

# Identifying Patients for Clinical Studies from Electronic Health Records: The TREC Medical Records Track

William Hersh, MD  
Professor and Chair  
Department of Medical Informatics & Clinical Epidemiology  
School of Medicine  
Oregon Health & Science University  
Email: [hersh@ohsu.edu](mailto:hersh@ohsu.edu)  
Web: [www.billhersh.info](http://www.billhersh.info)  
Blog: [informaticsprofessor.blogspot.com](http://informaticsprofessor.blogspot.com)

## OHSU Research Week – VERY SHORT

You can find a longer version of this talk:

<http://oninformatics.com/?p=834>

These slides on my web site:

<http://www.billhersh.info>

There is also more about the TREC Medical Records Track on my blog:

<http://informaticsprofessor.blogspot.com>

## References

- Bedrick, S., Ambert, K., et al. (2011). Identifying Patients for Clinical Studies from Electronic Health Records: TREC Medical Records Track at OHSU. *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, Gaithersburg, MD. National Institute for Standards and Technology.
- Blumenthal, D. (2011a). Implementation of the federal health information technology initiative. *New England Journal of Medicine*, 365: 2426-2431.
- Blumenthal, D. (2011b). Wiring the health system--origins and provisions of a new federal program. *New England Journal of Medicine*, 365: 2323-2329.
- Demner-Fushman, D., Abhyankar, S., et al. (2011). A knowledge-based approach to medical records retrieval. *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, Gaithersburg, MD. National Institute for Standards and Technology.
- Hersh, W. (2009). *Information Retrieval: A Health and Biomedical Perspective (3rd Edition)*. New York, NY. Springer.
- Hersh, W. and Voorhees, E. (2009). TREC genomics special issue overview. *Information Retrieval*, 12: 1-15.
- Ide, N., Loane, R., et al. (2007). Essie: a concept-based search engine for structured biomedical text. *Journal of the American Medical Informatics Association*, 14: 253-263.
- King, B., Wang, L., et al. (2011). Cengage Learning at TREC 2011 Medical Track. *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, Gaithersburg, MD. National Institute for Standards and Technology.

Safran, C., Bloomrosen, M., et al. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *Journal of the American Medical Informatics Association*, 14: 1-9.

Voorhees, E. and Harman, D., eds. (2005). *TREC: Experiment and Evaluation in Information Retrieval*. Cambridge, MA. MIT Press.

# Identifying Patients for Clinical Studies from Electronic Health Records: The TREC Medical Records Track

William Hersh, MD  
Professor and Chair  
Department of Medical Informatics & Clinical Epidemiology  
School of Medicine  
Oregon Health & Science University  
Email: [hersh@ohsu.edu](mailto:hersh@ohsu.edu)  
Web: [www.billhersh.info](http://www.billhersh.info)  
Blog: [informaticsprofessor.blogspot.com](http://informaticsprofessor.blogspot.com)



1

## Caveat

- I cannot even scratch the surface of this project in 10 minutes
- BUT
- You can find a longer version of this talk
  - <http://oninformatics.com/?p=834>
- And these slides on my web site
  - <http://www.billhersh.info>
- As well as materials on my blog
  - <http://informaticsprofessor.blogspot.com>



2

## Motivations for secondary use of clinical data

- Many “secondary uses” or re-uses of electronic health record (EHR) data, including (Safran, 2007)
  - Personal health records (PHRs)
  - Clinical and translational research – generating hypotheses and facilitating research
  - Health information exchange (HIE)
  - Public health surveillance for emerging threats
  - Healthcare quality measurement and improvement
- Little controlled research on how best to do it
- Opportunities facilitated by growing incentives for “meaningful use” of EHRs in the HITECH Act (Blumenthal, 2011; Blumenthal, 2011)

3



## Information retrieval (IR) evaluation (Hersh, 2009)

- Assessed with *test collections*, which consist of
  - Content – fixed yet realistic collections of content
  - Topics – statements of information need
  - Relevance judgments – by expert humans for which content items should be retrieved for which topics
- Evaluation consists of *runs* using a specific IR approach with output for each topic measured and averaged across topics
- Variety of measures to assess retrieval of “relevant” information, e.g., recall, precision, and aggregations thereof

