

Development and Evaluation of a Study Design Typology for Human Research

Simona Carini, MA¹, Brad H. Pollock, PhD², Harold P. Lehmann, MD, PhD³, Suzanne Bakken, RN, DNSc⁴, Edward M. Barbour, MS⁵, Davera Gabriel, RN⁶, Herbert K. Hagler, PhD⁷, Caryn R. Harper, MS, CCRC⁷, Shamim A. Mollah, MA⁵, Meredith Nahm, MS⁸, Hien H. Nguyen, MD⁶, Richard H. Scheuermann PhD⁷, Ida Sim, MD, PhD¹

¹University of California, San Francisco, CA; ²University of Texas Health Science Center, San Antonio, TX; ³Johns Hopkins University, Baltimore, MD; ⁴Columbia University, New York, NY; ⁵The Rockefeller University, New York, NY; ⁶University of California, Davis, CA; ⁷UT Southwestern Medical Center, TX; ⁸Duke University, Durham, NC

Abstract

A systematic classification of study designs would be useful for researchers, systematic reviewers, readers, and research administrators, among others. As part of the Human Studies Database Project, we developed the Study Design Typology to standardize the classification of study designs in human research. We then performed a multiple observer masked evaluation of active research protocols in four institutions according to a standardized protocol. Thirty-five protocols were classified by three reviewers each into one of nine high-level study designs for interventional and observational research (e.g., N-of-1, Parallel Group, Case Crossover). Rater classification agreement was moderately high for the 35 protocols (Fleiss' kappa = 0.442) and higher still for the 23 quantitative studies (Fleiss' kappa = 0.463). We conclude that our typology shows initial promise for reliably distinguishing study design types for quantitative human research.

Introduction

Study designs are by their nature complex, with attributes such as populations, eligibility criteria, interventions, outcomes, study processes, and analyses. Each of these attributes is crucial in establishing the “epistemological warrant” of data collection (that is, what can you learn from the data-collection process), in making a study relevant to a researcher or decision maker, or in permitting aggregation of data across studies (data sharing). A clear syntax and semantics about study design is an absolute requirement for computers to help with data sharing or decision support of clinical research; i.e., an ontology, an abstract representation, of clinical research study designs is needed. The ontology modeling task is made more challenging by the breadth of studies that are important in biomedicine and must be included in the scope: from case studies and qualitative studies that help us formulate the

issues of concern, to observational population-based studies that establish prevalence and incidence, to prevention studies, to N-of-1 trials, to definitive efficacy studies, and to practical effectiveness studies. There exist many methodology textbooks and articles for each of these study types in a wide array of disciplines, each grounded in their own epistemological view and using their own terminology for common concepts. Modern clinical and translational research, however, requires an integrated typology that spans all these disciplines and study design types. While schemes exist that define the quality of research based on research design, there is no agreed-upon typology of such studies. We therefore set out to define and evaluate a general integrated typology of study design type.

Our context for needing and using a Study Design Typology is the Human Studies Database Project, which is a multi-institutional project among 14 Clinical and Translational Science Award (CTSA) institutions to federate study design descriptors of our human research portfolio over a grid-based architecture [1]. Central to this task is the clear definition and classification of human research study types. We therefore worked towards a Study Design Typology, bringing together investigators from the Informatics as well as Biostatistics, Epidemiology, and Research Design (BERD) CTSA Key Function Committees. This modeling work is part of a larger consortial effort to define the Ontology of Clinical Research (OCRe), a broad ontology of the concepts and artifacts of clinical research [2].

We report in this paper on the Study Design Typology and its pilot evaluation for classifying quantitative human research studies. Our objective was to gain experience using the classification, and to assess classification adequacy and reproducibility in preparation for another iteration and a more comprehensive evaluation.

Study Design Typology

Use cases for using a Human Studies Database were first elicited from our CTSA colleagues. One of those use cases, “Research Characterization—Report research design,” had a high priority, motivating the need for a Study Design Typology. We defined this typology [3] through an iterative process over 8 months. The design goal was to identify the most parsimonious set of unambiguous factors that will correctly classify any study on individual humans into a set of common study design types.

