

Text-mining the Estonian Electronic Health Record

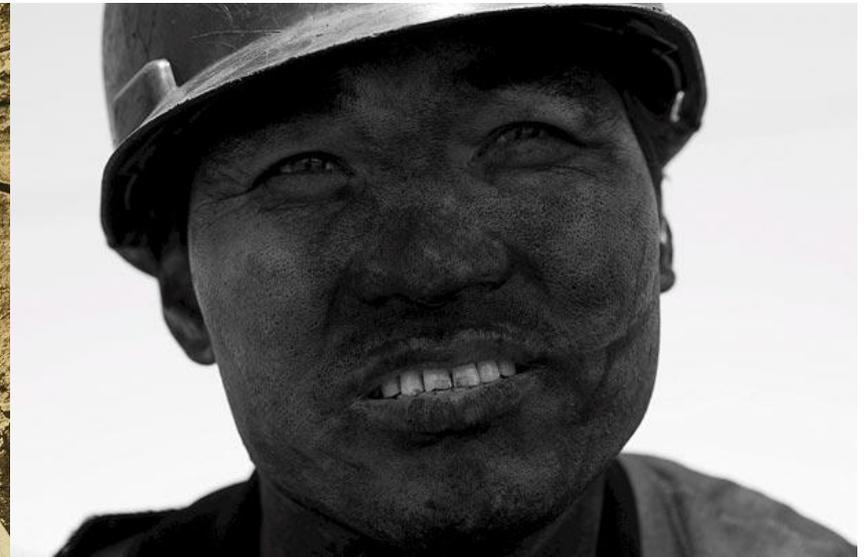
27.09.2012 @ Tartu

Raul Sirel

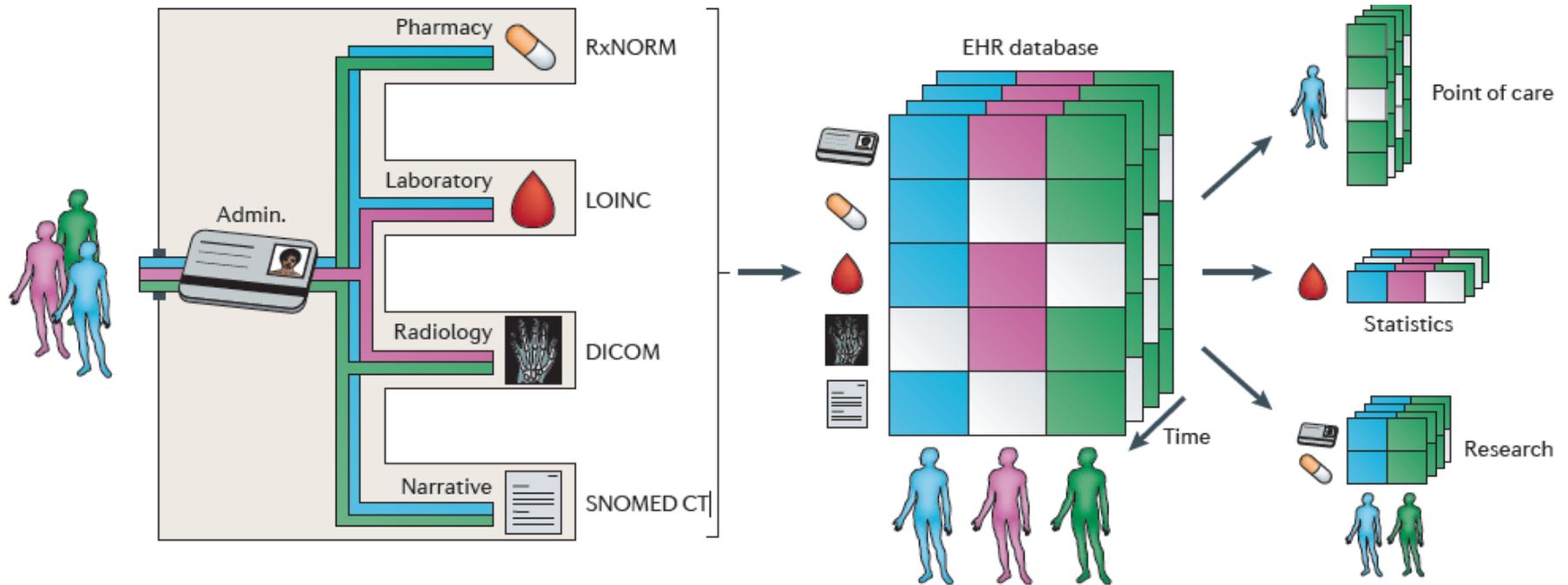
rsirel@ut.ee

Text Mining

- **Text mining is a process of deriving high-quality information from texts**



Electronic Health Record (EHR)



Peter B. Jensen, Lars J. Jensen and Søren Brunak. 2012. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics 13, 395-405.

Yet some data is quite stubborn

- Lot's of data is “locked away” as narrative texts
- Processing textual data is difficult:



Yet some data is quite stubborn

- Lot's of data is “locked away” as narrative texts
- **Processing textual data is difficult:**
 - Morphology
 - Syntax
 - Semantics
 - Pragmatics



What is this stubborn data like?

- **Abbreviations**
 - is Ca *Calcium* or *Circa*?
 - is Vas. *vascular* or *vasectomy*?

What is this stubborn data like?

- **Abbreviations**

- is Ca *Calcium* or *Circa*?
- is Vas. *vascular* or *vasectomy*?



What is this stubborn data like?

- **Abbreviations**

- is Ca *Calcium* or *Circa*?
- is Vas. *vascular* or *vasectomy*?

- **Latin terminology**

- Estonian and foreign terms are used intermittently

What is this stubborn data like?

- **Abbreviations**
 - is Ca *Calcium* or *Circa*?
 - is Vas. *vascular* or *vasectomy*?
- **Latin terminology**
 - Estonian and foreign terms are used intermittently
- **Almost no syntax**
 - Parsers do lousy job on clinical texts

What is this stubborn data like?

- **Abbreviations**

- is Ca *Calcium* or *Circa*?
- is Vas. *vascular* or *vasectomy*?

