Semi-Automating Biomedical Evidence Synthesis via Machine Learning and NLP

Byron C Wallace

byron@ccs.neu.edu | byronwallace.com | @byron_c_wallace
This talk
Evidence-Based Medicine *n.*
The conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients
Evidence-Based Medicine $n$. What you kind of would have hoped all medicine is
… only 20 percent of medical practices are based on rigorous research evidence … The rest are based on a kind of folklore.
Patients from a target population

Randomize

Treatment A

Treatment B
The trouble with individual trials
Meta-analysis

Studies | Estimate (95% C.I.)
--- | ---
Overall | 0.358 (0.152, 0.565)
- Carroll, 1997 | 0.411 (0.202, 0.621)
- Grant, 1981 | 0.370 (0.126, 0.615)
- Peck, 1987 | 0.353 (0.113, 0.593)
- Donat, 2003 | 0.254 (0.093, 0.416)
- Stewart, 1990 | 0.337 (0.094, 0.580)
- Young, 1995 | 0.397 (0.169, 0.626)
The problem (or, how CS can help)

Trial results are, insanely, disseminated via unstructured articles
The median length to complete a single review: 1110 person-hours.

Allen & Olkin, JAMA, 1998
On average, 75 articles describing results from clinical trials are published every day.

Estimated time to complete and publish a systematic review: 67.3 weeks

Borah et al, BMJ, 2017
1 formulate question, protocol & query

2 search database

4 extract data

3 screen retrieved citations

5 synthesize extracted data
Wallace et al. KDD 2010; SDM 2011; KDD 2013; BMC Bioinf. 2010; Genetics in Medicine 2012 …
The automation of systematic reviews
Would lead to best currently available evidence at the push of a button
1. Formulate question, protocol & query

2. Search database

3. Screen retrieved citations

4. Extract data

5. Synthesize extracted data
Automating Biomedical Evidence Synthesis: RobotReviewer

Iain J. Marshall,1 Joël Kuiper,2 Edward Banner3 and Byron C. ... Toward this end, this paper describes RobotReviewer (RR; Figure 1), an open-source system that automates aspects of producing SRs from thousands of articles, including deep systematic reviews (EBM) and meta-analysis (Moss et al., 1999; Sackett et al., 1997). Despite advances in technology, the process of synthesising, structuring, and disseminating healthcare evidence remains slow and laborious (Olkin, 2010). Several free-text evidence synthesis tools exist, such as RoboSummarizer (Bastian et al., 2010) and Synthesizer (Koenen, 2010). But none automate this process, and they require human inspection to properly extract pertinent information. Researchers need methods to produce SRs efficiently and at scale, to help other researchers identify new evidence and areas for future research (Sack et al., 2010).

This paper describes RobotReviewer (RR; Figure 1), an open-source system that automates aspects of producing SRs from thousands of articles, including deep systematic reviews (EBM) and meta-analysis (Moss et al., 1999; Sackett et al., 1997). Despite advances in technology, the process of synthesising, structuring, and disseminating healthcare evidence remains slow and laborious (Olkin, 2010). Several free-text evidence synthesis tools exist, such as RoboSummarizer (Bastian et al., 2010) and Synthesizer (Koenen, 2010). But none automate this process, and they require human inspection to properly extract pertinent information. Researchers need methods to produce SRs efficiently and at scale, to help other researchers identify new evidence and areas for future research (Sack et al., 2010).

Thus, this paper describes RobotReviewer (RR; Figure 1), an open-source system that automates aspects of producing SRs from thousands of articles, including deep systematic reviews (EBM) and meta-analysis (Moss et al., 1999; Sackett et al., 1997). Despite advances in technology, the process of synthesising, structuring, and disseminating healthcare evidence remains slow and laborious (Olkin, 2010). Several free-text evidence synthesis tools exist, such as RoboSummarizer (Bastian et al., 2010) and Synthesizer (Koenen, 2010). But none automate this process, and they require human inspection to properly extract pertinent information. Researchers need methods to produce SRs efficiently and at scale, to help other researchers identify new evidence and areas for future research (Sack et al., 2010).

http://www.robotreviewer.net/
RobotReviewer automatically extracts and synthesises data from Randomized Controlled Trials. Drag and drop PDFs here, or click to select files to upload!
PICO extraction