A series of use cases were created to establish the scope of studies to be characterized. Authoritative methodology sources were consulted [4-6]. Reference models of clinical research (e.g., CDISC [7], BRIDG [8], OBI [9]) were consulted but we found little formal modeling of study design types. The typology was presented for feedback to biostatisticians, epidemiologists, and informaticians and at CTSA and other institutions.

The Typology scopes Human Studies as any study collecting or analyzing data about individual humans, whole or in part, living or dead. Studies on populations, organisms, decision strategies, etc. are not considered human studies in our typology. This scope includes all studies that would require human subjects approval and beyond: e.g., autopsy, chart review, expression profiles of surgical pathology, shopping habits, utility assessment, scaled instrument development using data from individual humans, etc.

Human studies are next classified as Qualitative or Quantitative (Figure 1). Qualitative research techniques seek in-depth understanding through loosely structured mainly verbal data and field observations rather than measurements. Quantitative research studies are systematic investigations of quantitative properties, phenomena, and their relationships, and involve the gathering and analysis of data that are expressed in numerical form.

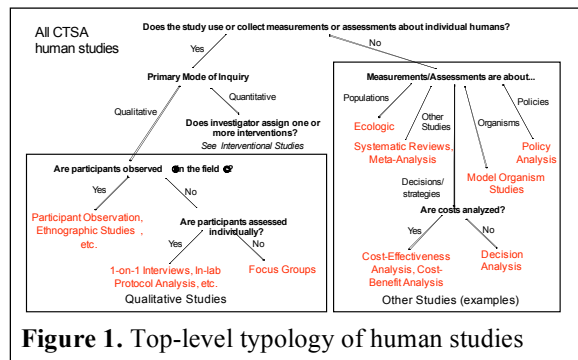


Figure 1. Top-level typology of human studies

Qualitative studies were tentatively broken out into Field Studies, Lab-based studies, or Focus Groups. This branch of the typology is under development and was not evaluated in the work we report here.

Quantitative studies are split into Interventional (Figure 2) and Observational (Figure 3) Studies. Interventional studies are ones where the investigator assigns one or more interventions to a study participant. Observational studies are ones where the investigator has no role in what treatment or exposures a study participant receives, but only observes participants for outcomes of interest.

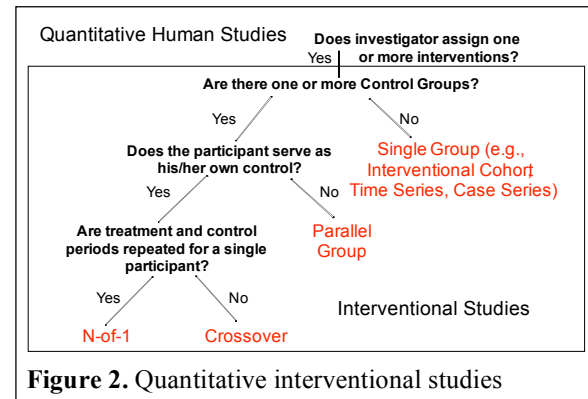


Figure 2. Quantitative interventional studies

We defined four high-level study design types for interventional studies, based on the number of control groups, whether the participant serves as his/her own control, and whether treatment and control periods are repeated for a single participant. These factors are factual and independent of judgment, and the permissible values are exhaustive and mutually exclusive, facilitating full partitioning of all interventional studies into one of these four study design types.

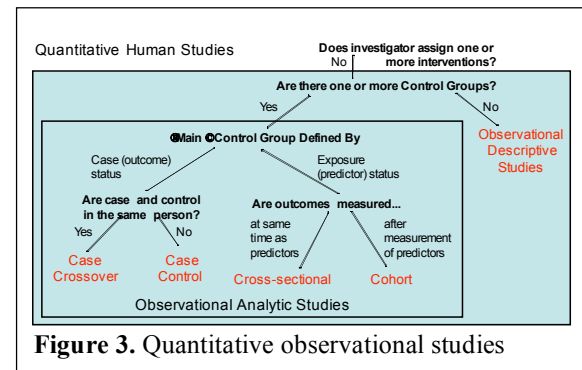


Figure 3. Quantitative observational studies

These high-level study types (in red in the Figures) represent distinct approaches to human investigations that are each subject to distinct sets of biases and interpretive pitfalls. Additional descriptors elaborate on secondary design and analytic features that introduce or mitigate additional validity concerns (Table 1).

