Data quality: Summary of the i~HD data quality assessment methods and improvement strategies

Pascal Coorevits¹, Carlos Sáez², Juan M García-Gómez², Veli Stroetmann³, Diane Whitehouse⁴, Geert Byttebier⁵, Dipak Kalra⁶

¹Ghent University & EuroRec, ²Universitat Politècnica de València, ³empirica, ⁴EHTEL, ⁵Janssen Pharmaceuticals, ⁶i~HD
The importance of good data quality

Patients
- Assurance
- Enhanced treatment
- Easier access to new drugs
- Better workflows
- ...

Hospitals
- Increased quality of care
- Optimised clinical trials
- Cost savings
- ...

Financial benefits
- Shorter trial process
- Improvements in several sectors
- ...

Other industrial stakeholders
- ...

Other industrial stakeholders

- Financial benefits
- Shorter trial process
- Improvements in several sectors
- ...

Other industrial stakeholders

Quality triangle
i~HD Data Quality Taskforce

Pascal Coorevits
Certification and Labelling Adviser
EuroRec Representative

Juan M Garcia-Gomez
Universitat Politècnica de València

Carlos Sáez
Universitat Politècnica de València

Veli Stroetmann
Empirica

Diane Whitehouse
EHTEL

Geert Byttebier
Janssen Pharmaceuticals

Louis Schilders
Janssen Pharmaceuticals

Bart Vannieuwenhuyse
Janssen Pharmaceuticals

Dipak Kalra
President i-HD
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

Focus on three areas:
- Healthcare
- Clinical trials
- Big data
Workshops and conferences

i~HD Hospital Network of Excellence Data Quality Workshop
Towards better data quality in hospitals
Tuesday 23rd May 2017 - Wednesday 24th May 2017
Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research

Nicole Gray Weiskopf, Chunhua Weng

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

Michael G. Kahn
University of Colorado Anschutz Medical Campus, michael.kahn@ucdenver.edu

Tiffany J. Callahan
University of Colorado Anschutz Medical Campus, tiffany.callahan@ucdenver.edu

Juliana Barnard
University of Colorado Anschutz Medical Campus, juliana.barnard@ucdenver.edu

Alan E. Bauck
Kaiser Permanente Northwest, alan.bauck@kpchr.org
Data quality dimensions

- Completeness
- Consistency
- Correctness
- Uniqueness
- Timeliness
- Stability
- Relevance
- Contextualization
- Trustworthiness
## Data quality dimensions

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>Data values are <strong>present</strong></td>
</tr>
<tr>
<td>Consistency</td>
<td>Data satisfy <strong>constraints</strong> (format, allowable ranges and values, domain rules, relations)</td>
</tr>
<tr>
<td>Correctness</td>
<td>Values are <strong>true and unbiased</strong> with respect to their real-world state</td>
</tr>
<tr>
<td>Uniqueness</td>
<td>Records representing a single patient are <strong>not replicated</strong></td>
</tr>
<tr>
<td>Timeliness</td>
<td>Data is <strong>up-to-date</strong> to their real world state for the task at hand</td>
</tr>
<tr>
<td>Stability</td>
<td>Data inherent concepts and statistics are <strong>comparable among sources</strong> (hospitals, professionals, etc) and over time</td>
</tr>
<tr>
<td>Relevance</td>
<td>Data are <strong>useful</strong> for their task</td>
</tr>
<tr>
<td>Contextualization</td>
<td>Data are annotated with the acquisition context, their <strong>meaning and semantics</strong></td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>Data can be trusted based on the <strong>reputation</strong> of the stakeholders involved in their acquisition</td>
</tr>
</tbody>
</table>
Data stability assessment for realising the value of Health Data for Research & Process improvement
Biomedical Data Science Lab

- Research lines:

- Contact:
  Carlos Sáez
carsaesiiupv.es

Biomedical Data Sources
- Electronic Health Records,
- Public Health Registries,
- MR Imaging,
- MR Spectroscopy,
- Biomedical signals,
- Laboratory measurements, etc.

Healthcare disciplines
- CNS tumors (e.g. Glioblastoma),
- Breast cancer (e.g. DCIS),
- Chronic diseases (Diabetes),
- In-Vitro Fertilization, Depression,
- Case Management, etc.
DQ experiences: the power of assessing data “stability”

- Methodological stability assessment* facilitated:
  1. Discovering other DQ dimension problems
  2. “Harvesting” value data for research


Selected by the IMIA as best-of papers published in 2016 in the area “Learning from Experience: Secondary Use of Patient Data” (Yearbook 2017)

- 2012 – Initial definition of dimensions
  - DQV framework
- 2013-2017 – DQ advising for Hospital Virgen del Castillo, Spain
  - Internal collaborations & Project by Spanish Ministry of Health + Hospital 12 de Octubre + VeraTech SL
- 2014 – Multi-source stability metric
- 2014 – Temporal stability metric
- 2016-today – i~HD Data Quality TaskForce
- 2017-today – Stability assessment of Hospital La Fe for reuse in predictive modelling
Stability assessment for data reuse in predictive modelling in Hospital Universitari i Politècnic La Fe

- New hospital opened between 2010-2011

Yearly numbers:
- 700k outpatients
- 50k hospitalizations
- 250k emergencies

- Hospital project for predicting 30-day patient readmission*
  - Anticipating high-cost patients
  - Saving 500,000€ monthly
  - 2010-16 data available

