Detection of Evidence in Clinical Research Papers

Patrick Davis-Desmond Diego Mollá

Department of Computing,
Macquarie University

HIKM, 1 Feb 2012

Contents

Contents

Contents

Evidence Based Medicine

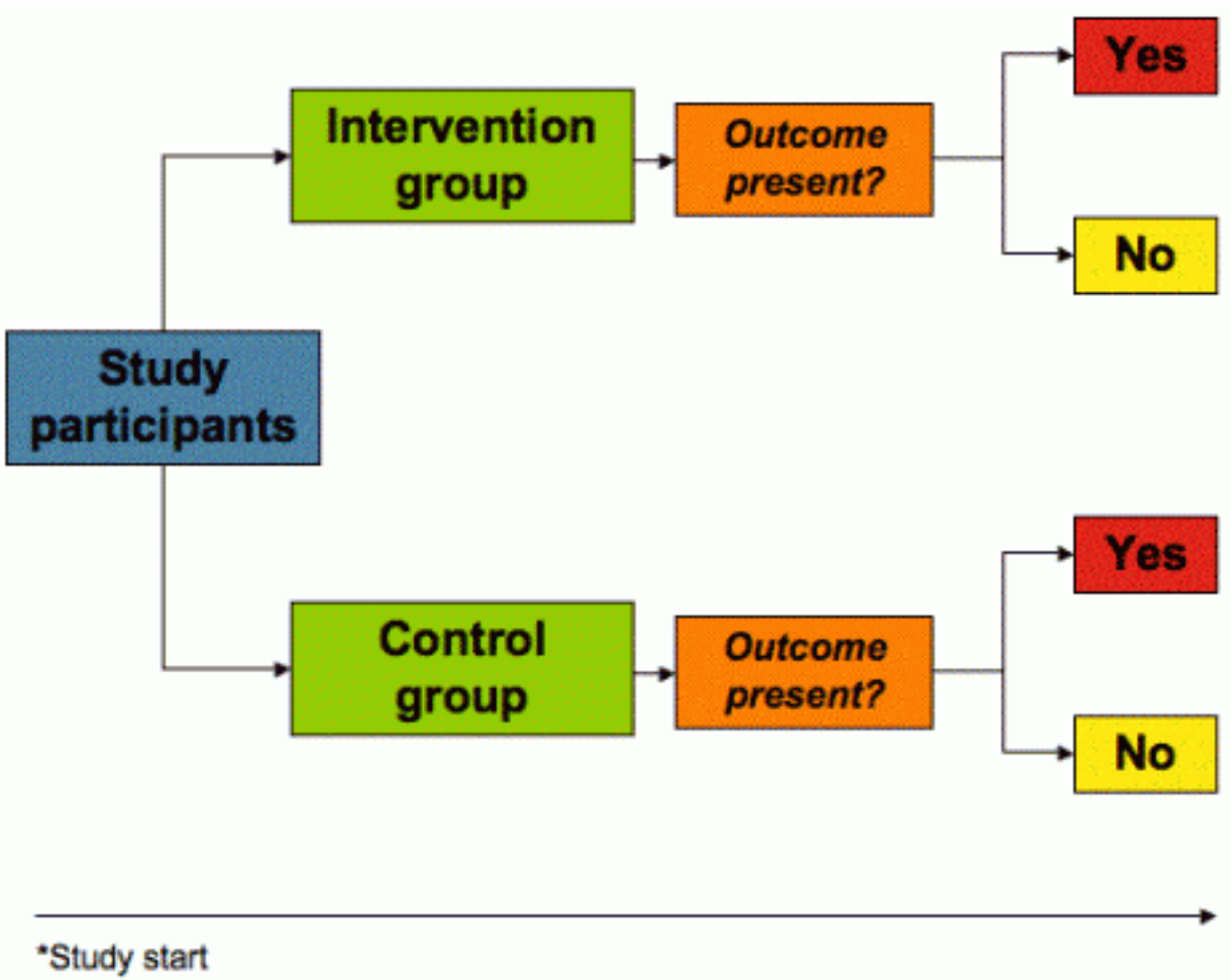
http://laikaspoetnik.wordpress.com/2009/04/04/evidence-based-medicine-the-facebook-of-medicine/

