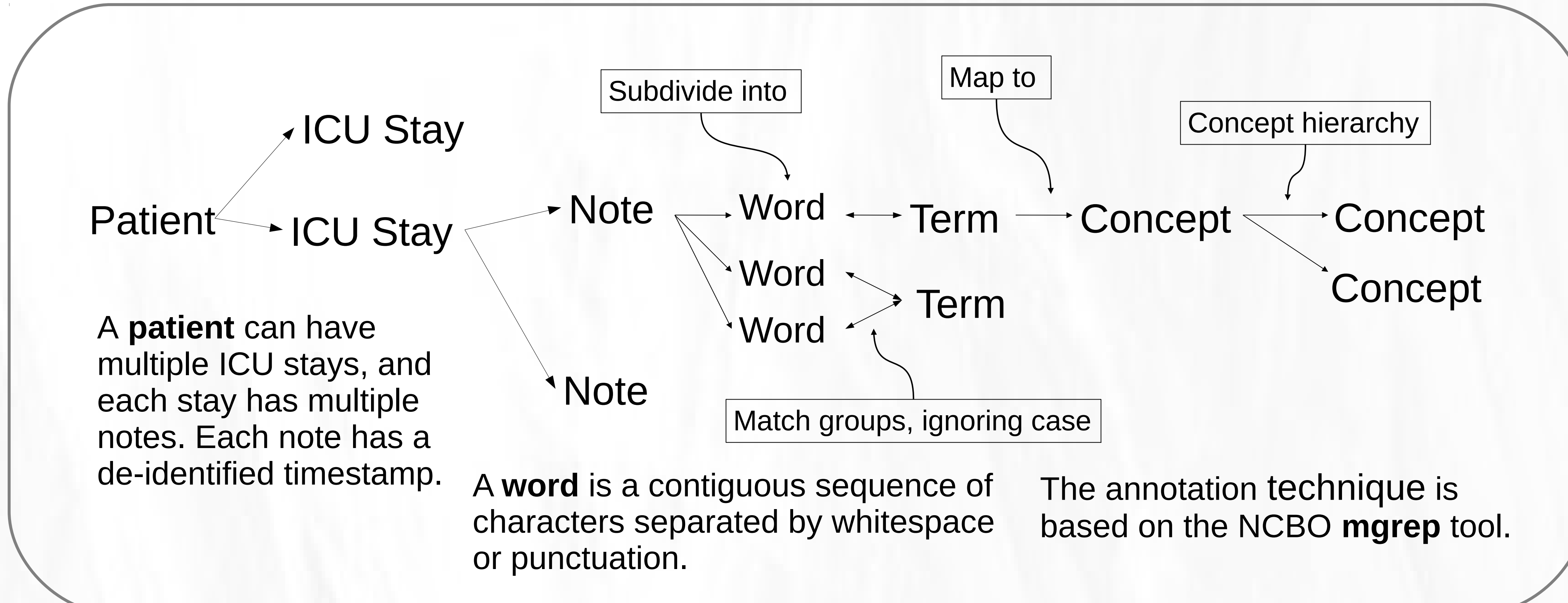


# Relationships between the biomedical terminology used by clinicians and medical test measurement results in intensive care units

Kenneth Baclawski  
College of Computer and Information Science  
Northeastern University

