LONGITUDINAL DATA ANALYSIS IN HEALTHCARE

HOW TO SET UP RELEVANT AND ACCESSIBLE DATABASES

Prof. Dr. Hans-Ulrich Prokosch

PD Dr. med. Thomas Ganslandt, Dr. Martin Sedlmayr, Jan Christoph
Chair of Medical Informatics, Friedrich-Alexander Universität Erlangen-Nürnberg

02.05.2017

ZD.B Symposium: The Future of Healthcare – Big Data Driven Healthcare
Make EHR Data Accessible

Data Warehousing

Perspectives for Medical Informatics
Reusing the Electronic Medical Record for Clinical Research


Challenge 1:
- Establish comprehensive clinical data warehouses by tapping and integrating clinical patient data from a multitude of currently separated clinical databases.
- Apply data and process mining methodologies as well as innovative tools for strategic reporting in order to gain new information and knowledge from such a data warehouse.
Make EHR Data Accessible: Instrumenting the health care enterprise for discovery research in the genomic era


Data Integration – international efforts

PCORnet

33 Partner Networks

• **13 Clinical Data Research Networks (CDRNs)**
  • based in healthcare systems such as hospitals, integrated delivery systems, and federally qualified health centers

• **20 Patient-Powered Research Networks (PPRNs)**
  • operated and governed by groups of patients and their partners.
First pragmatic clinical trial based on the PCORnet networks

- identify the optimal dose of aspirin for secondary prevention in patients with atherosclerotic cardiovascular disease (ASCVD)
- identify patients who are at high risk for ischemic events (phenotyping)
- 20,000 patients are randomly assigned to receive an aspirin dose of 81 mg/day vs. 325 mg/day
Data Integration – international efforts

OHDSI

Observational Health Data Sciences and Informatics

(OHDSI, pronounced "Odyssey")

- multi-stakeholder, interdisciplinary collaborative
to bring out the value of health data through large-scale analytics.
Characterizing treatment pathways at scale using the OHDSI network


Data Integration – industrial efforts
IBM Watson Health

IBM Watson Health

Our mission is to empower leaders, advocates and influencers in health through support that helps them achieve remarkable outcomes, accelerate discovery, make essential connections and gain confidence on their path to solving the world’s biggest health challenges.

IBM Watson for Genomics
Bringing the promise of precision medicine to more cancer patients, Watson can interpret genetic testing results faster and with greater accuracy than manual efforts. Our partnership with Quest Diagnostics means that all providers can potentially benefit, regardless of access to in-house sequencing.

IBM Watson for Drug Discovery
Help researchers identify novel drug targets and new indications for existing drugs. The platform can help researchers uncover new connections and develop new treatments ahead of the competition.
Learn more about Drug Discovery

IBM Watson Health Patient Engagement
Identify patients with care gaps and automate personalized interventions, keeping patients engaged and helping them manage their own care between visits.
Learn more about Patient Engagement

IBM Watson for Oncology
Spend less time searching literature and more time caring for patients. Watson can provide clinicians with evidence-based treatment options based on expert training by Memorial Sloan Kettering (MSK) physicians.

IBM Watson Care Manager
Use personalized care plans, automated care management workflows, and integrated patient engagement capabilities to help create more informed action plans.
Data Integration – industrial efforts
IBM Watson Health

Watson for Oncology

Spend less time searching literature and the EMR, and more time caring for patients. Watson can provide clinicians with evidence-based treatment options based on expert training by Memorial Sloan Kettering (MSK) physicians.

Consume and share large amounts of dynamically curated data.

• Ability to keep pace with the growing and changing body of relevant guidelines, trials, articles, and patient data.
• Confidently derive key insights from the relevant information and medical advances that are applicable to a patient’s condition.
• Develop a more individualized, patient-centric approach to oncology while helping to increase time for patient-physician interactions.
Data Integration – national initiative
BMBF Medical Informatics Funding Scheme

Core Elements

- **Data Integration Centres**
  - at University Hospital Sites
  - within funded consortia
  - across consortia (Germany wide cooperation)

- **currently seven funded consortia**
  (http://www.gesundheitsforschung-bmbf.de/de/6685.php)
  - have submitted proposal for networking and development phase (28.4.17 ⇒ 1.1.2018)

- **National Steering Committee (NSC), Working Groups**
  - Patient Consent / Data Use / Data Governance / Data Protection / Data Security
  - Use and Access Policies / Committees
  - Interoperability

https://www.bmbf.de/pub/Medical_Informatics_Funding_Scheme.pdf
The MIRACUM example: Components of a Data Integration Center

