

# A clinical epidemiologist's perspective on digital medicine

Does digital medicine represent an epistemic shift in clinical epidemiology?

Prof. Dr. Ulrich Mansmann <u>ulrich.mansmann@lmu.de</u> Department of Medical Informatics, Biometry, and Epidemiology, LMU



## The perspective of clinical epidemiology

Clinical Epidemiology is ...

- .... a methodological discipline
  - which develops research strategies to produce high quality medical evidence. It is the basic science of Evidence-based medicine.
- .... a medical science

that studies the frequency and determinants of disease development, as well as the diagnostic and therapeutic approaches to disease management in clinical practice.













### The perspective of clinical epidemiology

Design Analysis Reporting Evidence Synthesis Implementation Decision Making













MEDIZIN

NFORMATH

Bundesministerium für Bildung und Forschung

### **Evidence Synthesis**





4



DIFUTURE Data Integration for Future Medicine



### **Multi-Resolution**

MEDIZIN

INITIATIVE

Data Integration for Future Medicine

Bundesministerium für Bildung und Forschung



### **Multi-Resolution**

Comprehensive pheno- and omics-typing

Data Sharing

**Distributed Computing** 







MEDIZIN

NITIATIVE

### Low risk group

	0+	<b>O-</b>	
T <sub>old</sub>	500	1500	2000
T <sub>new</sub>	250	750	1000

0 - outcome T - treatment

### High Risk group

	0+	О-	
T <sub>old</sub>	100	900	1000
T <sub>new</sub>	200	1800	2000



INITIATIVE

### Low risk group

### High Risk group

	0+	<b>O-</b>		P[O+]		0+	<b>O-</b>		P[O+]
T <sub>old</sub>	500	1500	2000	0.25	T <sub>old</sub>	100	900	1000	0.10
T <sub>new</sub>	250	750	1000	0.25	T <sub>new</sub>	200	1800	2000	0.10

Risk status is related

to treatment outcome and to treatment allocation.

Both treatments are equally effective in both risk populations.





MEDIZIN

NITIATIVE

#### Low risk group High Risk group 0-0-0+ 0+ 1500 900 500 2000 100 $\mathsf{T}_{\mathsf{old}}$ T<sub>old</sub> 1800 2000 $\mathsf{T}_{\mathsf{new}}$ 250 750 1000 200 T<sub>new</sub> All together **P[O+]** 0+ 0-600 2400 2000 0.30 T<sub>old</sub> 450 2550 1000 0.45 T<sub>new</sub>





### Experimental







Sir Bradford-Hill: 1897-1991 Paul Martini: 1889-1964



Jerzy Newman 1894-1981 Donald Rubin James Robins



## Observational







# Treatment Response Modifying Predictive Factors



**RESEARCH ARTICLE** 

Gender-Specific Differences in Low-Dose Haloperidol Response for Prevention of Postoperative Nausea and Vomiting: A Register-Based Cohort Study

Florian Brettner<sup>1®</sup>\*, Silke Janitza<sup>2,3®</sup>, Kathrin Prüll<sup>4</sup>, Ernst Weninger<sup>1</sup>, Ulrich Mansmann<sup>2</sup>, Helmut Küchenhoff<sup>3</sup>, Alexander Jovanovic<sup>1</sup>, Bernhard Pollwein<sup>1</sup>, Daniel Chappell<sup>1</sup>, Bernhard Zwissler<sup>1</sup>, Vera von Dossow<sup>1</sup>

1 Department of Anaesthesiology, University Hospital of Munich, Munich, Germany, 2 Department of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany, 3 Statistical Consulting Unit Department of Statistics, Ludwig-Maximilians-Universität München, Munich, Germany, 4 Anaesthesia and Intensive Care Medicine Clinic, Klinikum Dritter Orden, Munich, Germany



© These authors contributed equally to this work. \* Florian.Brettner@med.uni-muenchen.de







### Learning health system

### Better Health = D2K + K2P



Charles P. Friedman University of Michigan



### Multi-Phase

There is a long way from the natural object to the respective quantitative data

When data have been pre-processed by multiple teams, one typically doesn't know in the end what assumptions were made or what cleaning processes were used, but whatever things previous people have done would have an impact.

Are we making inferences about the real nature, or about the second natures created about the pre-processors, or a bit of both?











### Multi-Model



Died 5<sup>th</sup> of July 2005





GERÖNDERT VOM Bundesministerium für Bildung und Forschung

### The Algorithmic Modeling Culture

The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function  $f(\mathbf{x})$ —an algorithm that operates on  $\mathbf{x}$  to predict the responses  $\mathbf{y}$ . Their black box looks like this:



Model validation. Measured by predictive accuracy. *Estimated culture population*. 2% of statisticians, many in other fields.

Leo Breiman (2001) Statistical Modeling: The Two Cultures, Statistical Science 16: 199–231

## Insight from machine learning algorithms

- A barrier to the adoption of machine learning: Computers usually don't explain their predictions.
- Model-free approaches as a starting point to gain insight (Breiman, 2001)
- Interpretation, subject matter information

GEFÖRDERT VON

Bundesministeriun für Bildung und Forschung

- Concepts of interpretability and its evaluation
- Identification of important variables
- Feature Interaction
- Representation of partial dependencies
- Etc.

Data Integration for Future Medicine



Bernd Bischl LMU, Stat.



https://christophm.github.io/interpretable-ml-book/

# What when we try to understand the world through data?

Statisticians have spent the past 200 years figuring out what traps lie in wait when we try to understand the world through data.

We must not pretend that the traps have all been made safe.

Clinical Epidemiology is doing this job in medical research since 80 years.



Harford T (2014) Big Data: Are we making a mistake? Significance 11(5):14-19





## Quality versus Quantity

"Which one should I trust more: a 1% survey with 60% response rate or a self-reported administrative dataset covering 80% of the population?"

How to compare two datasets with different quantities and different qualities?



XIAO-LI MENG (2018) STATISTICAL PARADISES AND PARADOXES IN BIG DATA (I) The Annals of Applied Statistics Vol. 12, No. 2, 685-726

$$\overline{G}_n - \overline{G}_N = \underbrace{\rho_{R,G}}_{\text{Data Quality}} \times \underbrace{\sqrt{\frac{1-f}{f}}}_{\text{Data Quantity}} \times \underbrace{\sigma_G}_{\text{Problem Difficulty}}$$
N: Population size; n; size of data set N > n, f= n/N



MEDIZIN

### Social Value of health research

Cultural value

Biotech + Al

Yuval Noah Harari



Scientific value

Causality - Association



Health related value

Patient centered evaluation Patient participation



Daniel Strech, Marcel Mertz, Forschungsethische Grundprinzipien, in Raspe, Hüppe, Strech, Taupitz (eds) Empfehlungen zur Begutachtung klinischer Studien durch Ethik-Kommissionen, 2. Auflage, 2012





MEDIZIN

NFORMATI

## Summary

- Medical Informatics provides technologies and infrastructure to booster medical research.
- Cooperation between experts in Medical Informatics and Clinical Epidemiology will allow to explore the full potential of the digitalization of medicine
- Evaluation concepts are needed to qualify the added value of specific projects in digital medicine
- Implementing elements of digital medicine into the health care system represent complex interventions which can be evaluated using strategies of clinical epidemiology (RCTs, Cluster randomized studies, edge wedge trials)
- Does digital medicine represent an epistemic shift in clinical epidemiology? No!





### DIZKUM







MEDIZIN

INITIATIVE