Automating *risk of bias* assessment
PICO

Participants / Population

Intervention + Comparator

Outcomes
Antibiotic Treatment of Exacerbations of COPD*

A Randomized, Controlled Trial Comparing Procalcitonin-Guidance With Standard Therapy

Daiana Stolz, MD; Mirjam Christ-Crain, MD; Roland Bingisser, MD; Jörg Lettner, MD; David Miedinger, MD; Christian Müller, MD; Peter Huber, PhD; Beat Müller, MD; and Michael Tamir, MD

Background: Therapy with antibiotics influences recovery only in selected cases of COPD exacerbations. We evaluated the efficacy and safety of procalcitonin guidance compared to standard therapy with antibiotic prescriptions in patients experiencing exacerbations of COPD. Methods: A total of 208 consecutive patients requiring hospitalization for COPD exacerbation were randomized at the index exacerbation to procalcitonin-guided or standard antibiotic therapy. Patients receiving procalcitonin-guided therapy were treated with antibiotics according to serum procalcitonin levels; standard-therapy patients received antibiotics according to the attending physician. The primary outcome was the antibiotic exposure at the index exacerbation and the subsequent antibiotic requirement for COPD exacerbation within 6 months. Secondary outcomes were clinical recovery, symptom scores, length of hospitalization, ICU stay, death, lung function, exacerbation rate, and time to next exacerbation.

Results: At the index exacerbation, procalcitonin guidance reduced antibiotic prescription (40% vs 72%, respectively; p < 0.0001) and antibiotic exposure (relative risk [RR], 0.56; 95% confidence interval [CI], 0.43 to 0.73; p < 0.0001) compared to standard therapy. Moreover, procalcitonin guidance at the index exacerbation allowed a significant sustained reduction in total antibiotic exposure for up to 6 months (RR, 0.76; 95% CI, 0.64 to 0.92; p = 0.004). Clinical course and improvement in FEV1 at 14 days and 6 months did not differ between groups. Within 6 months, the exacerbation rate (0.62 vs 0.64, respectively), the rehospitalization rate (0.21 vs 0.24, respectively), and mean (± SD) time to the next exacerbation (70.0 ± 46.1 vs 70.4 ± 51.9 days, respectively; p = 0.523) were similar in both groups.

Conclusions: Procalcitonin guidance for exacerbations of COPD offers a sustained advantage.
Patients (n = 24, 15 females) with neck pain of > 3 months’ duration, who had pain in one or more cervical (C3-C7) zygapophysial joints …

The significant rate of response to the control treatment, even among patients who had been tested with placebo-controlled diagnostic blocks to confirm their perceptions of pain …

A journal paper
Author 1, Author 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit…

… study patients were selected from among patients whose cervical zygapophyseal-joint …

Curabitur convallis elementum lorem, sed vestibulum nunc …
Supervised/Distant Supervision
Wallace et al., Journal of Machine Learning Research, 2016

- structured data
- learned mapping
- f(x)
- distantly labeled data
- small amount of manually labeled unstructured data
- lots of unstructured data
AIMS: We examined whether Dementia Care Mapping (DCM) or the VIPS practice model (VPM) is more effective than education of the nursing home staff about dementia (control group) in reducing agitation and other neuropsychiatric symptoms as well as in enhancing the quality of life among nursing home patients with dementia.

http://www.ccs.neu.edu/home/benmye/EBM-NLP/
Current SOTA for Tagging

CRF

LSTM

word embeddings

char embeddings

enrolled men diagnosed with diabetes

enrolled men diagnosed with diabetes
How can we exploit the huge set of unlabeled biomedical abstracts?
Exploiting structured abstracts

- We capitalize on 50,000 segments from structured abstracts that explicitly contain PICO subsections
Minimally supervised pattern learning: AutoSlog-TS

Population texts

Interventions/Comparators & Outcomes texts

<NP_PREP>: patients with
<Act VP_PREP>: admitted with
...

High precision patterns
Exploiting structured abstracts

- We capitalize on 50,000 segments from structured abstracts that explicitly contain PICO subsections

- Example patterns:
  - P  women who, years of, diagnosed with
  - I/C patients received, performed after
  - O  scale of, patients reported, rate of
Exploiting structured abstracts

• We capitalize on 50,000 segments from structured abstracts that explicitly contain PICO subsections

• Example patterns:
  
  \( P \)  women who, years of, diagnosed with
  \( I/C \)  patients received, performed after
  \( O \)  scale of, patients reported, rate of

• How to exploit these as features in an LSTM-CRF?
Including pattern embeddings in an LSTM-CRF

- One way: add indicator of pattern type for current and adjacent tokens – concatenate one-hot and feed to LSTM layer
enrolled men diagnosed with diabetes

CRF

LSTM

word embeddings

char embeddings

enrolled men diagnosed with diabetes
enrolled men diagnosed with diabetes ...