- **Latin terminology**

- Estonian and foreign terms are used intermittently

- **Almost no syntax**

- Parsers do lousy job on clinical texts

- **Misspellings and misprints**

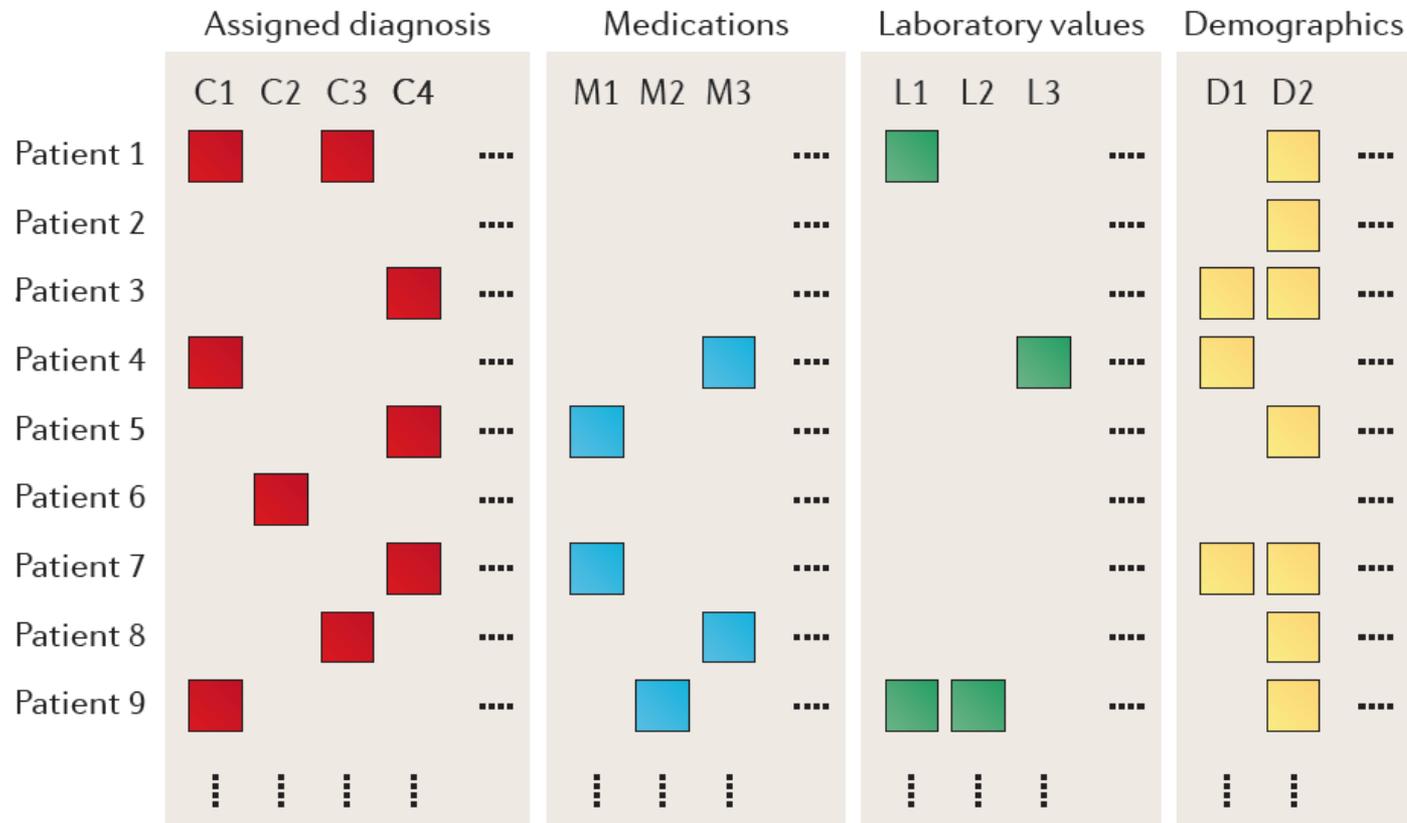
- Many practitioners not fluent in Estonian
- Very little time to type



What kind of questions can we ask from the data?