Levels of Evidence

Levels of evidence defined in the Strength Of Recommendation Taxonomy (SORT)

Study quality	**Diagnosis**	**Treatment / prevention / screening**	**Prognosis**
Level 1: good-quality patient-oriented evidence	Validated clinical decision rule; SR/meta-analysis of high-quality studies; high-quality diagnostic cohort study	SR/meta-analysis of RCTs with consistent findings; high-quality individual RCT; all-or-none study	SR/meta-analysis of good-quality cohort studies; prospective cohort study with good follow-up
Level 2: limited-quality patient-oriented evidence	Unvalidated clinical decision rule; SR/meta-analysis of lower-quality studies or studies with inconsistent findings; lower-quality diagnostic cohort study or diagnostic case-control study	SR/meta-analysis of lower-quality clinical trials or of studies with inconsistent findings; lower-quality clinical trial; cohort study; case-control study	SR/meta-analysis of lower-quality cohort studies or with inconsistent results; retrospective cohort study or prospective cohort study with poor follow-up; case-control study; case series
Level 3: other evidence	Consensus guidelines, extrapolations from bench research, usual practice, opinion, disease-oriented evidence (intermediate or physiologic outcomes only), or case series for studies of diagnosis, treatment, prevention, or screening		

Clinical Evidence in Randomised Controlled Trials



http://ebp.lib.uic.edu/dentistry/?q=node/48

Contents

NegEx

NegEx

- Aims to detect negated findings and diseases in discharge summaries
- List of expressions indicating negation
- Additional list of expressions indicating pseudo-negation (e.g. double negations)
- Negation is limited to a context window of five words either side of the target concept

Contents

Basic Approach

Key Idea

We frame the approach of detecting (lack of) evidence as one of detecting negation

Method

We modify and simplify NegEx

Contents

Corpus Gathering

Issues

- PubMed identifies RCTs but it does not provide full text
- PubMed Central provides full text in XML

Corpus Gathering

Issues

- PubMed identifies RCTs but it does not provide full text
- PubMed Central provides full text in XML

Process

1. Identify RCTs in PubMed
2. Select those RCTs from PubMed that appear in PubMed Central

Process for Corpus Gathering

Process, more detailed

1. Visit PubMed
2. Look at recent Randomised Control Trials (RCT)
3. Identify those that are completed (visual inspection)
4. Identify those that have a PMCID
5. Extract the PICO details (manually)
6. Save the full XML source from PubMed Central

