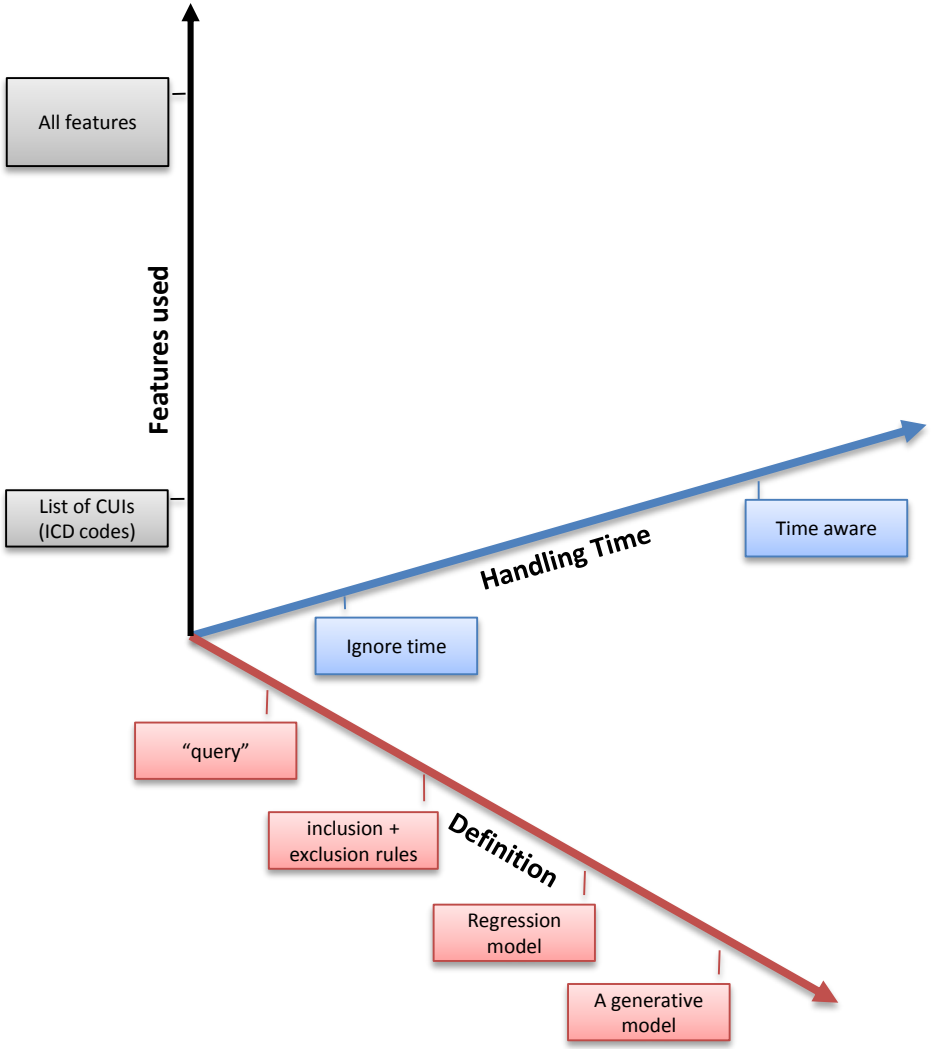
Data Mining for Medical Informatics: Electronic Phenotyping

