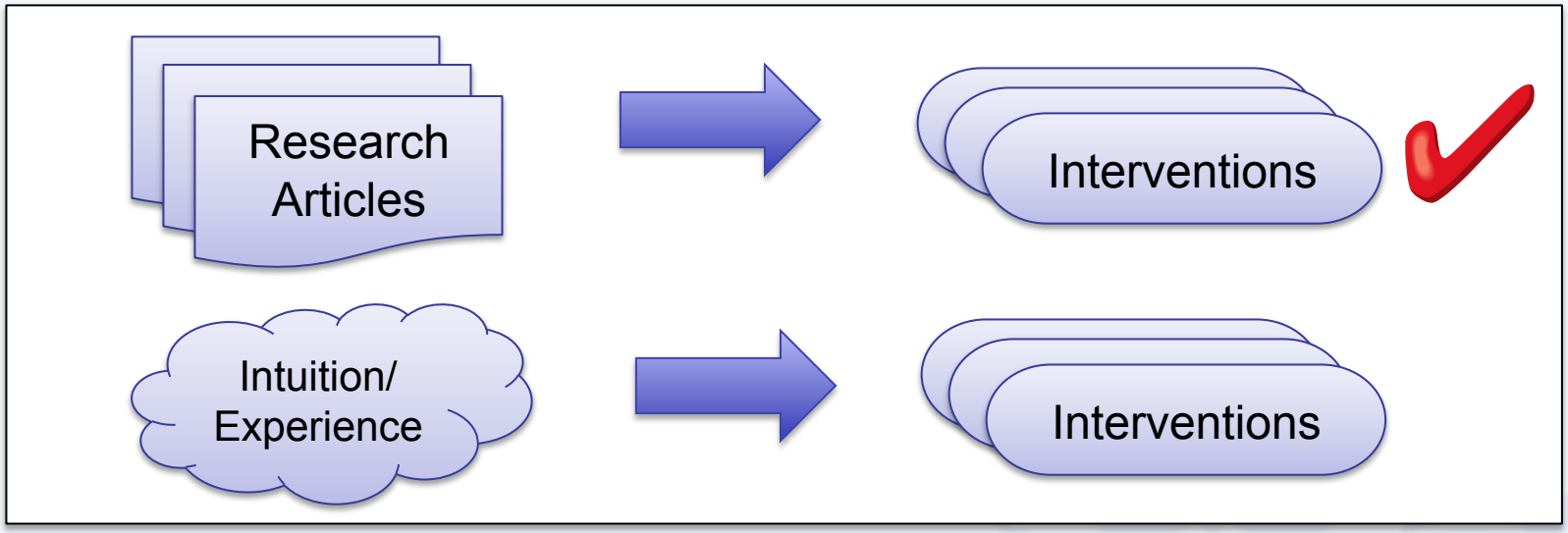
Improving Search for Evidence-based Practice using Information Extraction

**Jin Zhao, Min-Yen Kan, Paula M. Procter,
Siti Zubaidah, Wai Kin Yip, Goh Mien Li**

Evidence-based Practice (EBP)



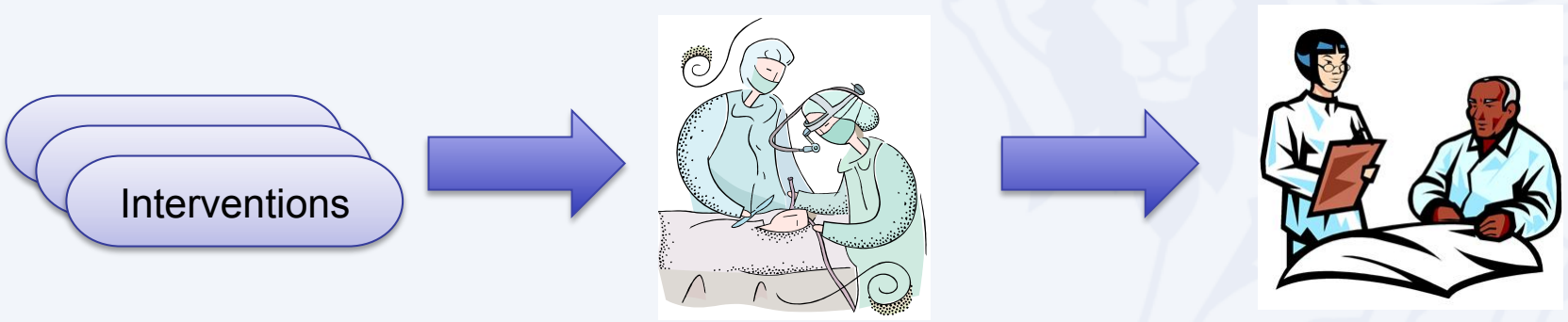
- **Advantages**
  - Reliable
  - Efficient
  - Updated