4



## Challenge evaluations

- A common approach in computer science, not limited to IR
- Develop a common task, data set, evaluation metrics, etc., ideally aiming for real-world size and representation for data, tasks, etc.
- In IR, oldest and largest is Text Retrieval Conference (TREC, [trec.nist.gov](http://trec.nist.gov); Voorhees, 2005) – sponsored by National Institute for Standards and Technology (NIST)
  - Many “tracks” of interest, such as routing/filtering, Web searching, question-answering, etc.
  - Operates on annual cycle of test collection release, experiments, and analysis of results
  - Non-medical, with exception of Genomics Track (Hersh, 2009)

5



## TREC Medical Records Track

- Facilitated by availability of a large-scale, de-identified corpus of medical records from University of Pittsburgh Medical Center (UPMC)
- Task: identify patients for possible inclusion in clinical research studies
- Topic development and relevance assessment carried out by OHSU
- Participation of 29 research groups who could submit up to 8 runs each
  - Total of 127, with 109 automatic and 18 manual
  - Each run scored with bpref averaged over topics

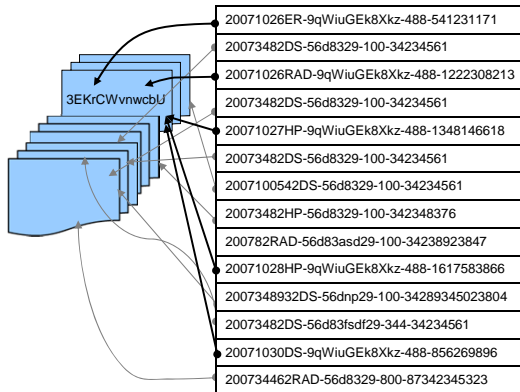
6



# Test collection

VISIT LIST

RECORD-VISIT MAP



DISCHARGE SUMMARY  
 ...  
 PRINCIPAL DIAGNOSES:  
 1. Urinary tract infection.  
 2. Gastroenteritis.  
 3. Dehydration.  
 4. Hypoglycemia.  
 5. Diabetes mellitus.  
 6. Osteoarthritis.  
 7. History of anemia.  
 8. History of tobacco use.  
 HOSPITAL COURSE: The patient is a \*\*AGE[in 40s]  
 -year-old insulin-dependent diabetic who  
 presented with nausea,...

Report Extract

17,198 visits

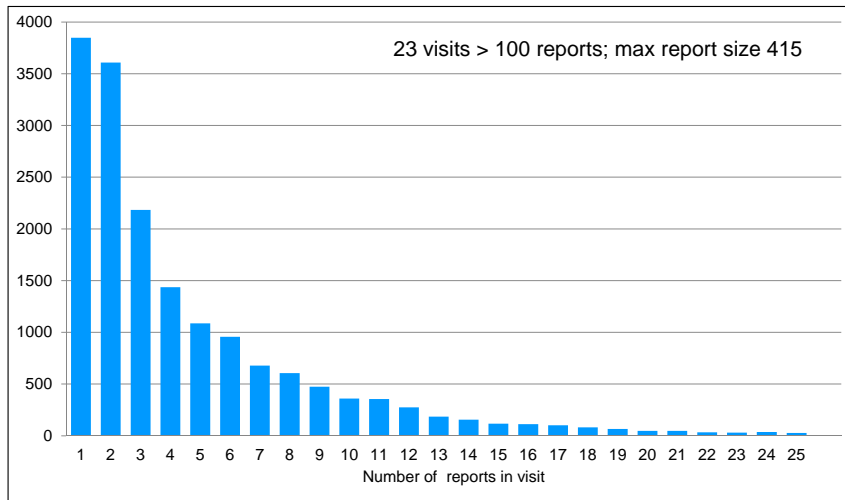
101,712 reports (93,552 mapped to visits)

7

(Courtesy, Ellen Voorhees, NIST)