CRF

LSTM

word embeddings

char embeddings

pattern indicators

enrolled men diagnosed with diabetes ...

⊕ ⊕ ⊕ ⊕ ⊕ ⊕
Results: LSTM-CRF

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>61.87</td>
<td>41.45</td>
</tr>
<tr>
<td>+patterns</td>
<td>75.00</td>
<td>64.80</td>
</tr>
<tr>
<td>baseline</td>
<td>38.65</td>
<td>23.80</td>
</tr>
<tr>
<td>+patterns</td>
<td>58.25</td>
<td>43.39</td>
</tr>
<tr>
<td>baseline</td>
<td>39.90</td>
<td>25.47</td>
</tr>
<tr>
<td>+patterns</td>
<td>39.66</td>
<td>24.20</td>
</tr>
</tbody>
</table>
Including pattern embeddings in an LSTM-CRF

- One way: add indicator of pattern type for current and adjacent tokens – concatenate one-hot and feed to LSTM layer
- Another way: directly embed the patterns!
enrolled men \{diagnosed with\} diabetes ...

CRF

LSTM

word embeddings

char embeddings

enrolled men \{diagnosed with\} diabetes ...

\oplus \oplus \oplus \oplus
Learned pattern embeddings
Learned pattern embeddings

Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron C. Wallace. NAACL, 2018. Syntactic Patterns Improve Information Extraction for Medical Search
Learning PICO representations of abstracts

Sarthak Jain, Edward Banner, Jan-Willem van de Meent, Iain Marshall and Byron C. Wallace.
Learning Disentangled Representations of Texts with Application to Biomedical Abstracts.
Automating risk of bias assessment

Aspirin has been shown to alleviate symptoms ...

A statistician generated a random sequence ...

This study concerns the use of ...

Sentence predictions & rankings

MT SVM sentence module

RA-CNN sentence module

Final document prediction

Aggregate Sentence Rankings via Borda count

Document embedding

sparse feature vector

sentence module

document prediction

sentence module

document prediction

sentence module

document prediction
Patients from a target population

Randomize

Treatment A

Treatment B
Risks of Bias (RoB)

- A key step in evidence appraisal: assessing the reliability of individual trials
Risks of Bias (RoB)

- A key step in evidence appraisal: assessing the reliability of individual trials

- Formalized in the Cochrane Risk of Bias (RoB) tool
  - Four core criteria: random sequence generation; allocation concealment; blinding of participants/personnel; blinding of outcomes assessment
Risks of Bias (RoB)

• A key step in evidence appraisal: assessing the reliability of individual trials

• Formalized in the Cochrane Risk of Bias (RoB) tool
  – Four core criteria: *random sequence generation*; *allocation concealment*; *blinding of participants/personnel*; *blinding of outcomes assessment*

• For each, assess overall risk as “low” or “high/unknown” and provide support for this judgement
Bias: Allocation concealment

Authors judgement: Low risk

Support for judgement: Quote: "The Family Practice Research Coordinator at the University of British Columbia held this sequence independently and remotely"
The machine learning task

Input: a full-text paper

Machine Learning

Output: RoB assessments and supporting quotes
Prior work

RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials

Iain J Marshall\textsuperscript{1}, Joël Kuiper\textsuperscript{2} and Byron C Wallace\textsuperscript{3}
RA-CNN: Extending CNNs to exploit rationales


[GitHub](https://github.com/bwallace/rationale-CNN)
The film, however, is all good.

Films adapted from comic books... 

Now onto from hell’s appearance: it’s...