Two Stages of EBP

- **Stage 1: Search and Appraise**

Crucial yet difficult!

- **Stage 2: Apply and Evaluate**



Usability Gap between Search Engines and Healthcare Professionals

- **EBP-related information**
  - PICO elements
    - Patient
      - Sex, CO-morbidity, Race, Age, Pathology (SCORAP)
    - Intervention
    - Condition
    - Outcome
  - Strength of evidence
    - Systematic review of Randomized Control Trials (RCTs)
    - At least one RCT
    - Pseudo-RCT, Cohort studies, …

Usability Gap between Search Engines and Healthcare Professionals

- **EBP-related information missing in search results**

Effects of Intermittent Electrical Stimulation on Superficial Pressure, Tissue Oxygenation, and Discomfort Levels for the Prevention of Deep Tissue Injury.

Solis LR, Gyawali S, Seres P, Curtis CA, Chong SL, Thompson RB, Mushahwar VK.

Rehabilitation Science Program, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada.

Abstract

The overall goal of this project is to develop effective methods for the prevention of deep tissue injury (DTI). DTI is a severe type of pressure ulcer that originates at deep bone-muscle interfaces as a result of the prolonged compression of tissue. It afflicts individuals with reduced mobility and sensation, particularly those with spinal cord injury. We previously proposed using a novel electrical stimulation paradigm called intermittent electrical stimulation (IES) for the prophylactic prevention of DTI. IES-induced contractions mimic the natural repositioning performed by intact individuals, who subconsciously reposition themselves as a result of discomfort due to prolonged sitting. In this study, we investigated the effectiveness of various IES paradigms in reducing pressure around the ischial tuberosities, increasing tissue oxygenation throughout the gluteus muscles, and reducing sitting discomfort in able-bodied volunteers. The results were compared to the effects of voluntary muscle contractions and conventional pressure relief maneuvers (wheelchair push-ups). IES significantly reduced pressure around the tuberosities, produced significant and long-lasting elevations in tissue oxygenation, and significantly reduced discomfort produced by prolonged sitting. IES performed as well or better than both voluntary contractions and chair push-ups. The results suggest that IES may be an effective means for the prevention of DTI.

Patient details?

Interventions considered?

Outcomes?

Study Design?

