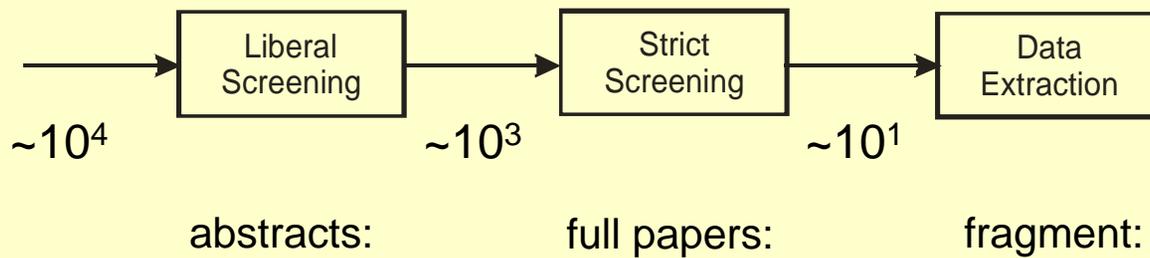


# Classifying medical abstracts

- Evidence-based medicine:  
“is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients”
- Systematic Reviews  
the foundation of evidence-based medicine is the systematic review of evidence for particular treatments, mainly randomized controlled trials

# How is it done?



- Trialstat
  - SRS: Web-based tool

# First observation

- Liberal screening is a text classification problem
- Let's apply off-the shelf methods
- EMISAR
- Low-hanging fruit?

# Data

- Abstracts and titles
- 20-40K instances
- No authors, references, dates
- Imbalance ~ 10-90

# Requirements

- 100% recall (!)
  - Is this realistic?
  - Does the human-manual system achieve this?
  - Perhaps recall at least as good as as good as human (90-95%)
- ...and “decent” precision (at least as good as as good as human – 80-85%)
- Output a score, so that abstracts can be sorted by their relevance - ranking

# Course project

- We will work with much easier (smaller) data
- We will follow the project described in Cohen 2006, Cohen 2008, and our 2010 paper (see course website under papers)