- **How do events correlate in EHR data?**

One possible way of analysing EHR data

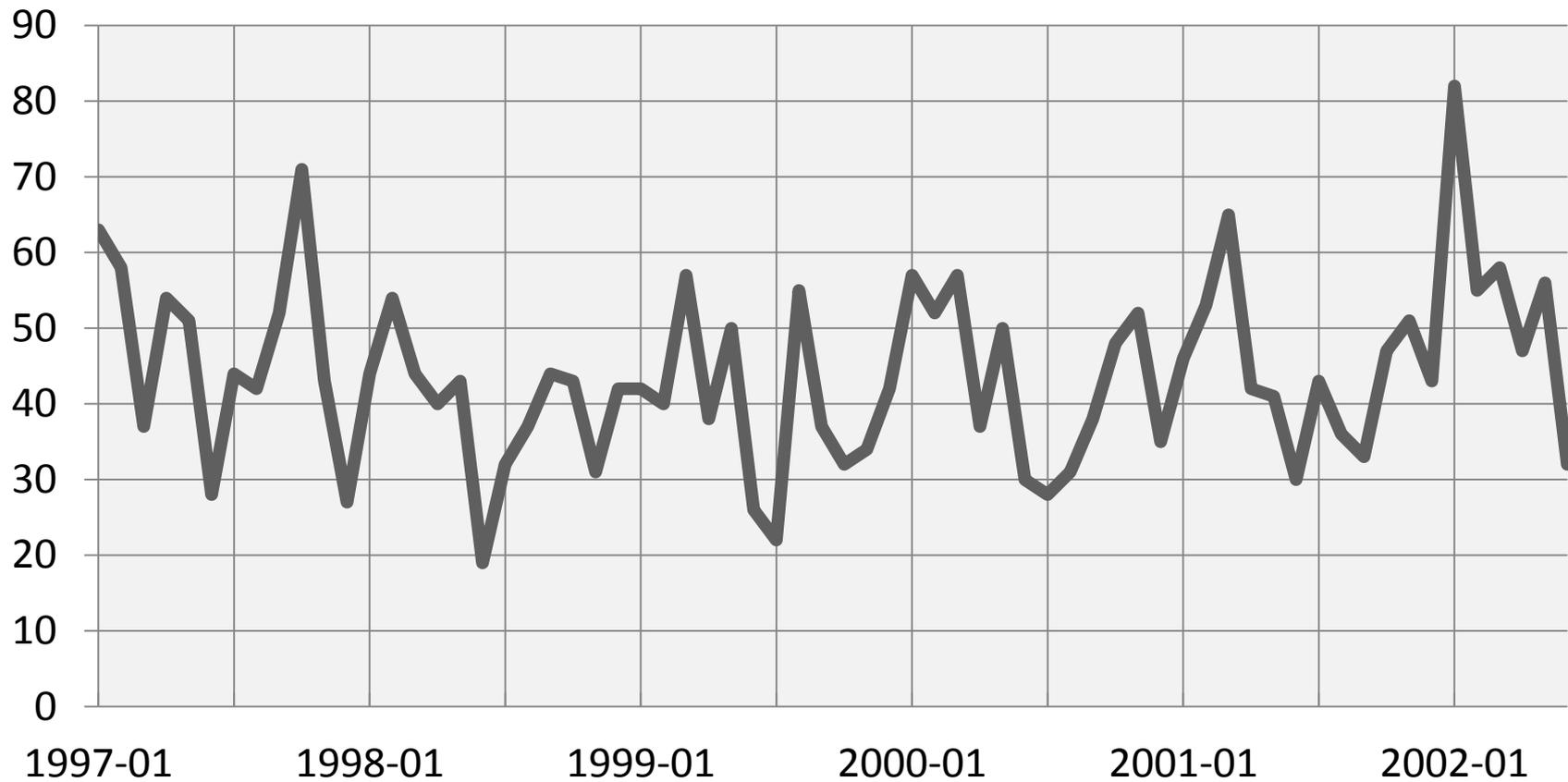


Peter B. Jensen, Lars J. Jensen and Søren Brunak. 2012. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics 13, 395-405.

What kind of questions can we ask from the data?

- **How do facts correlate in EHR data?**
- **Injuries occur... How many and when?**

