

Clinical Relevance of the Doctor's Dilemma Question Set

Daniel R. Schlegel Ph.D., Sashank Kaushik MD, and Peter L. Elkin MD

University at Buffalo Department of Biomedical Informatics

Introduction

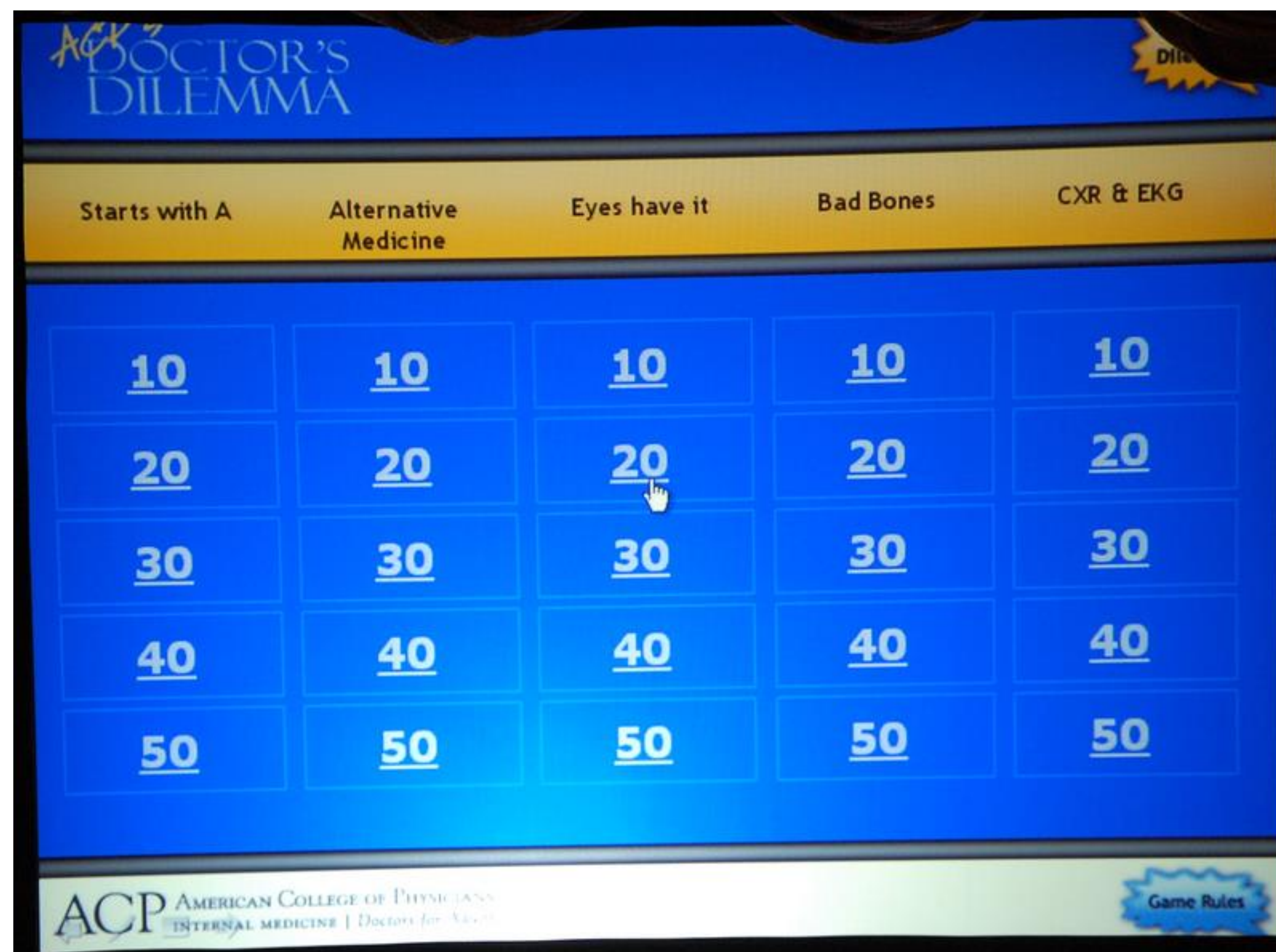
It is our eventual goal to create a computational system which answers all types of textual clinical questions. To do so, we need a set of questions to train on. Such questions must be topically diverse within medicine, and be clinically relevant. We have examined the Doctor's Dilemma [1] question set to determine if it meets those criteria.