*La Fe Hospital project by S. Tortajada, J. Pérez, C. Sáez, A. Conejero, JM García-Gómez, B. Valdivieso
Temporal stability assessment of 2010-16 data H.U. La Fe

Temporal stability probability heatmap (N=108,347)

- RELOCATION OF FACILITIES
  - Reduction of planned admissions
  - >Urgent profiles

- START OF NEW BED ALLOCATION POLICY

- ARRIVAL OF NEW POPULATION FROM HOSP. DR. PESET
  - >Admissions
  - >Home hospitalization beds

Length of stay of more than 2 days was abruptly reduced

HIGHLY PREDICTIVE VARIABLE FOR 30-DAY READMISSION!
What data timespan has the largest VALUE for building a proper 30-day patient readmission statistical model?
Conclusions

- Stability assessment applied in more than 7 DQ projects
  - Temporal stability
  - Multi-source stability
  - Crucial when reusing Big RWD of long periods and multiple data-creators (hospitals, professionals…)

- Stability metrics and exploratory methods can be directly incorporated in DQ assessment methodologies and tools
Data quality improvement strategies
Data quality survey

- What are the best ways to enhance data quality?
- Which methods are the most motivating or incentivising?
- What are the important dimensions underpinning data quality?
Data quality improvement

Incentives strategies for Data Quality Improvement

Motivation
Reputation
Positive reinforcement
Feedback
Leadership
Benchmarking
Patient access / input
User experience
Comparison
Competition
Audits

Benefits
• Improved care quality
• Cost savings
• Increased productivity / efficiency
• Improved research opportunities
• …

Enablers
• EHR / EMR system
• IT Infrastructure
• Software tools & applications
• Support services to improve DQ (e.g. consultancy, education, policies, guidelines, …)
i~HD Hospital NoE Data Quality Workshop

Wednesday 24th May 2017, Brussels

Brainstorming session on
the data quality improvement value case

The value of better data quality -
why should a hospital invest in improving data quality

Group work on arguments and example evidence
to support a data quality improvement value / business case

Session lead: Veli Stroetmann, Head of Unit eHealth Research & Policy, empirica Technology Research, Bonn, Germany

Group leaders and rapporteurs:

- Healthcare: Mark Jackson, Director of Research & Informatics, Liverpool Heart and Chest Hospital, United Kingdom
- Clinical Trials: Christel Daniel, Directrice adjointe du département WIND en charge des donnée, Assistance Publique - Hôpitaux de Paris (AP-HP), Paris, France
- Big Data: Gurparkash Singh, Janssen Research & Development LCC, United States
<table>
<thead>
<tr>
<th>Benefits</th>
<th>Financial implications</th>
<th>Non-financial implications</th>
<th>Evidence / examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare (quality)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Improved quality &amp; safety of care</td>
<td>- Better billing</td>
<td>- Transparency in data capture, reporting and usage</td>
<td>- KPMG study</td>
</tr>
<tr>
<td>- Patient &amp; professional confidence</td>
<td>- Better spending</td>
<td>- Time and effort required (also financial impact)</td>
<td>- Liverpool Heart and Chest Hospital</td>
</tr>
<tr>
<td>- Higher efficiency</td>
<td>- Improvement in cost-efficiency – input vs. output</td>
<td>- Staff engagement processes</td>
<td>- NCBC (National Cardiovascular Benchmarking Collaborative), UK</td>
</tr>
<tr>
<td>- Better research</td>
<td>- Potential for outcome-based reimbursement</td>
<td></td>
<td>- Swedish Neuro Registries</td>
</tr>
<tr>
<td>- Better decisions – organisationally and for patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- For hospitals (financial &amp; non-financial: incomes and cost savings, improved cost-efficiency, new business, visibility, publications)</td>
<td>- Costs for business model definition</td>
<td>- Fostering of innovation in collaboration with patients</td>
<td>Evidence from AP-HP:</td>
</tr>
<tr>
<td>- Benefits for patients: access to new (free) drugs and better follow up</td>
<td>- Data quality tools, infrastructure, training, certification, audits</td>
<td>- Improved quality of care for patients</td>
<td>- Feasibility studies</td>
</tr>
<tr>
<td></td>
<td>- Stakeholder involvement (healthcare professionals and directors)</td>
<td>- Personal motivation for change management</td>
<td>- Patient recruitment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Data collection (standards, interoperability, conformance, fine-grained data)</td>
</tr>
<tr>
<td>Big Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Translation research</td>
<td>- Time-saving to market for drugs</td>
<td>- Data sharing</td>
<td>- CPRD (Clinical Practice Research Datalink)</td>
</tr>
<tr>
<td>- Better / more efficient multicentre studies</td>
<td>- High costs of investment</td>
<td>- Data (quality) standards</td>
<td>- CALIBER (Clinical reseArch using Linked Bespoke studies and Electronic health Records)</td>
</tr>
<tr>
<td>- Reduced costs</td>
<td>- Access to new resources in technology and talent</td>
<td>- Data harmonization</td>
<td></td>
</tr>
<tr>
<td>- Use of free-text information</td>
<td></td>
<td>- Patient rights (concern or positive?)</td>
<td></td>
</tr>
</tbody>
</table>
Future work

▪ Extending work on survey (methodological work, validation of preliminary results, …)
▪ Publications of the results of i~HD Data Quality taskforce in international peer-reviewed journals
▪ Communication of a collection of good practices
▪ Develop first i~HD service on data quality
▪ Organise next workshop(s)
Next workshop

Next i~HD Data Quality workshop to be held on November 30th 2017