Additional Descriptors and Value Sets
Statistical intent <ul style="list-style-type: none"> • Superiority, non-inferiority, equivalence
Sequence generation <ul style="list-style-type: none"> • Random, non-random
Allocation concealment method <ul style="list-style-type: none"> • Centralized, pre-numbered containers, etc.
Assignment to study intervention <ul style="list-style-type: none"> • Factorial, non-factorial
Unit of randomization <ul style="list-style-type: none"> • Individual, cluster
Blinding/Masking <ul style="list-style-type: none"> • Participant, investigator, care provider, outcomes assessor, statistician
Control group type <ul style="list-style-type: none"> • Active, placebo, sham, usual care, dose comparison
Study phase <ul style="list-style-type: none"> • Phase 0, 0/1, 1, 1/2, 2, 2/3, 3, 4

Table 1. Additional Descriptors for Interventional Study Designs. Each of the four design types can be further described by some or all of these descriptors.

For the quantitative observational studies subtree, we defined five high-level study design types, based on the number of control groups, whether the main control group was defined by case (outcome) or exposure (predictor) status, whether the case and control are in the same person, and whether outcomes are measured at the same time as predictors or after. Like the discriminatory factors for interventional studies, these factors exhaustively and mutually exclusively partition observational studies into one of these five design types. Additional descriptors also apply to these observational study types (not shown), although some descriptors apply only to some study types.

Mixed method studies are accommodated in this typology by classifying all methods applicable to a study (e.g., interventional with qualitative outcomes assessment).

Evaluation Methods

We evaluated the Quantitative Study subtree of the Study Design Typology. Four institutions (Columbia, Duke, UC Davis, UT Southwestern) were asked to identify and collect documentation for 7 to 10 studies from their own institutions involving individual humans. Although this evaluation study was exempted from IRB approval (Duke: Pro00015270), we nevertheless restricted the studies to completed studies and studies for which the institutions had cleared rights for such use, to reduce any perceived risks from disclosure of protocol information.

Moreover, we sought permission to use these protocols from their principal investigators, according to the local institutional practice.

The institutions were asked to obtain documents that had to have all the required information to describe the study design, e.g., grant proposals, ethics board submissions, and formal study protocols. The institutions collected protocols divided roughly equally among the following categories: 1) Translational studies: early phase studies, more biomedical, genome/phenome (e.g., GWAS); 2) Clinical studies: often focused on efficacy, e.g., phase II/III interventions, etc.; 3) Community studies: more focused on implementation of evidence into practice.

The Study Design Typology was translated into a web-based tool and deployed using SurveyMonkey, an online survey service provider. At the beginning of the survey, the reviewers identified their background experience (clinical researcher, statistician, etc.), identified the protocol, and answered questions about the protocol based on the classification scheme (Figures 1–3). Once the high-level study type was identified (e.g., interventional parallel group), the reviewers selected details regarding additional, type-specific study descriptors (e.g., statistical intent), as in Table 1 for interventional studies. For each classification question, the choice “Not Clear/Not Specified” was available. Reviewers were encouraged to provide comments about the classification questions, the study type that resulted from traversing the decision tree and the list of additional descriptors.

Each reviewer was asked to classify the protocols for his or her institution and those from two other institutions, chosen at random, so each protocol was classified three times. There was only one reviewer per institution, chosen to be roughly representative of clinical researchers who have some study design expertise but were not biostatisticians or epidemiologists.

In order to assess the reliability of agreement between the three raters for all protocol study design classification ratings we calculated the Fleiss’ kappa. Fleiss’ kappa represents the degree of agreement in classification over what would be expected by chance alone, and is anchored by the value 0 for representing no agreement to 1 for perfect agreement.

