHD mission in health data quality, it was quite obvious for us to select i~HD as a preferred partner for our hospital data quality strategy. It will make us an even more attractive partner for European research activities."* #healthdata

Joan Comella
Director of Vall d'Hebron
Institut de Recerca

Miguel Angel Mayer
Hospital del Mar

Dipak Kalra@DipakKalra . Oct 17

Dr Miguel Angel Mayer, from Hospital del MAr and i~HD Member, explains to the SCOPe audience how integrating clinical data, and assessing its quality, has enabled them to accelerate their bio-informatics research. #Datasaveslives

i~HD

Thank you!

pascal.coorevits@ugent.be

www.linkedin.com/pascalcoorevits