The film, however, is all good.
sentence model
\[ p(y^{iC}_{\text{sen}} = k) \propto \exp(\mathbf{w}_{\text{sen}}^k \cdot \mathbf{x}^{i0}_{\text{sen}}) \]
\[ \cdots \]
\[ p(y^{iR}_{\text{sen}} = k) \]
\[ \cdots \]
\[ p(y^{iN}_{\text{sen}} = k) \]

sentence vectors
\[ \mathbf{x}^{i0}_{\text{sen}} \]
\[ \cdots \]
\[ \mathbf{x}^{i1}_{\text{sen}} \]
\[ \cdots \]
\[ \mathbf{x}^{iN}_{\text{sen}} \]

Films adapted from comic books...
\[ \cdots \]
Now onto from hell’s appearance: it’s...
\[ \cdots \]
The film, however, is all good.

Films adapted from comic books...
\[ \cdots \]
Now onto from hell appearance: it’s...
\[ \cdots \]
The film, however, is all good.
document vector

\[ x'_{\text{doc}} = \sum_{j=1}^{N_j} \max\{p(y_{\text{sen}}^j = \text{positive rationale}), p(y_{\text{sen}}^j = \text{negative rationale})\} \cdot x_{\text{sen}}^{ij} \]

sentence model

\[ p(y_{\text{sen}}^{iC} = k) \propto \exp(w_{\text{sen}}^i \cdot x_{\text{sen}}^{iu}) \]

sentence vectors

\[ x_{\text{sen}}^{i0}, \ldots, x_{\text{sen}}^{ij}, \ldots, x_{\text{sen}}^{iN_j} \]

Films adapted from comic books...

Now onto from hell's appearance: it's...

The film, however, is all good.

Films adapted from comic books...

Now onto from hell's appearance: it's...

The film, however, is all good.

neutral

positive rationale
document model

\[ p(y^i_{doc} = k) \propto \exp(W^k_{doc} \mathbf{x}^i_{doc}) \]

document vector

\[ x^i_{doc} = \sum_{j=1}^{N_i} \max\{p(y^j_{sen} = \text{positive rationale}), p(y^j_{sen} = \text{negative rationale})\} \cdot x^j_{sen} \]

sentence model

\[ p(y^i_{sen} = k) \propto \exp(W^k_{sen} \mathbf{x}^i_{sen}) \]

sentence vectors

\[ \mathbf{x}^{i0}_{sen}, \ldots, \mathbf{x}^{iN_i}_{sen} \]

 Films adapted from comic books...  
Now onto from hell’s appearance: it’s...  
The film, however, is all good.

\[ N_i \text{ sentences} \]
Results on RoB datasets

RA-CNN

our prior approach
Example rationales (RSG)

Study deemed at low risk of bias

*The study was performed double blind.*

Study deemed at high/unknown risk of bias

*The present study is retrospective, there is a risk that the woman did not properly recall how and what they experienced ....*
1. Whole-breast irradiation with or without a boost for patients treated with breast-conserving surgery for early breast cancer: 20-year follow-up of a randomised phase 3 trial

Evidence Updates 2015

Risk of bias: low

Robot Reviewer estimate of bias caused by problems in:

- randomisation: low
- allocation concealment: low
- blinding: high/unclear

For further information on the assessment of bias click here

https://www.tripdatabase.com/
Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool

Allison Gates, Ben Vandermeer, Lisa Hartling*

Alberta Research Centre for Health Evidence (ARCHE), Department of Pediatrics, University of Alberta, 4-472, Edmonton Clinic Health Academy, 11405-87 Avenue NW, Edmonton, Alberta T6G 1C9, Canada

Accepted 14 December 2017; Published online 28 December 2017

Abstract

Objectives: To evaluate the reliability of RobotReviewer’s risk of bias judgments.

Study Design and Setting: In this prospective cross-sectional evaluation, we used RobotReviewer to assess risk of bias among 1,130 trials. We compared reliability with human reviewers using Cohen’s kappa coefficient and calculated sensitivity and specificity. We investigated differences in reliability by risk of bias domain, topic, and outcome type using the chi-square test in meta-analysis.

Results: Reliability (95% CI) was moderate for random sequence generation (0.48 [0.43, 0.53]), allocation concealment (0.45 [0.40, 0.51]), and blinding of participants and personnel (0.42 [0.36, 0.47]); fair for overall risk of bias (0.34 [0.25, 0.44]); and slight for blinding of outcome assessors (0.10 [0.06, 0.14]), incomplete outcome data (0.14 [0.08, 0.19]), and selective reporting (0.02 [0.01, 0.03]). Reliability for blinding of participants and personnel (P < 0.001), blinding of outcome assessors (P = 0.005), selective reporting (P < 0.001), and overall risk of bias (P < 0.001) differed by topic. Sensitivity and specificity (95% CI) ranged from 0.20 (0.18, 0.23) to 0.76 (0.72, 0.80) and from 0.61 (0.56, 0.65) to 0.95 (0.93, 0.96), respectively.