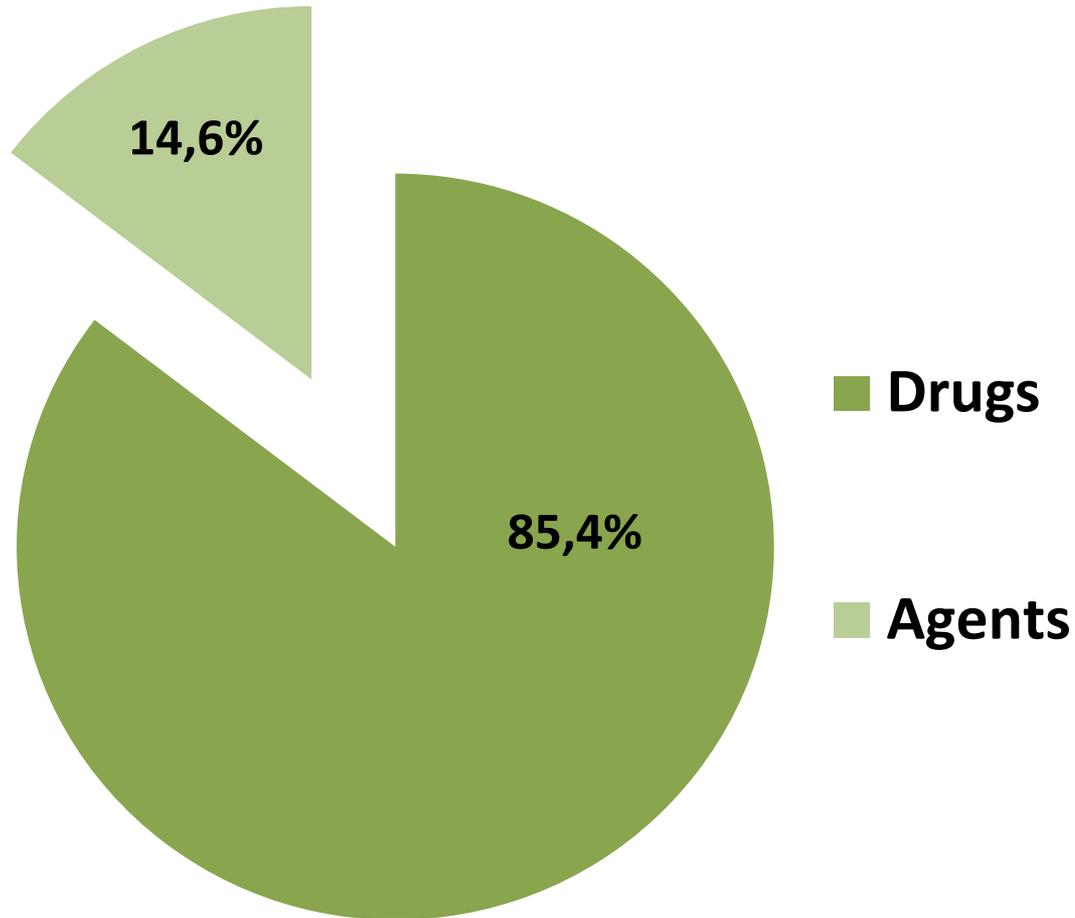
Falling-related injuries in Tartu county in 1997-2002



What kind of questions can we ask from the data?

- **How do facts correlate in EHR data?**
- **Injuries occur... How many and when?**
- **Practitioners should be referring to chemical agents rather than drugs. So do they?**

EHR records from 2008-2011

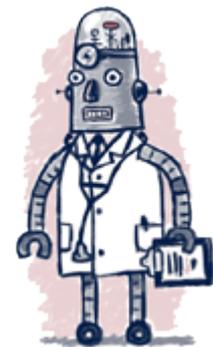


Why are we doing it?

- **Benefits of mining the EHR data:**
 - Revealing unknown disease correlations
 - Revealing unknown side-effects of medicines
 - Evaluating the effectiveness of medicines

Why are we doing it?

- **Benefits of mining the EHR data:**
 - Revealing unknown disease correlations
 - Revealing unknown side-effects of medicines
 - Evaluating the effectiveness of medicines
- **Possible new technologies:**
 - Improved diagnostic algorithms
 - Predictive interfaces for practitioners
 - Personalised online self-diagnosis systems for patients



Why are we doing it?

- **Benefits of mining the EHR data:**
 - Revealing unknown disease correlations
 - Revealing unknown side-effects of medicines
 - Evaluating the effectiveness of medicines
- **Possible new technologies:**
 - Improved diagnostic algorithms
 - Predictive interfaces for practitioners
 - Personalised online self-diagnosis systems for patients
- **Integration with genetic data**

The future

- **New data is produced every day!**
- **Ongoing research**
- **More and more researchers all over the world interested in clinical texts**

Thank You!