Corpus Annotation

Annotation

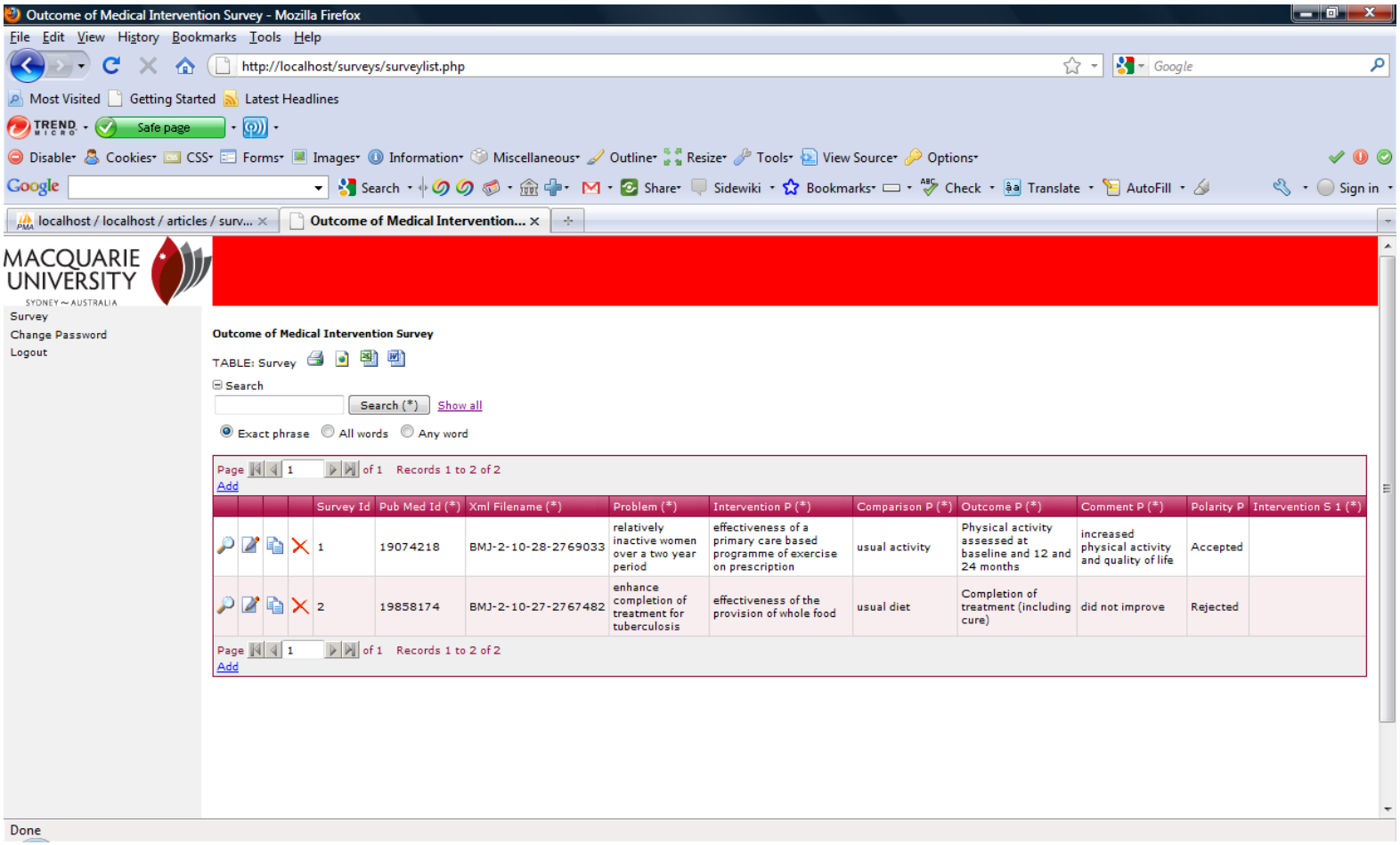
- Three annotators
- Web-based annotation tool