Conclusion: Risk of bias appraisal is subjective. Compared with reliability between author groups, RobotReviewer’s reliability with human reviewers was similar for most domains and better for allocation concealment, blinding of participants and personnel, and overall risk of bias. © 2018 Elsevier Inc. All rights reserved.
ORIGINAL ARTICLE

Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool

Allison Gates, Ben Vandermeer, Lisa Hartling*

Alberta Research Centre for Health Evidne (ARCHE), Department of Pediatrics, University of Alberta, 4-472, Edmonton Clinic Health Academy, 11405-87 Avenue NW, Edmonton, Alberta T6G 1C9, Canada

Accepted 14 December 2017; Published online 28 December 2017

Abstract

Objectives: To evaluate the reliability of RobotReviewer’s risk of bias judgments.

Study Design and Setting: In this prospective cross-sectional evaluation, we used RobotReviewer to assess risk of bias among 1,130 trials. We compared reliability with human reviewers using Cohen’s kappa coefficient and calculated sensitivity and specificity. We investigated differences in reliability by risk of bias domain, topic, and outcome type using the chi-square test in meta-analysis.

Results: Reliability (95% CI) was moderate for random sequence generation (0.48 [0.43, 0.53]), allocation concealment (0.45 [0.40, 0.51]), and blinding of participants and personnel (0.42 [0.36, 0.47]); fair for overall risk of bias (0.34 [0.25, 0.44]); and slight for blinding of outcome assessors (0.10 [0.06, 0.14]), incomplete outcome data (0.14 [0.08, 0.19]), and selective reporting (0.02 [-0.02, 0.05]). Reliability of RobotReviewer’s risk of bias appraisal was similar for most domains and better for allocation concealment, blinding of participants and personnel, and overall risk of bias. © 2018 Elsevier Inc. All rights reserved.
A randomized user study of semi-automated risk of bias prediction (nearing completion)
Prospective trial

41 Participants, recruited via personal networks

Each performed RoB assessment for four RCT articles, all using our interface, but only 2/4 randomly selected to present ML predictions
Efficacy and Safety of Fluticasone Furoate/Vilanterol Compared With Fluticasone Propionate/Salmeterol Combination in Adult and Adolescent Patients With Persistent Asthma

A Randomized Trial

Ashley Woodcock, MD; Eugene R. Bleecker, MD, FCCP; Jan Lötvall, PhD; Paul M. O’Byrne, MD, FCCP; Eric D. Bateman, MD; Hilary Medley, PGDip; Anna Ellsworth, BS; Loretta Jacques, PhD; and William W. Busse, MD

Background: The combination of fluticasone furoate (FF), a novel inhaled corticosteroid (ICS), and vilanterol (VI), a long-acting β agonist, is under development as a once-daily treatment of asthma and COPD. The aim of this study was to compare the efficacy of FF/VI with fluticasone propionate (FP)/salmeterol (SAL) in patients with persistent asthma uncontrolled on a medium dose of ICS.

Methods: In a randomized, double-blind, double-dummy, parallel group study, 806 patients received FF/VI (100/25 μg, n = 403) once daily in the evening delivered through ELLIPTA (GlaxoSmithKline) dry powder inhaler, or FP/SAL (250/50 μg, n = 403) bid through DISKUS/ACCUMULAIR (GlaxoSmithKline). The primary efficacy measure was 0- to 24-h serial weighted mean (wmg) FEV, after 24 weeks of treatment.

Results: Improvements from baseline in 0- to 24-h wmgFEV, were observed with both FF/VI (341 mL) and FP/SAL (377 mL); the adjusted mean treatment difference was not statistically significant (~37 mL; 95% CI, 0.88 to 15, P = 0.162). There were no differences between 0- to 4-h serial wmgFEV, trough FEV,, and asthma control and quality-of-life questionnaire scores. There was no difference in reported exacerbations between treatments. Both treatments were well tolerated, with no clinically relevant effect on urinary cortisol excretion or vital signs and no treatment-related serious adverse events.

Conclusions: The efficacy of once-daily FF/VI was similar to bid FP/SAL in improving lung function in patients with persistent asthma. No safety issues were identified.