## Extracting terminology from ICU notes



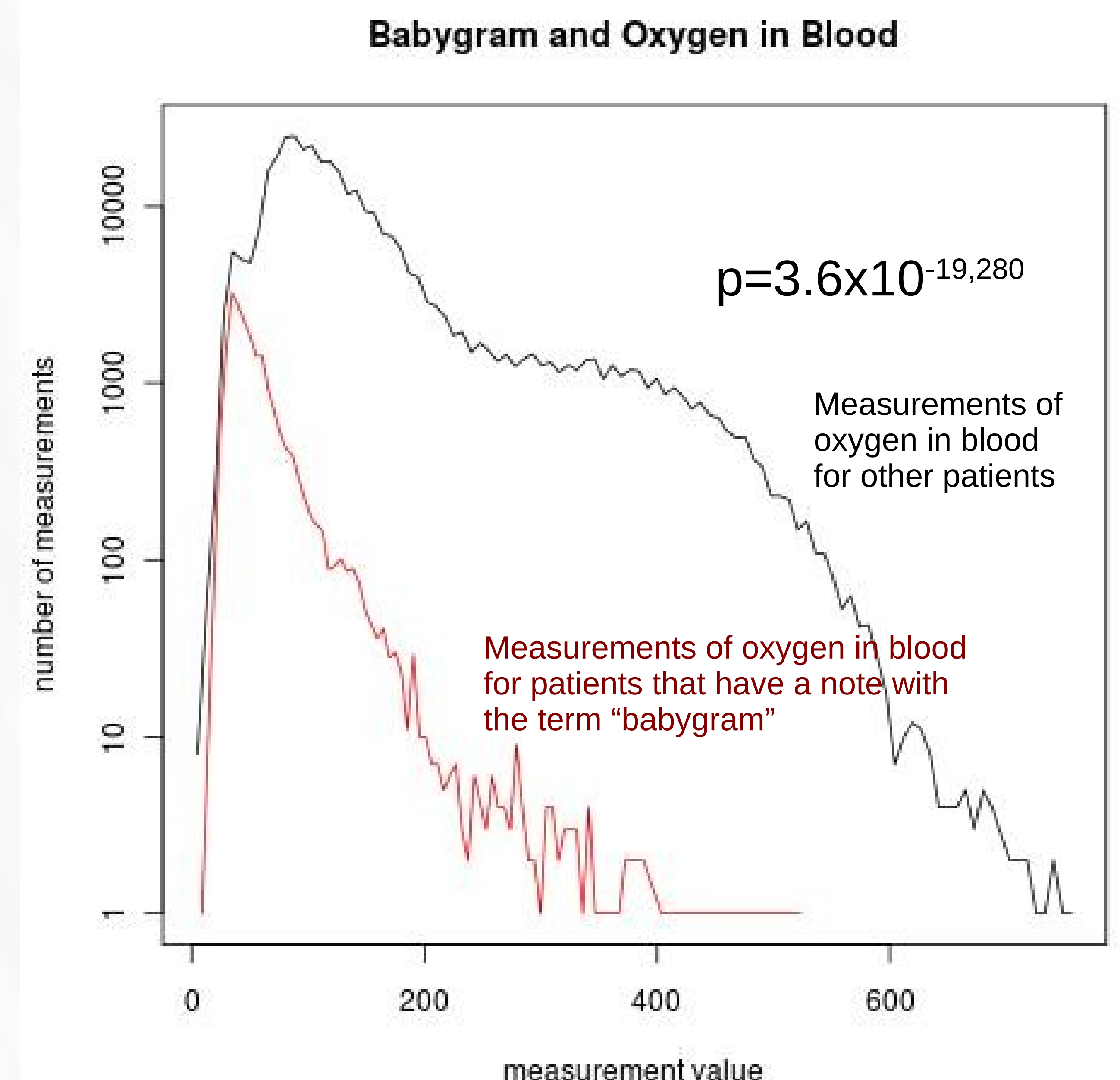
## Vocabularies

SNOMED CT®  
Medical Subject Headings (MeSH)  
Foundational Model of Anatomy (FMA)  
Current Procedural Terminology (CPT)  
Veterans Health Administration National Drug File  
RxNorm normalized names for clinical drugs  
RadLex controlled terminology for radiology  
12dixts English vocabulary  
MIMIC-II Vocabulary