Instructions to annotators

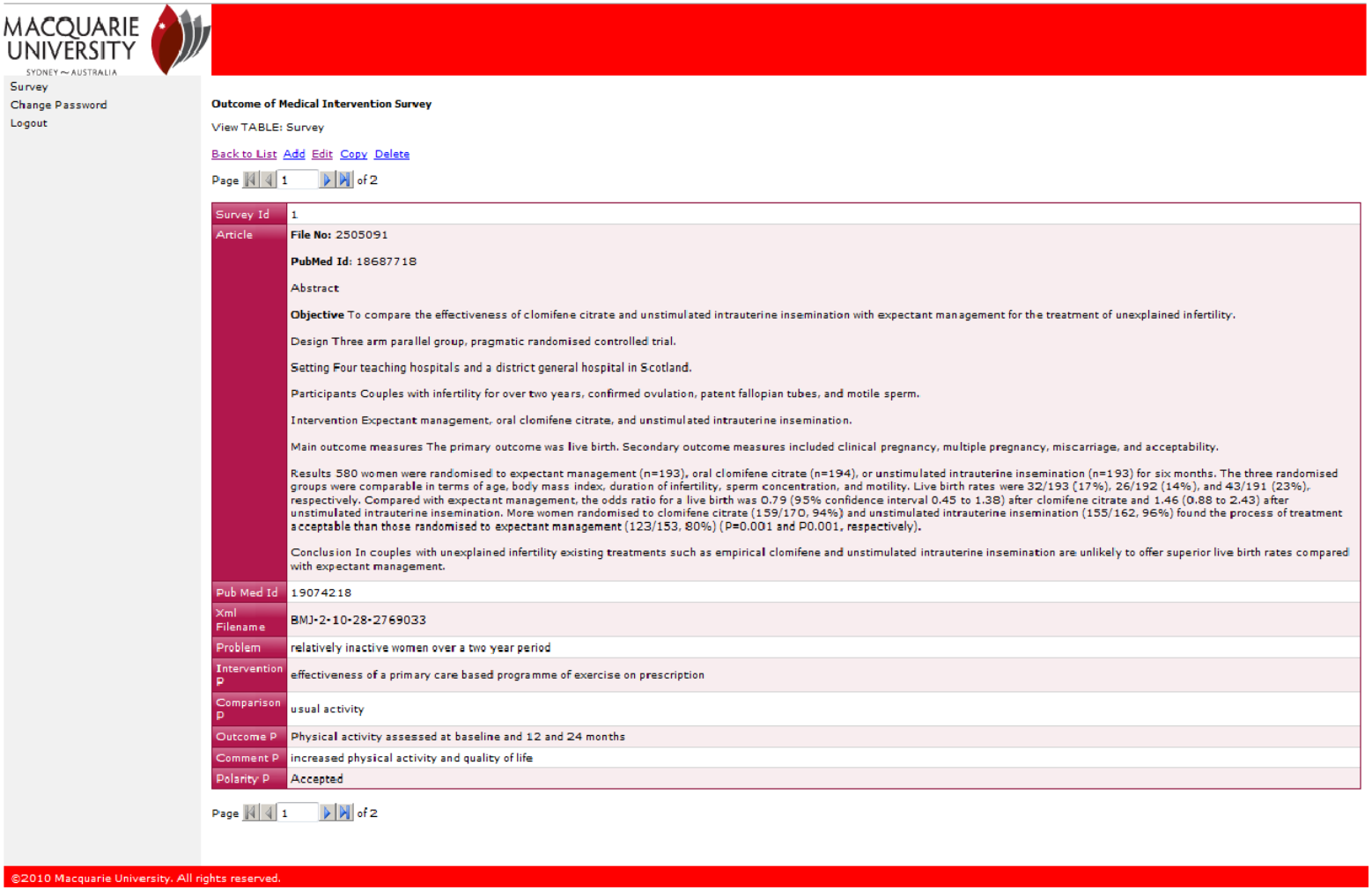
Read the abstract and assign one of these options:

Accepted	A difference is reported between the intervention and the control group
Rejected	No difference is reported
Unknown	Unable to tell (e.g. no results are provided)

Summary Listing Page



View Annotations Details Page



MACQUARIE UNIVERSITY
SYDNEY ~ AUSTRALIA

Survey
Change Password
Logout

Outcome of Medical Intervention Survey

View TABLE: Survey

Back to List Add Edit Copy Delete

Page 1 of 2

Survey Id	1
Article	**File No:** 2505091 **PubMed Id:** 18687718 Abstract **Objective** To compare the effectiveness of clomifene citrate and unstimulated intrauterine insemination with expectant management for the treatment of unexplained infertility. Design Three arm parallel group, pragmatic randomised controlled trial. Setting Four teaching hospitals and a district general hospital in Scotland. Participants Couples with infertility for over two years, confirmed ovulation, patent fallopian tubes, and motile sperm. Intervention Expectant management, oral clomifene citrate, and unstimulated intrauterine insemination. Main outcome measures The primary outcome was live birth. Secondary outcome measures included clinical pregnancy, multiple pregnancy, miscarriage, and acceptability. Results 580 women were randomised to expectant management (n=193), oral clomifene citrate (n=194), or unstimulated intrauterine insemination (n=193) for six months. The three randomised groups were comparable in terms of age, body mass index, duration of infertility, sperm concentration, and motility. Live birth rates were 32/193 (17%), 26/192 (14%), and 43/191 (23%), respectively. Compared with expectant management, the odds ratio for a live birth was 0.79 (95% confidence interval 0.45 to 1.38) after clomifene citrate and 1.46 (0.88 to 2.43) after unstimulated intrauterine insemination. More women randomised to clomifene citrate (159/170, 94%) and unstimulated intrauterine insemination (155/162, 96%) found the process of treatment acceptable than those randomised to expectant management (123/153, 80%) (P=0.001 and P0.001, respectively). Conclusion In couples with unexplained infertility existing treatments such as empirical clomifene and unstimulated intrauterine insemination are unlikely to offer superior live birth rates compared with expectant management.
Pub Med Id	19074218
Xml Filename	BMJ-2-10-28-2769033
Problem	relatively inactive women over a two year period
Intervention P	effectiveness of a primary care based programme of exercise on prescription
Comparison P	usual activity
Outcome P	Physical activity assessed at baseline and 12 and 24 months
Comment P	increased physical activity and quality of life
Polarity P	Accepted

Page 1 of 2

©2010 Macquarie University. All rights reserved.