Bridging the Gap…

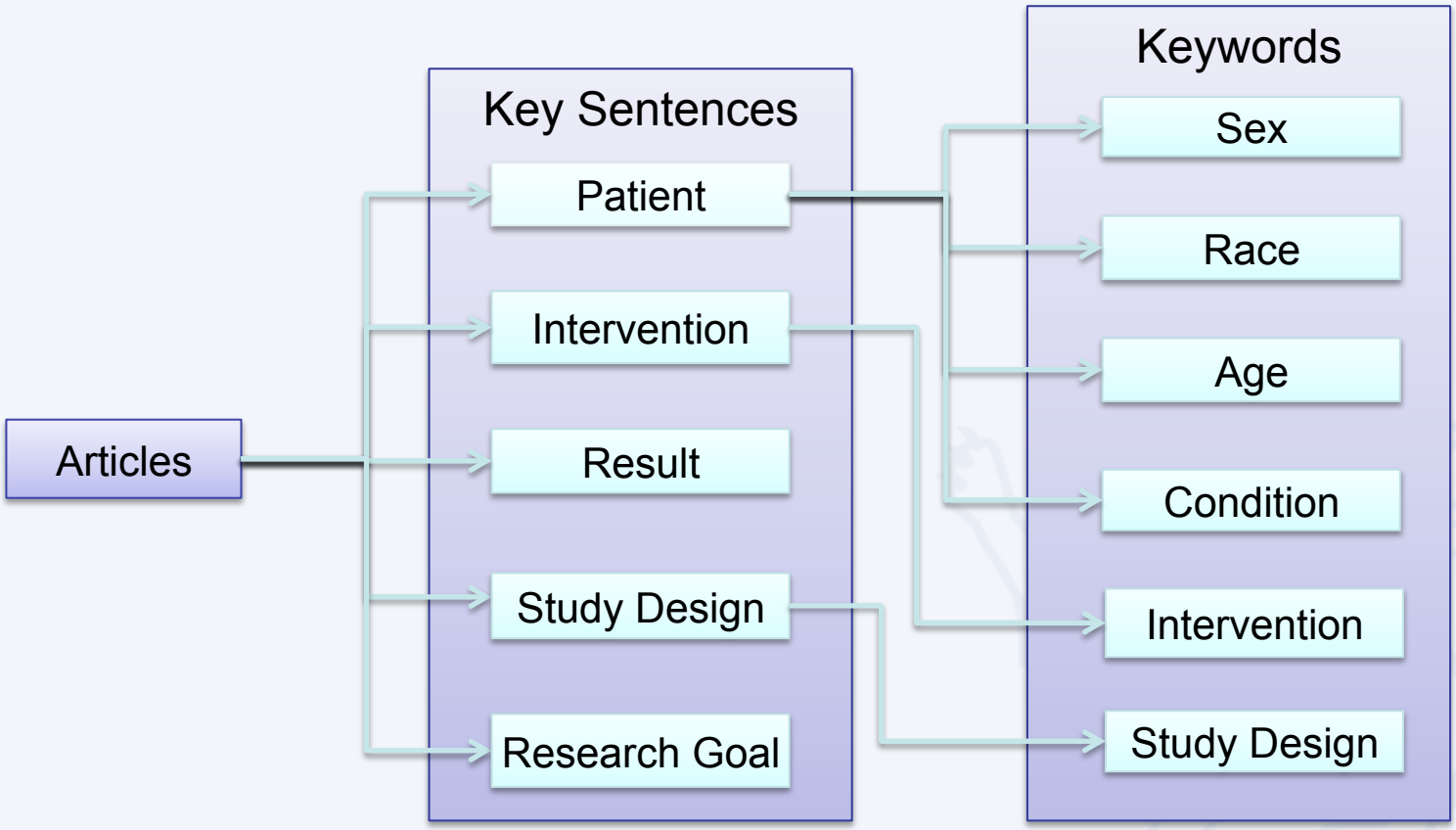
- **Automated extraction of EBP-related information**

Intervention, Patient, Research Goal, Study Design	We performed an **open**, **prospective**, **randomized** **clinical** **trial** in 51 patients receiving mechanical ventilation for more than 72 h, in order to evaluate the impact of using **noninvasive** (**quantitative** **endotracheal** **aspirates** [**QEA**]) diagnostic method on the morbidity and mortality of **ventilator-associated** **pneumonia** (**VAP**).

Sex
Condition
Race
Age
Intervention
Study Design

- **Integration into the search process**
  - Filtering and sorting function
  - Re-ranking based on key information

Two-stage Extraction Pipeline



Extraction via Supervised Classification

- **Binary Maximum Entropy Classifiers**
  - One for each class
  - Extracted if the corresponding classifier reports positive
  - Text features only

Feature Group	Definition
Token	The N-grams of the sentence
Sentence	The length and position of the sentence
Named Entity	Whether the sentence contains person, location, organization names
MeSH	Whether the sentence contains a MeSH term of a particular category
Lexica	Whether the sentence contains a word from the manually compiled wordlist for sex/age/race.