## Relating terminology usage with medical test measurements

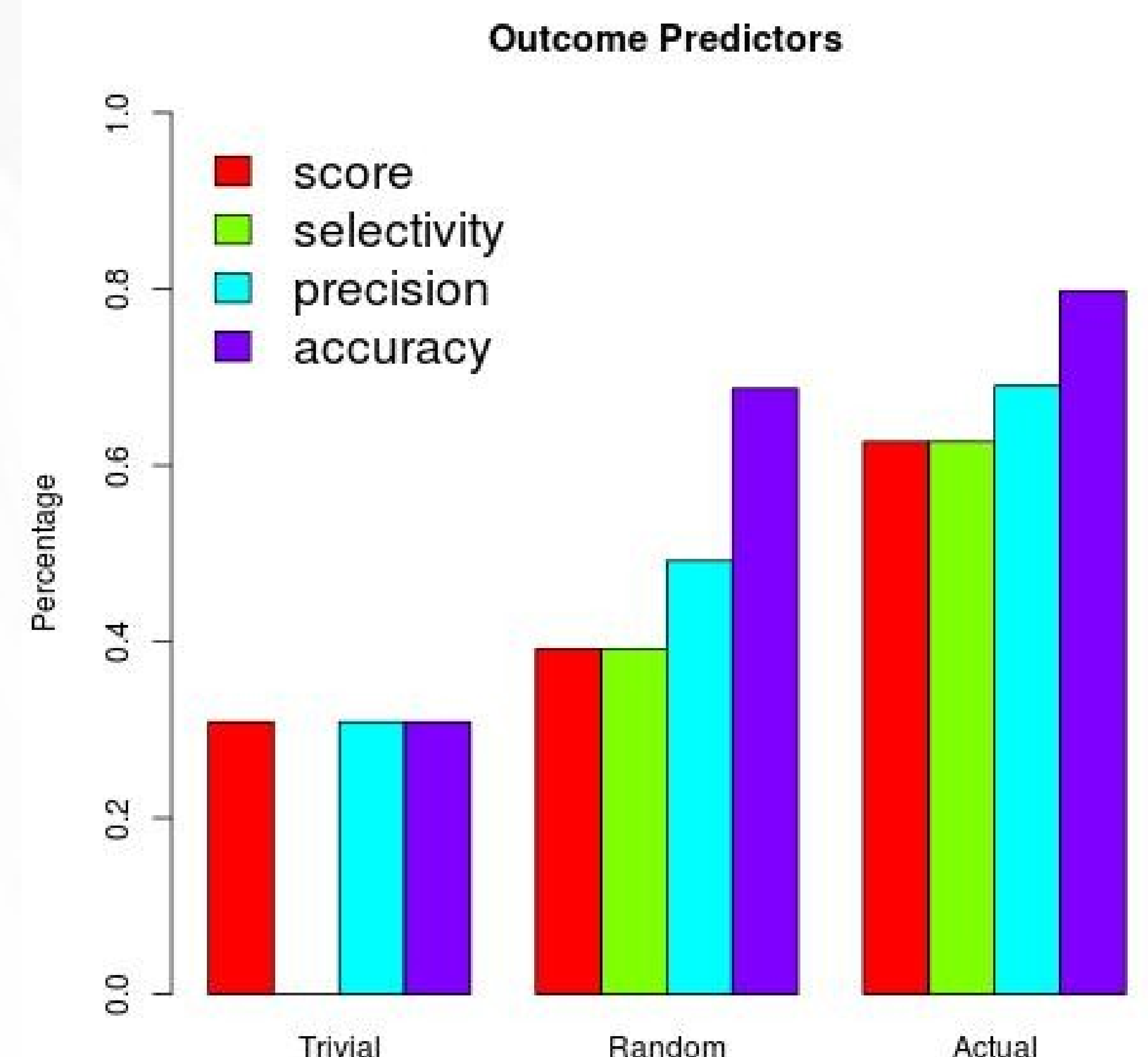
### Statistical Methods

- Comparison is between two samples:
  - Lab measurements of one type done on behalf of a patient for whom a biomedical term was used in some note
  - Lab measurements of the same type done on behalf of other patients
- Statistical hypothesis tests:
  - When the samples are large (>100), the two-sided t-test was used (Welch's t-test).
  - For smaller samples, the two-sided Mann-Whitney U test was used (Wilcoxon rank-sum test).
- Corrections:
  - Bonferroni (multiple tests on the same data)
  - Compare with random case (because measurements are correlated)



## Predicting ICU stay outcomes using terminology usage

- A Support Vector Machine was used to predict the ICU stay outcome.
- Only notes with timestamp at least 48 hours before the end of the ICU stay were used.
- The libsvm (with kernel) and liblinear (without kernel) tools were used.
- The liblinear tool produced much better models.
- Terms with very small coefficients were dropped.
- Demographic data did not have a significant effect.
- A 10-fold cross-validation was used.
- Three models were compared:
  - Trivially predicting a one outcome in all cases.
  - Using randomly generated term usage data.
  - Using the actual term usage data.



## References

Clement Jonquet, Nigam H. Shah, Mark A. Musen, The Open Biomedical Annotator, AMIA Summit on Translational Bioinformatics, p. 56-60, March 2009, San Francisco, CA, USA.

Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification Journal of Machine Learning Research 9(2008), 1871-1874.

### Some of the coefficients

0.268 Malignant tumor (disorder)  
0.256 Ordained priest (occupation)  
...  
-0.241 Cramping  
-0.233 Infant (person)  
...  
0.193 Non-Reactive (qualifier value)  
0.192 Antineoplastic agent  
0.192 Niece (person)  
...