Annotation Consistency

Agreement policy

- Whenever there was disagreement, the annotators were asked to review the abstract
- The annotators were not influenced to select any class or to change their decisions

Final Agreement

$\kappa = 70.6\%$ "good agreement beyond chance"

Contents

Statistical Classification

Corpus Splitting

	Accepted	Rejected	Total
(1) Training	66	61	127
(2) Test	33	34	67
(1)+(2) Total	99	95	194

Baselines

Statistical Classifiers

1. Decision Trees (J48)
2. Support Vector Machine (SVM)
3. Naïve Bayes (NB)

Features

1. All words in the abstract
2. All words in the conclusion section
3. Selected words in the abstract
4. Selected words in the conclusion section

Selected Words

The Selected Words

achieved, decrease, decreased, difference, effect, effective, effects, efficacy, improve, improvement, increase, increased, no, not, provide, provided, reduce, reduced, significant

Contents

Simplifications of NegEx

Simplifications

1. Different set of negation triggers
2. Two classes: "Accepted", "Rejected"
3. Detection of concepts was disabled
4. Detection of conjunctions and pseudonegation was disabled
5. Modified input-output processing (see paper)
6. Other minor changes (see paper)

Other Particularities

- Negation phrases are mostly bigrams and a few trigrams
- The algorithm only processed the conclusion section
  - All abstracts were structured

List of Negation Phrases

been overestimated, cannot endorse, cannot recommend, did not reduce, does not reduce, effectiveness overestimated, failed to, ineffective in, low probability, neither altered, no advantage, no advantageous, no beneficial, no benefit, no certain, no conclusive, no convincing, no definite, no detectable, no difference, no effect, no evidence, no favourable, no findings, no important, no improved, no increase, no irrefutable, no major, no meaningful, no more, no new, no novel, no overall benefit, no overall benefits, no overall effect, no positive, no proof, no reduction, no significant, no statistically, no strong, no substantial, no suggestion, nonsignificant improvement, non-significant improvement, nonsignificant reduction, non-significant reduction, nor protected, not affect, not appear to, not appreciate, not associated, not be, not beneficial, not change, not clinically, not confirm, not confirmed, not demonstrate, not differ, not exhibit, not find, not had, not have, not improve, not increase, not influence, not know, not known, not lead, not lend support, not likely, not meaningful, not meaningfully, not met, not necessarily, not observed, not offer, not prevent, not produce, not promote, not prove, not provide, not result, not reveal, not see, not show, not shown, not significant, not significantly, not slow, not statistically, not superior, not suppress, not to, not,, remains unproved, similarly effective, unlikely to

Contents

Results

Accuracy with 95% confidence intervals

	J48	SVM	NB
Baseline 1	49% (37%–61%)	66% (54%–76%)	69% (57%–79%)
Baseline 2	82% (71%–89%)	78% (67%–86%)	71% (59%–80%)
Baseline 3	54% (42%–65%)	63% (51%–73%)	58% (46%–69%)
Baseline 4	**84% (73%–91%)**	80% (69%–88%)	78% (67%-86%)
Rule-based		**95% (88%–98%)**	

Errors Explained

The main source of errors is the incorrect scope of the negation

- Secundary outcomes in the conclusions section
- The conclusions section did not include information about quality of evidence

Conclusions

Conclusions

1. An adaptation of NegEx produces very good results
2. ML methods not as good, though they may improve with more data
3. Focusing on the conclusions section improves the results
4. May need to detect the scope of negation

Further Work

Further Work

1. Test ML on larger data
2. Test other clinical study types, e.g. systematic reviews
3. Apply automated text structuring techniques to detect conclusion sentences
4. Detect secundary outcomes
5. Integrate into an evidence grading system

That's All

Questions?

`http://sourceforge.net/p/clinevidence/`