Journal of Biomedical Informatics xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research

Ida Sim [a,*], Samson W. Tu [b], Simona Carini [a], Harold P. Lehmann [c], Brad H. Pollock [d], Mor Peleg [e], Knut M. Wittkowski [f]

[a] Department of Medicine, University of California, San Francisco, CA, United States
[b] Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, United States
[c] Division of Health Sciences Informatics, Johns Hopkins University, Baltimore, MD, United States
[d] Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States
[e] Department of Information Systems, University of Haifa, Haifa, Israel
[f] Department of Research Design and Biostatistics, The Rockefeller University, New York, NY, United States

ARTICLE INFO

Article history:
Received 10 July 2013
Accepted 3 November 2013
Available online xxxx

Keywords:
Ontology
Clinical research
Eligibility criteria
Clinical research informatics
Evidence-based medicine
Trial registration

ABSTRACT

To date, the scientific process for generating, interpreting, and applying knowledge has received less informatics attention than operational processes for conducting clinical studies. The activities of these scientific processes – the science of clinical research – are centered on the study protocol, which is the abstract representation of the scientific design of a clinical study. The Ontology of Clinical Research (OCRe) is an OWL 2 model of the entities and relationships of study design protocols for the purpose of computationally supporting the design and analysis of human studies. OCRe's modeling is independent of any specific study design or clinical domain. It includes a study design typology and a specialized module called ERGO Annotation for capturing the meaning of eligibility criteria. In this paper, we describe the key informatics use cases of each phase of a study's scientific lifecycle, present OCRe and the principles behind its modeling, and describe applications of OCRe and associated technologies to a range of clinical research use cases. OCRe captures the central semantics that underlies the scientific processes of clinical research and can serve as an informatics foundation for supporting the entire range of knowledge activities that constitute the science of clinical research.

1. Introduction

Interventional and observational human studies are crucial for advancing our understanding of health, disease, and therapy. Clinical research informatics (CRI) is "the use of informatics in the discovery and management of new knowledge relating to health and disease" [1]. To date, CRI has focused on facilitating conduct of clinical trials and management of clinical data for secondary research use. Yet clinical research is fundamentally a scientific pursuit, and foundational methods for CRI should support the *science* of clinical research: asking the right question, designing rigorous protocols, conducting protocol-adherent studies, fully reporting all results, and finally, making inferences and applying research results to care decisions and policy.

The underpinning of this broad range of knowledge tasks is the *study protocol* as the study's conceptual scientific structure. The *planned study protocol* drives all key scientific and biomedical activities during study execution and analysis, while the *executed study protocol* represents the study activities that actually took place. Early CRI work relegated support of protocols to the electronic sharing of text-based study protocol documents. More recently, study protocols have been reified into data models (e.g., BRIDG [2,3]) geared towards supporting the execution of clinical trials intended for submission to the U.S. Food and Drug Administration (FDA) for regulatory approval of therapeutic products, or supporting data management of clinical trial results (e.g., OBX [4] and CDISC [5]).

To provide knowledge-based support for the *scientific* tasks of clinical research, the study protocol should be modeled in a knowledge representation formalism with clear, consistent and declarative semantics that support drawing clinical conclusions from study observations. The Ontology of Clinical Research (OCRe) is such a model. OCRe is an OWL 2 ontology of human studies, defined as any study collecting or analyzing data about humans that explore questions of causation or association [6,7]. OCRe models the entities and relationships of study designs to serve as a common semantics for computational approaches to the design and analysis of human studies.

* Corresponding author. Address: 1545 Divisadero St., Suite 308, San Francisco, CA 94143-0320, United States. Tel.: +1 415 514 8657; fax: +1 415 514 8666.
E-mail address: ida.sim@ucsf.edu (I. Sim).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.11.002

In this paper, we describe the motivation methods behind OCRe, present highlights of the OCRe model, and review examples of how OCRe supports the science of clinical research. Our use cases illustrate why clinical and research informatics need to be more deeply integrated [8], to create a "learning health system" [9] that generates best evidence and also "drive[s] the process of discovery as a natural outgrowth of patient care" [10]. We posit that the study protocol, representing the essence of clinical research, is the epistemological foundation for a learning health system and that OCRe, representing study protocol elements, is a core informatics foundation for clinical research science.

2. Motivating use cases and background

To show the value of OCRe across the breadth of clinical research science, we describe its role in the five phases of a human study's idealized scientific lifecycle: (1) review and interpretation of results of previous studies to refine a scientific question; (2) design of a new study; (3) study execution; (4) results reporting; and (5) interpretation and application of the results to clinical care or policy (Fig. 1). In a learning health system, clinical practice completes the cycle as a source of new scientific questions. The five phases of a study's lifecycle are closely related and iterative (Fig. 1).