Neils Peek, Nigam Shah, Gregor Stiglic, Fei Wang

STANFORD
SCHOOL OF MEDICINE

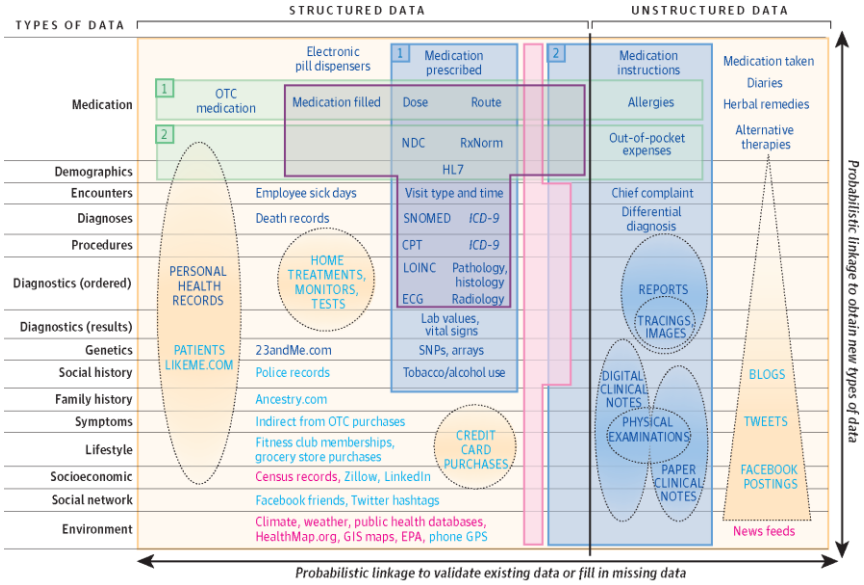
Electronic phenotyping

- Identifying a set of patients:
  - For observational research
  - For clinical trial eligibility,
  - As Numerators or denominators of quality metrics
  - For whom a decision support reminder should “fire”
  - Who are “similar” based on whom a clinical decision should be based.
  - Who progress along similar paths
- The main problems:
  - the need for a gold standard
  - poor portability across sites and studies



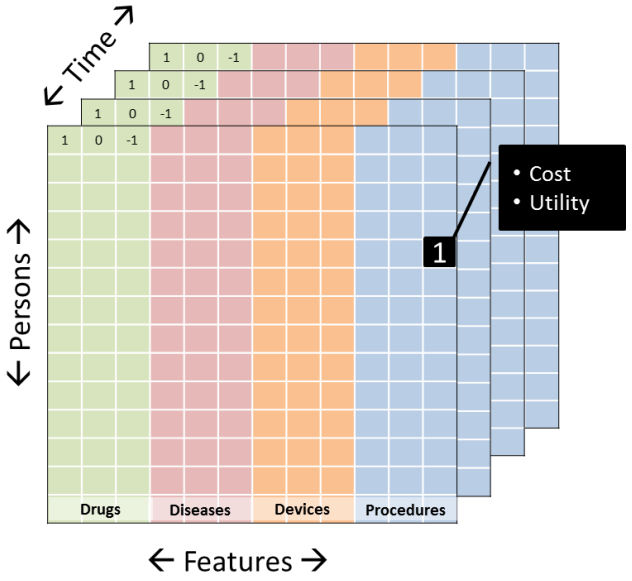
Features used in phenotyping

The source of features



Weber et al, JAMA 2014

The choice of features



Shah NH, Using Big Data, in
Translational Informatics: Realizing the Promise of Knowledge-Driven Healthcare,
Springer-Verlag, London

Agenda for the day

- Opening keynote – Joshua Denny
- Panel on ‘state of the art’ – Shawn Murphy, Patrick Ryan, Jyoti Pathak, Maryan Zirkle
- Four workshop papers

Lunch

- Four short papers
- Open problems – George Hripcsak
- Closing keynote – Iain Buchan

Acknowledgements

- **Mohsen Bayati**, Stanford University
- **Riccardo Bellazzi**, University of Pavia
- **Adam Davey**, Temple University
- **Joydeep Ghosh**, University of Texas, Austin
- **Tudor Groza**, The University of Queensland
- **Joyce Ho**, University of Texas, Austin
- **John Holmes**, University of Pennsylvania
- **Jianying Hu**, IBM T.J. Watson Research Center
- **Siddhartha Jonnalagadda**, Northwestern University
- **Jin-Dong Kim**, Database Center for Life Science
- **Robert Moskovitch**, Columbia University
- **Zoran Obradovic**, Temple University
- **Mykola Pechenizkiy**, Eindhoven University of Technology
- **Mattia Prosperi**, University of Manchester
- **Lucia Sacchi**, University of Pavia
- **Suchi Saria**, Johns Hopkins University
- **Nicholas Tatonetti**, Columbia University