Results

There were nine quantitative high-level study design types to classify into (four interventional and five observational).

Thirty-six protocols were collected from the four institutions. One protocol was classified only twice and was therefore excluded from the analysis. Although the evaluation is focused on quantitative studies, our study set inadvertently included some qualitative or non-human studies. For all 35 analyzed protocols, the Fleiss' kappa, an overall measure of agreement, was 0.442. For the subset of 23 quantitative studies, the kappa was slightly higher at 0.463.

Reviewers took advantage of the ability to enter comments and did so in reference to the branching questions, the final study type classification, and the additional descriptors. Some comments related to the typology's language being less appropriate for behavioral trials than for drug studies. A recurring theme in comments throughout the survey was the need for better definitions (e.g., what is a control group). Yet another set of comments related to study phases (e.g., Phase I, Phase II) for interventional trials, primarily about confusion regarding the definition of each phase. There were several comments on the question asking whether the participant served as his/her own control. More complex situations also arose, as when one reviewer had difficulty categorizing a dose escalation study as N-of-1 vs. Crossover Study. This ambiguity likely arose because a study of several patients all undergoing the same N-of-1 repeated within-patient treatment episodes can also be construed as a crossover study.

Discussion

This study is the first of a series to validate a Study Design Typology for classifying human research study designs. Our results show initial promise for classifying quantitative studies reproducibly using our typology. The most apparent weaknesses of the current typology were primarily related to tool presentation (e.g., missing definitions) than to fundamental conceptual errors of the typology. However, it was clear from the comments that users were particularly confused about several distinctions. One was whether a study is on individual humans. For example, is a study collecting blood samples a human study? By our definition, if the data is collected about an individual human, then it is a human study, but there was not a uniform understanding about this. There was confusion about qualitative versus quantitative studies. Questionnaires on "qualitative" phenomena (e.g., mood) that use quantitative scales analyzed using quantitative approaches (e.g., Likert scales) are quantitative studies, but were sometimes classified as qualitative. Also, studies following "quantitative" design types

(e.g., interventional parallel group) could include qualitative outcomes and analyses. The typology was weak in handling hybrid study designs and mixed methods. Likewise, a clearer distinction between N-of-1 and Crossover study types is needed. Finally, the notion of a control group was problematic, for example in dose ranging studies, or in observational studies where there is no one clear control exposure. There were also cases where terminology was an impediment to reviewers from different epistemological backgrounds. More careful assessment of the comments is needed to inform the continued development of the typology in terms of its content and structure, and of its presentation.

A validated and usable Study Design Typology would play a critical role in any clinical research management system. A study's basic design type greatly influences the relevant protocol details that need to be captured. It also influences the type and extent of ethics review required, the degree and nature of regulatory oversight, and the types of challenges faced in executing the study. After the study data are collected, the choice of an analytic strategy depends on the research design, as do the types of potential bias that would need to be considered during analysis and reporting. Thus, a valid Study Design Typology has a profound impact throughout the clinical research life cycle.

Once further validated, our Study Design Typology has many uses. It can be used manually by investigators early on in the study design process to highlight critical methodological features of their planned protocols, and to describe them in a consistent way. In particular, investigators in training will benefit from thinking through the key distinctions between, for example, defining an observational control group using case status or exposure status. Furthermore, if the factors in our typology can be collected in standardized fashion across studies, then retrieval of studies by methodology and design can be greatly facilitated for researchers who need to know the methodological status of prior studies, systematic reviewers who use methodological filters to screen studies, systematic reviewers and meta-analysts who use methodological knowledge to weight the epistemological strength of evidence available, and research administrators who need to characterize the portfolio. Indeed, the Study Design Typology would be most useful if it can be represented computationally within clinical research informatics systems. We explicitly tried to identify discriminatory factors that would be assessable either by human reviewers or by an automated process by a clinical research management system.