This remainder of this section presents use cases for each of the five phases. Based on these use cases, Section 3 presents the foundational capabilities that would transform informatics support for clinical research and describes the OCRe model. Section 4 then applies OCRe to selected use.

2.1. Pose a scientific question, retrieve and interpret prior studies

The first step of a clinical study's lifecycle is highly iterative: potential scientific questions are posed and revised many times as prior studies are retrieved and interpreted over time.

2.1.1. Retrieve prior studies

Investigators often frame their research questions using the "PICO" mnemonic (for **P**atient, **I**ntervention, **C**omparison, **O**utcome [11], sometimes with a **T** added for outcome timing [12]). For example, a broad initial question about vitamin D and cardiovascular risk (e.g., "does vitamin D supplementation reduce LDL cholesterol levels?") could be phrased as Intervention = vitamin D and Outcome = cardiovascular endpoints (e.g., hypertension, high cholesterol, body weight). Running this PICO query at the PubMed PICO interface [13] returned 265 studies on vitamin D's effect on high cholesterol (hyperlipidemia) at the writing of this paper.

These results are not directly helpful to an investigator because the PICO structures of these studies are buried within PDFs. A better search interface might be CTSearch with its interactive tag cloud PICO display [14], or interactive visualizations of the scientific structure of human studies like the tools that biomedical researchers have for visual exploration and query of gene sequences, pathways, and protein structures.

Even so, PICO elements alone are insufficient to support the full retrieval task. Different study objectives are best addressed by different study design types [15]. PubMed Clinical Queries [16] allows narrowing a search to appropriate study designs (e.g., prospective cohort studies to explore the association of vitamin D levels with cardiovascular outcomes, randomized controlled trials (RCTs) to explore questions of therapy), but two major problems attend this approach. First, this interface returns 2247 citations for a Narrow search for Therapy studies of vitamin D. An investigator would still need visualizations that reveal PICO and study-design features of large numbers of studies. Second, because there is no established study-design taxonomy, design types are poorly indexed in PubMed entries and searches by design type are correspondingly inaccurate [17]. There have only been a handful of published study-design taxonomies [18–20], including the Cochrane Collaboration's taxonomy [21] and Hartling's which showed a reliability of $\kappa = 0.45$ and is the basis for an AHRQ taxonomy [17]. In Section 3.2.1, we describe our OCRe-based study-design typology, which showed a moderate inter-rater agreement of Fleiss' kappa of 0.46 in a preliminary evaluation [22].

2.1.2. Interpret prior studies

Once investigators have retrieved a set of relevant studies whose designs are appropriate for the scientific question, they need to assess their evidentiary strength [23]. In statistical terms, they need to appraise the "internal validity" of each study, including the comparability of comparison groups and the existence and nature of follow-up bias [24], Critical appraisal remains somewhat of an art, as many study quality scales and bias instruments are poorly correlated, imprecise, and irreproducible [25,26]while not predictive of observed effects [27].

Many researchers lack the methodological skills embodied in guides like those from *JAMA* [28], *BMJ* [29], the Cochrane Collaboration [30], and others [31] to carry out these appraisal tasks.

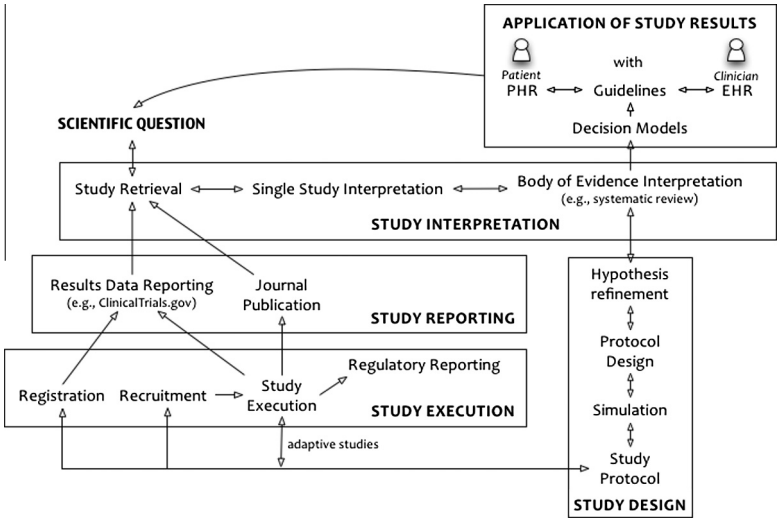

Fig. 1. Idealized scientific lifecycle of a human study within a learning health system.

Please cite this article in press as: Sim I et al. The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research. J Biomed Inform (2013), http://dx.doi.org/10.1016/j.jbi.2013.11.002