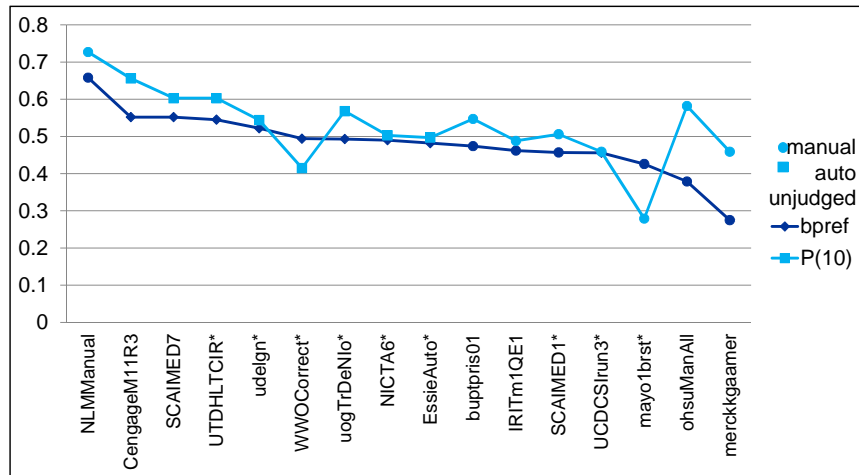
# Wide variations in number of documents per visit



8

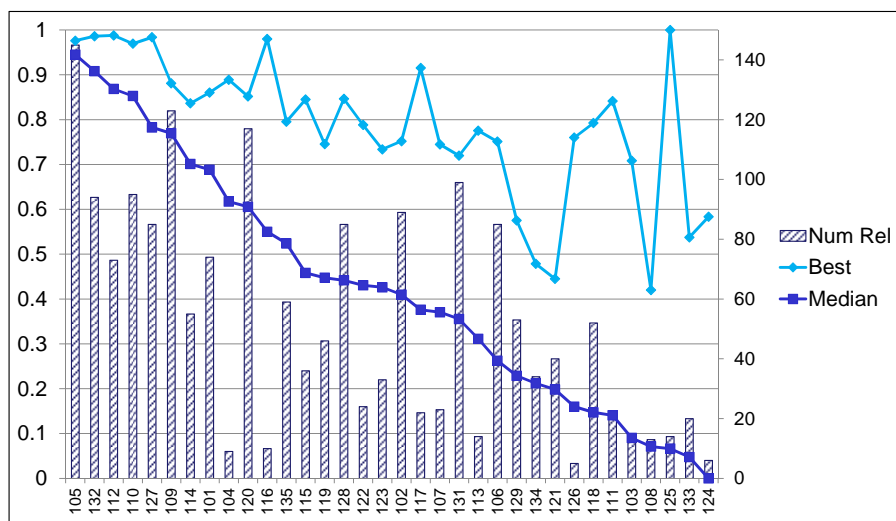
(Courtesy, Ellen Voorhees, NIST)

## Evaluation results for top runs ...



9

## ... BUT, wide variation among topics



10

## Easy and hard topics

- Easiest – best median bpref
  - 105: Patients with dementia
  - 132: Patients admitted for surgery of the cervical spine for fusion or discectomy
- Hardest – worst best bpref and worst median bpref
  - 108: Patients treated for vascular claudication surgically
  - 124: Patients who present to the hospital with episodes of acute loss of vision secondary to glaucoma
- Large differences between best and median bpref
  - 125: Patients co-infected with Hepatitis C and HIV
  - 103: Hospitalized patients treated for methicillin-resistant Staphylococcus aureus (MRSA) endocarditis
  - 111: Patients with chronic back pain who receive an intraspinal pain-medicine pump

11



## What approaches did (and did not) work?

- Best results obtained from NLM group (Demner-Fushman, 2011)
  - Top results from manually constructed queries using Essie domain-specific search engine (Ide, 2007) – BPref = 0.658
  - Other automated processes fared less well, e.g., creation of PICO frames, negation, term expansion, etc. – BPref = 0.4822
- Best automated results obtained by filtering by age, race, gender, admission status; terms expanded by UMLS Metathesaurus – BPref = 0.552 (King, 2011)
- Benefits of approaches commonly successful in IR did provided small or inconsistent value for this task
- Nor did manual queries and using ICD-9 codes (Bedrick, 2009)

12





## Conclusions and future directions

- TREC Medical Records Track extended IR challenge evaluation approach to a patient selection triage task
- Initial results show mixed success for different methods – common with a new IR task
- Future work will aim to expand test collection, tasks, and topics – aiming to develop best approaches for variety of tasks