Features for Sentence Extraction

Feature Group	Definition
Token	The word itself, its stem and its part-of-speech
Phrase	The position of the word in the phrase and the head noun in the phrase
Named Entity	Whether the word is part of person, location, organization names
MeSH	Whether the word is part of a MeSH term of a particular category
Lexica	Whether the word is part of an entry in the manually compiled wordlist for sex/age/race.

Features for Keyword Extraction

Evaluation

- **Corpus development**
  - 19,893 medical abstracts and full text articles from 17 journal websites
  - 2,000 randomly selected sentence annotated

Patient	Intervention	Result	Study Design	Research Goal	Others
8%	4.8%	22.5%	10.2%	3.5%	66.7%

  - All (6,754) words in sentences belonging to Patient, Intervention and Study Design annotated

Sex	Age	Race	Condition	Intervention	Study Design	Others
0.8%	2.6%	0.2%	5.4%	5.5%	3.8%	82.4%

Evaluation

- **Metrics**
  - Precision, Recall and F_1-Measure
    - Precision (P) = TP / (TP+FP)
    - Recall (R) = TP / (TP+FN)
    - F_1-Measure (F) = 2 * P * R / (P + R)

    where TP: true positive, FP: false positive, FN: false negative.

Results for Sentence Extraction

Class	All			No Token			No Sentence			No Named Entity			No MeSH			No Lexica		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Patient	.70	.24	.36	**-.13**	**+.08**	**+.05**	**-.10**	-.04	**-.06**	+.02	-.01	-.01	0	+.02	+.02	**-.05**	-.01	-.02
Intervention	.90	.28	.43	**-.40**	**-.10**	**-.16**	**-.14**	**-.05**	**-.07**	-.04	-.03	-.04	**-.07**	**-.07**	**-.09**	-.04	-.03	-.04
Result	.78	.56	.65	**-.17**	**-.23**	**-.22**	-.02	+.02	+.01	-.01	0	0	-.01	+.02	+.01	-.02	+.01	0
Study Design	.89	.39	.54	**-.56**	**-.29**	**-.39**	**-.09**	-.04	**-.06**	-.01	**-.05**	**-.05**	-.01	+.03	+.03	0	+.01	+.01
Research Goal	.92	.27	.42	**-.47**	**-.08**	**-.15**	+.03	-.01	-.01	-.05	+.02	+.02	+.04	+.03	+.04	+.03	-.01	-.01

- **Precise but much room for improvement in recall**
- **Difficulty in handling structural variation and short information**
- **Important features: Token and Sentence**

Results for Keywords Extraction

Class	All			No Token			No Phrase			No Named Entity			No MeSH			No Lexica		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Sex	.98	1	.99	0	0	0	0	0	0	0	0	0	0	0	0	0	**-.12**	**-.06**
Condition	.74	.62	.67	+.03	-.03	-.01	-.03	**-.15**	**-.11**	0	**-.10**	**-.06**	-.01	**-.17**	**-.12**	-.02	-.01	-.01
Race	1	.86	.92	**-.08**	0	-.04	**-.08**	0	-.04	**-.18**	0	**-.09**	**-.08**	0	-.04	0	**-.79**	**-.79**
Age	.84	.75	.79	-.04	**-.07**	**-.06**	+.04	**-.07**	-.03	-.02	.00	-.01	0	0	0	-.01	-.04	-.03
Intervention	.73	.55	.63	0	-.01	-.01	**-.07**	**-.20**	**-.17**	0	+.04	+.03	+.04	-.08	-.04	-.01	+.02	+.01
Study Design	.91	.71	.80	**-.15**	**-.23**	**-.21**	0	**-.08**	**-.05**	-.03	.02	0	-.11	+.02	-.03	-.04	+.01	-.01

- **Better performance in general**
- **Difficulty in handling lexical variations and diverse vocabulary**
- **Positive contribution from all feature groups**

Future Work

- **Joint inference**
  - Allows interactions between two extraction steps
- **Bigger knowledge source**
  - Addresses diverse vocabulary problem
- **Integration of extracted information into search process**

Conclusion

- **Two-stage pipeline for key information extraction**
  - Extraction using binary Maximum Entropy classifiers and text features
- **To improve extraction performance and integrate extracted information into search process**