Identifying treatments, groups, and outcomes in medical abstracts

Rodney L. Summerscales
Shlomo Argamon
(Illinois Institute of Technology)

Jordan Hupert
Allen Schwartz
(University of Illinois at Chicago Medical School)

Evidence Based Medicine (EBM)

- Make treatment decisions based on analysis of *current research*

- **Problem:**
  - A lot of medical research
  - Research results not presented in standard ways
    - Difficult to apply results

- **Solution:**
  - Summarize results
    - Extract key information
      - e.g. Treatments, group size, number of good/bad outcomes
    - Include decision tools
      - Risk Reduction Calculator (Schwartz, 2006)
    - Currently created by hand
      - Need to automate

Sample abstract

Use of probiotic Lactobacillus preparation to prevent diarrhoea associated with antibiotics: randomised double blind placebo controlled trial

Mary Hickson et al.

Objective: To determine the efficacy of a probiotic drink containing Lactobacillus for the prevention of any diarrhoea associated with antibiotic use and that caused by Clostridium difficile.

Design: Randomised double blind placebo controlled study.

Participants: 135 hospital patients (mean age 74) taking antibiotics. Exclusions included diarrhoea on admission, bowel pathology that could result in diarrhoea, antibiotic use in the previous four weeks, severe illness, immunosuppression, bowel surgery, artificial heart valves, and history of rheumatic heart disease or infective endocarditis.

Intervention: Consumption of a 100 g (97 ml) drink containing Lactobacillus casei, L bulgaricus, and Streptococcus thermophilus twice a day during a course of antibiotics and for one week after the course finished. The placebo group received a longlife sterile milkshake.

Main outcome measures: Primary outcome: occurrence of antibiotic associated diarrhoea. Secondary outcome: presence of C difficile toxin and diarrhoea.

Results: 7/57 (12%) of the probiotic group developed diarrhoea associated with antibiotic use compared with 19/56 (34%) in the placebo group (P=0.007). Logistic regression to control for other factors gave an odds ratio 0.25 (95% confidence interval 0.07 to 0.85) for use of the probiotic, with low albumin and sodium also increasing the risk of diarrhoea. The absolute risk reduction was 21.6% (6.6% to 36.6%), and the number needed to treat was 5 (3 to 15). No one in the probiotic group and 9/53 (17%) in the placebo group had diarrhoea caused by C difficile (P=0.001). The absolute risk reduction was 17% (7% to 27%), and the number needed to treat was 6 (4 to 14).

Conclusion: Consumption of a probiotic drink containing L casei, L bulgaricus, and S thermophilus can reduce the incidence of antibiotic associated diarrhoea and C difficile associated diarrhoea. This has the potential to decrease morbidity, healthcare costs, and mortality if used routinely in patients aged over 50.

EBM summary

Lactobacillus;BMJ;2007;335;80
http://www.bmj.com/cgi/content/full/335/7610/80

Groups: **Lactobacillus** and **placebo**. The patients were inpatients, over 50 years of age, taking antibiotics.

Treatment: Same

Endpoints:
- Primary Outcome: **Occurrence of antibiotic associated diarrhea**
- Secondary Outcome: **Presence of C. Difficile toxin and diarrhea**

Outcomes:
- Primary: **Diarrhea**
    ARR: 0.22 [0.07, 0.37], NNT: 5 [3, 16]
  Worse case scenario
    ARR: 0.15 [0.005, 0.29], NNT: 7 [4, 198]
  Worst case scenario
    ARR: 0.01 [-0.14, 0.17]

Calculation: http://araw.mede.uic.edu/cgi-bin/nntcalc.pl?EB=7;EG=50;EL=12;CB=19;CG=37;CL=10;2x2=Compute
...

First step: Identify key entities

- Need to identify sections of text that describe or refer to
  - Treatments
    - Examples:
      - *tamoxifen*
      - *placebo pill*
      - *acupuncture*
      - *surgical stabilisation of the spine*
  - Groups (treatment groups)
    - Examples:
      - *control arm*
      - *paroxetine group*
      - *the 760 patients who received the telephone intervention*
  - Outcomes
    - Examples:
      - *death*
      - *lower limb injuries*
      - *the number of days of infectious symptoms*

Approach

- View this task as a *named entity recognition* problem

- <u>Named entity recognition:</u>
  - Locate sections of a text that correspond to a specific type of information
    - Names:
      - people, organizations, or locations
    - Type of value:
      - email addresses, date/times, or monetary values

Challenges

- Little research focused on finding the types of entities we seek
  - Most biomed NER work concerned with genes/proteins

- Entities are sometimes *long and complex descriptive phrases* (i.e. not just names)

  conventional coronary artery bypass grafting surgery using cardiopulmonary bypass

- Entities may be referred to *indirectly*

  half had [additional advice on anxiety management]$_{treatment}$ and half $\underbrace{\text{[did not]}}_{?}$$_{treatment}$

Data

- No available corpus containing annotated treatments, groups, and outcomes
  - Had to create our own