Trial registry: ClinicalTrials.gov; No.: NCT01147848; URL: www.clinicaltrials.gov

CHEST 2013; 144(4):1222–1229

Abbreviations: AE = adverse event; AQLQ+12 = Asthma Quality of Life +18 Questionnaire; EQ-5D = European Quality of Life-5 Dimensions; FF = fluticasone furoate; FP = fluticasone propionate; ICS = inhaled corticosteroid; ICSA = inhaled corticosteroid; ICSL = inhaled corticosteroid; SABA = short-acting β agonist; SAMA = short-acting antimuscarinic; SAL = salmeterol; FF = fluticasone; VI = vilanterol.
Punchline: ML (semi-automation) reduced workload by 25%
Effectiveness and cost of selective decontamination of the digestive tract in critically ill intubated patients: A randomized, double-blind, placebo-controlled, multicenter trial.

We evaluated the effect of selective decontamination of the digestive tract (SDD) on the incidence of ventilator-associated pneumonia (VAP) and its associated morbidity and cost in a mixed population of intubated patients. Two hundred seventy-one consecutive patients admitted to the intensive care units (ICUs) of five teaching hospitals and who had an expected need for intubation exceeding 48 h were enrolled and received topical antibiotics or placebo. Uninfected patients additionally received ceftriaxone or placebo for 3 d. VAP occurred in 11.4% of SDD-treated and 29.3% of control-group patients (p < 0.001; 95% confidence interval [CI]: 7.8 to 27.9). The incidence of nonrespiratory infections in the two groups was 19.1% and 30.7%, respectively (p = 0.04; 95% CI: 0.7 to 22.7). Among survivors, the median length of ICU stay was 11 d (interquartile range: 7 to 21.5 d) for the SDD-treated group and 16.5 d (10 to 30 d) for the control group (p = 0.006). Mean cost per survivor was $11,926 for treated and $16,296 for control-group patients. Mortality was 38.9% and 47.1%, respectively (p = 0.57). In decontaminated patients, the prevalence of gram-negative bacilli fell within 7 d from 47.4% to 13.0% (p < 0.001), whereas colonization with resistant gram-positive strains was higher (p < 0.05) than in the placebo group. In a mixed population of intubated patients, SDD was associated with a significant reduction of morbidity at a reduced cost. Our findings support the use of SDD in this high-risk group.

With respect to overall mortality, reported difference between patients receiving topical plus systemic and those receiving no prophylaxis …

☐ Significantly increased
☐ Significantly decreased
☑ No significant difference
☐ Cannot tell based on the text
Abstract

Background

Whether epidural analgesia for labor prolongs the active-first and second labor stages and increases the risk of vacuum-assisted delivery is a controversial topic. Our study was conducted to answer the question: does lumbar epidural analgesia with lidocaine affect the progress of labor in our obstetric population?

Method

395 healthy, nulliparous women, at term, presented in spontaneous labor with a singleton vertex presentation. These patients were randomized to receive analgesia either, epidural with bolus doses of 1% lidocaine or intravenous, with meperidine 25 to 50 mg when their cervix was dilated to 4 centimeters. The duration of the active-first and second stages of labor and the neonatal Apgar scores were recorded, in each patient. The total number of vacuum-assisted and cesarean deliveries were also measured.

Results

197 women were randomized to the epidural group. 198 women were randomized to the single-dose intravenous meperidine group. There was no statistical difference in rates of vacuum-assisted delivery rate. Cesarean deliveries, as a consequence of fetal bradycardia or dystocia, did not differ significantly between the groups. Differences in the duration of the active first and the second stages of labor were not.

With respect to "Oxytocin augmentation", characterize the reported difference between patients receiving "Epidural analgesia" and those receiving "Control".

Please fill out the following information based on the statement above:

- Significantly increased ("Epidural analgesia" vs. "Control")
- Significantly decreased ("Epidural analgesia" vs. "Control")
- No significant difference ("Epidural analgesia" vs. "Control")
- Cannot tell based on the abstract

Support for Decision

Submit
Thanks. Questions / comments / complaints?

Informing patient care using all relevant evidence is hard, because evidence is unstructured.

Researchers in IR/NLP/ML can play a key role in fixing this!

byron@ccs.neu.edu | http://byronwallace.com | @byron_c_wallace