Clinical Relevance

A question is *clinically relevant* if it's one which could occur to a doctor during clinical practice, regardless of whether it's something they know already or might have to look up. In the past, the IBM Watson group has made use of a subset of the Doctor's Dilemma questions in developing Medical Watson [2], but they only used the diagnosis-type questions, which are clearly clinically relevant.

Doctor's Dilemma

Residents compete in what is essentially "medical Jeopardy" every year at the annual ACP Internal Medicine Meeting. Also called "Medical Jeopardy", up to 50 teams compete each year over a 3-day period for the Osler Cup. Just like in televised Jeopardy, harder questions have higher point values. For practice, the game can also be played online on the ACP's website.



November 2008 ACP Doctor's Dilemma competition. Photo credit: Flickr user grhemb04, used with permission.

References

- [1] Doctor's dilemma competition — ACP. http://www.acponline.org/residents_fellows/competitions/doctors_dilemma/. Accessed: February 20, 2015.
- [2] D. A. Ferrucci, A. Levas, S. Bagchi, D. Gondek, and E. T. Mueller. Watson: Beyond jeopardy! *Artif. Intell.*, 199:93–105, 2013.

Doctor's Dilemma Question Set

As of February 2015, there were 1110 questions in the Doctor's Dilemma question set, across 26 categories. The categories cover a diverse set of topics and subfields throughout medicine.

ICU
Allergy and Immunology
History
Poisoning
Endocrinology
Pregnancy
Nephrology
Gastroenterology
Hematology
Cardiology
Gynecology
Dermatology
Ethics
ENT
Preoperative Care
Neurology
Ophthalmology
Infectious Diseases
Biostatistics
Rheumatology
Epidemiology
Pharmacology
Physical Examination
Oncology
Pulmonary
Psychiatry

The set of categories in the question set.

The questions vary not only in topic, but in the questions which truly are clinical, in the type of clinical scenario they pose. For example, there are diagnostic, etiological, and more general knowledge type questions. One of the diagnostic questions is:

“Diagnosis for nasal congestion and rhinorrhea triggered by allergen exposure.”

Other questions are etiological, concerned with the cause of diseases or disorders:

“Most common causes of postpartum thyrotoxicosis.”

More general knowledge questions are included as well:

“Name of a low-power magnification device used to detect cervical dysplasia.”

From only examining the questions in a cursory way, we hypothesized that some questions would be irrelevant, especially from sections such as *History*, *Biostatistics*, and *Ethics*. Other categories could also contain questions which are more “trivia-like” rather than clinically relevant, so we analyzed the questions closely.

Acknowledgements

We would like to thank the ACP for allowing us to use the Doctor's Dilemma questions in this research.

Methods

Of the 1110 questions, we found that 171 referred to images, and therefore were excluded. The resulting 939 questions were divided into two sets randomly – one to be used for this study, the other set aside for later projects. We attempted to extract 10 questions in each of the 26 categories from the remaining 465 questions. Some categories did not have 10 questions. The result was a set of 229 questions.

Two clinicians examined and annotated each question as either clinically relevant or not. The two annotators were in near perfect agreement ($\kappa = .87$). Disagreements were mediated and a consensus was formed.

Results

After analysis, only 3.5% (8 of 229) questions were found to be irrelevant, covering 4 categories, as in the table below.

Category	Number Relevant	Number Irrelevant
History	5	5
Ethics	5	1
Biostatistics	5	1
Epidemiology	9	1

As expected, some questions from the History category were trivia, and not really related to clinical practice. For example:

“Nobel awardee for discovery of insulin.”

We were surprised to find that half of the *History* category was relevant though, including questions such as:

“Previous name for reactive arthritis.”

Questions from the *Ethics* and *Biostatistics* categories which were irrelevant had to do with IRB approval and equation usage. Our fears about questions of a trivia-like nature slipping into the more obviously clinical categories were overall unfounded. One question in the *Epidemiology* category was found to be irrelevant, though we believe it would have been better placed in the *Biostatistics* category to begin with.

“Greek letter that represents the probability of making a type I error.”

Since overall very few questions were found to be irrelevant, **we can conclude that the Doctor's Dilemma dataset is clinically relevant**, and is suitable for use as the training set for a clinical question answering system.

Toward Question Answering: Representing Questions as Graph Queries