Board of Directors
Ethics Committee

Use- & Access Committee

Data Transfer Unit

Software Test/Deployment Pipeline
Data Provenance
Anonymisation
Visualisation

Long Term Data Archive

Federation

Source system 1
Source system 2
Common server

Routine-Business Reporting

Data Warehouse

Harmonisation

Research Data Repository

Research Queries / Analysis

Data Provenance

Enrichment

NLP
Phenotyping

Security
Authentification
Audit-Trail

Metadata
MDR
Terminology service

Metadata
The challenge of data harmonization

OMOP Common Data Model
(Observational Medical Outcomes Partnership)

https://www.ohdsi.org/data-standardization/
The challenge of data harmonization

2.4. Overview Diagram

PCORnet Common Data Model v3.0

http://www.pcornet.org/pcornet-common-data-model/
The MIRACUM consortium
First Pilot Results

The challenge of data harmonization

The Metadata Repository (MDR)

The MIRACUM Metadata Repository serves the harmonization of data content in the MIRACUM consortium. For this purpose it contains a collection of metadata (definitions of data elements compatible to the ISO/IEC 11179 standard) for the data elements included in the MIRACUM DIC research repositories. Those are defined compatible to the definition of the core data set defined by the BMBF-MI-NSG working group interoperability. In defining the data elements, it is important that their meaning is described accurately and that existing data elements are used if possible.
MIRACUM: towards a Learning Health System
Distributed Data Analysis

DataSHIELD: taking the analysis to the data, not the data to the analysis

Amadou Gaye,1 Yannick Marcon,2 Julia Isaeva,3 Philippe LaFlamme,2
Andrew Turner,1 Elinor M Jones,4 Joel Minion,1 Andrew W Boyd,1
Christopher J Newby,5 Marja-Liisa Nuotio,6,7 Rebecca Wilson,1
Oliver Butters,7 Barnaby Murtagh,9 Ipek Demir,9 Dany Doiron,2
Lisette Giepmans,10 Susan E Wallace,9 Isabelle Budin-Ljönsne,3
Carsten Oliver Schmidt,11 Paolo Boffetta,12 Mathieu Boniol,12
Maria Bota,12 Kim W Carter,13 Nick deKlerk,13 Chris Dibben,14
Richard W Francis,13 Tero Hiekkaninna,6,7 Kristian Hveem,15
Kirsti Kvaløy,15 Sean Millar,16 Ivan J Perry,16 Annette Peters,17
Catherine M Phillips,16 Frank Popham,18 Gillian Raab,14 Eva Reischl,17
Nuala Sheehan,8 Melanie Waldenberger,17 Markus Perola,6,7,19
Edwin van den Heuvel,20 John Macleod,1 Bartha M Knoppers,21
Ronald P Stolk,10,22 Isabel Fortier,2 Jennifer R Harris,3
Bruce HR Woffendenbuttel,22,23 Madeleine J Murtagh,241
Vincent Ferretti2,251 and Paul R Burton2,241*
Establishing a first DIC Infrastructure at all 8 MIRACUM partner sites

- Applying the i2b2-Plattform (Informatics for Integrating Biology & the Bedside)
  - Open Source, US-, European and German AUG
- Based on billing data for inpatient stays
  - Covers 5 of 7 data types of the basis module in the NSC core data set
  - Standardized data structures across all sites
- Application of this infrastructure for
  - Distributed analysis
  - Federated cohort identification
MIRACUM DIC 0.9
Distributed Analysis

MIRACUM inpatient coverage

MIRACUM disease category distribution
Cross-consortial cooperation
MIRACUM – HD4CR
- Charité, Ulm, Würzburg, Vivantes
- Provision of i2b2 instances
- Provision of ETL routine

Joint geovisualisation
- Years 2015 - 2016
applies broker / connector concept

- developed in previous projects (DKTK, DZL, GBN: German Biobank Node)
  - based on i2b2 research data repository and -webclient
  - connector polls the broker (IT security: no open ports for access from outside)

- supports federated cohort identification (aggregated patient counts)

- proof-of-concept query on colorectal cancer patients across all participating MIRACUM/HD4CR sites
MIRACUM DIC 0.9
federated cohort query across 10 hospitals

http://www.miracum.de/
http://www.hd4cr.org/
MIRACUM DIC 1.0
next steps – adding omics data

TranSMART Development
Evolve your genetic analysis with the power of TranSMART platform!
MIRACUM DIC 1.0
next steps – adding omics data
tranSMART Open API supports the development of new Analyse-Plugins, e.g. Kaplan-Meier Plots

1Knell C. et al. Developing interactive plug-ins for tranSMART using the SmartR framework: The case of survival analysis. Stud Health Inform 2017

2Christoph J. et al. Two Years of tranSMART in a University Hospital for Translational Research and Education. Stud Health Inform 2017
MIRACUM DIC 1.0
next steps – adding imaging data


MIRACUM DIC 1.0 - next steps: Adding Patient Generated Data

Quantified Self

Mandl KD, Kohane IS. Time for a Patient-Driven Health Information Economy?

Quantified Self
Wearables
Sensor Data

The MIRACUM example:
How to set up relevant and accessible Databases

The largest challenge?
The MIRACUM example:
How to set up relevant and accessible Databases

The largest challenge!

Data Quality
The MIRACUM example:
How to set up relevant and accessible Databases

Thank you very much!

ulli.prokosch@uk-erlangen.de