- Our corpus
  - 100 BMJ abstracts
    - Randomized controlled trials electronically published in 2005 and 2006
    - 1344 sentences
    - All treatments, groups and outcomes were annotated
      - Annotations verified by a physician at UIC

	Occurrences	Unique
Treatments	1208	230
Groups	363	168
Outcomes	1131	494

Number of annotated and unique entities in the corpus

Data

Annotation example:

[Mortality]$_{outcome}$ was higher in [the [quinine]$_{treatment}$ group]$_{group}$ than in [the [artemether]$_{treatment}$ group]$_{group}$

	Treatment	Group	Outcome
Avg. length	3.1	3.0	3.6
Std. dev.	2.3	1.5	2.2
Min. length	1	1	1
Max. length	15	12	14

Entity length statistics (in terms of number of tokens)

Methods

- To find entities in a sentence
  - Generate features for each word in a sentence
  - Use a Conditional random field (CRF) classifier to label each word as an entity word or not
  - Group together consecutive words with entity label
    - These word groupings are <u>*detected*</u> entities

- Performance evaluation
  - Match detected entities to annotated entities using various matching criteria (Exact, Left, Right, Left/Right, Partial)
  - Within an abstract, group detected entities together that refer to the *same* unique entity, and match this detected unique entity to one of the <u>true</u> (annotated) unique entities in the abstract.
    - Detected entities grouped using Exact or Partial match
    - Detected unique entities matched using Exact match

CRF classifier

- Designed for segmenting and labeling sequential data
  - e.g. labeling words in a sentence
- For CRF classifier we used MALLET SimpleTagger (McCallum, 2002)
- Features used:
  - word itself
  - part of speech (POS) tag
  - Medical Subject Heading (MeSH) Id
  - Semantic tag:
    - anatomy, time, disease, symptom, drug, procedure and measurement
  - The four words to the right and left of the current word along with their POS and semantic tags

CRF classifier

- For comparison, also used the Stanford CRF-based Named Entity Recognizer (Finkel et al., 2005)

- Features used:
  - word itself
  - character n-grams of lengths 2 to 6 within the word
  - identities of the four words on either side of the word
  - various word shape features such as "contains digit" or "all uppercase".

Results - Token/Occurrence recognition

	Token			Exact			Left			Right			Left/Right			Partial		
	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F
Treatments:																		
All features	.51	.75	.60	.40	.60	.48	.47	.70	.56	.49	.74	.59	.55	.82	.66	.56	.84	.67
Stanford NER	.43	.74	.54	.39	.63	.48	.45	.72	.55	.44	.71	.54	.49	.79	.61	.50	.81	.62
Groups:																		
All features	.70	.93	.80	.67	.91	.77	.69	.94	.79	.67	.92	.78	.69	.94	.80	.69	.94	.80
Stanford NER	.71	.95	.81	.67	.89	.76	.69	.92	.79	.68	.90	.78	.70	.93	.80	.70	.93	.80
Outcomes:																		
All features	.62	.75	.68	.45	.58	.51	.54	.69	.60	.57	.73	.64	.63	.80	.71	.64	.82	.72
Stanford NER	.52	.71	.60	.42	.59	.49	.48	.66	.55	.50	.70	.58	.54	.75	.63	.56	.78	.65

Results - Unique entity recognition

	Exact			Partial		
	R	P	F	R	P	F
Treatments:						
All features	.78	.34	.47	.73	.54	.62
Stanford NER	.67	.34	.46	.62	.51	.56
Groups:						
All features	.80	.83	.81	.77	.87	.82
Stanford NER	.77	.82	.80	.74	.88	.81
Outcomes:						
All features	.69	.43	.53	.61	.61	.61
Stanford NER	.61	.42	.50	.55	.60	.57

Results - feature analysis

- Usefulness of each type of feature depends on entity type

- Most useful features:
  - Treatments and Outcomes:
    - context features (features of neighboring tokens)
    - POS
    - section label
  - Groups:
    - word itself
    - context features

- Least useful features:
  - Treatments and Outcomes:
    - MeSH Id and semantic tags
  - Groups:
    - POS and section labels

Results - error analysis

- Missing an entity is most common error
  - More common than finding invalid entities
- Groups:
  - Groups of the form “<Treatment> group” are rarely missed
  - Difficulty detecting mentions that do not contain “group” or “arm”
    - e.g. participants referred to the clinic
- Treatments and outcomes:
  - Recognizing the wrong type of entity (i.e. labeling a treatment as an outcome or a disease as a treatment)

Conclusions

- Recognizing treatments, groups, and outcomes is challenging, but important
- Features that contribute outside information about a word (POS tags, abstract section title, MeSH Id, semantic tag) appear to be slightly more useful than "word shape" features (character n-grams, binary word shape features)
- Domain information about a word (i.e. semantic tags and MeSH Id) are not as helpful as might be expected

Future work

- Increase size of data set
  - Expand corpus to include full paper, not just abstracts
- Add features based on syntactic relations
- Handle difficult cases such as
  - half had treatment X and half did not