Limitations

One limitation of this study is that the test set of studies included qualitative studies when we had intended to test only the typology for quantitative studies. Because we had not finalized the typology for distinguishing qualitative from quantitative studies, some studies were misclassified, leading to the lower kappa for agreement across all 35 protocols. Another limitation is the relatively small test set, drawn entirely from academic medical centers, which may not be representative of protocol documents or study designs used in other sectors of clinical research. Finally, in this preliminary analysis, we have not analyzed agreement on the additional descriptors, which may show less agreement than overall study design type.

Current work

Current work on the Study Design Typology includes incorporating into the quantitative study subtree the feedback gathered during this pilot evaluation, finalizing the qualitative studies subtree, and validating the typology with a broader sample of protocols with investigators from a wider range of backgrounds. An in-progress version of the typology is at CTSpedia.org, the CTSA BERD group's knowledge management website [3]. The typology is also being incorporated in the study design module of the Ontology of Clinical Research (Figure 4) and is already publicly available [2]. It will thereby serve as the semantic standard for study design types within the Human Studies Database Project as it federates human study designs among CTSA institutions.

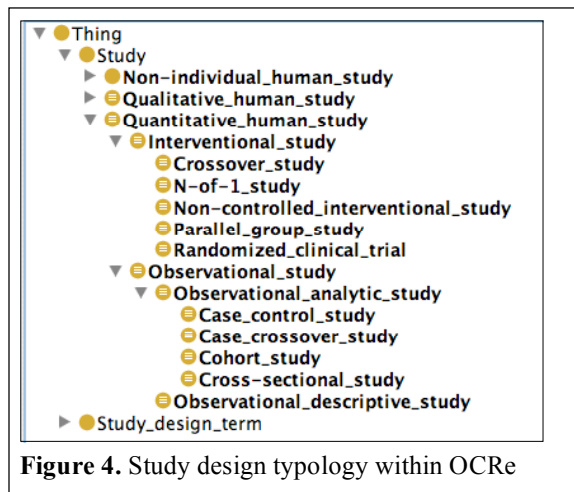


Figure 4. Study design typology within OCRE

Conclusions

We have defined a typology of study design types for human research. For quantitative studies, this typology defines unambiguous discriminatory factors

with exhaustive and mutually exclusive value sets that partition quantitative research studies into one of nine high-level study design types. Our pilot evaluation of this typology demonstrates initial promise for reliably classifying study design types in human research. Further development and validation is needed, especially through wider use by investigators, who should benefit from a clear and standardized tool for accurately characterizing their study's design.

Acknowledgements

Funded in part by UL1RR024156 (Columbia), UL1RR024128 (Duke), UL1RR025005 (Johns Hopkins), UL1RR024143 (Rockefeller), UL1RR024146 (UC Davis), UL1RR024131 (UCSF), UL1RR025767 (UTHSC-San Antonio), and UL1RR024982 (UT Southwestern) from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. We thank Yuanyuan Liang for assistance with the statistical analysis.

References

1. Human Studies Database Project. 2009 [cited 2009 March 13]; Available from: <http://rctbank.ucsf.edu/home/hsdb.html>
2. Ontology of Clinical Research. 2009 [cited 2009 March 13]; Available from: <http://rctbank.ucsf.edu/home/ocre.html>
3. CTSpedia. 2009 [cited 2009 July 27; Available from: <http://www.ctspedia.org/do/view/CTSpedia/StudyDesignCategorizer>
4. Hulley S, Cummings S (eds). Designing Clinical Research: An Epidemiologic Approach. New York, NY: Lippincott; 2001.
5. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* 2002;359(9300):57-61.
6. Shadish W. Experimental and Quasi-Experimental Designs for Generalized Causal Inference: Houghton Mifflin; 2002.
7. CDISC -- Clinical Data Interchange Standards Consortium. 2009 [cited 2009 March 3]; Available from: <http://www.cdisc.org>
8. The BRIDG Group. BRIDG. 2009 [cited 2009 March 10]; Available from: <http://bridgmodel.org>
9. Ontology for Biomedical Investigations. 2009 [cited 2009 March 13]; Available from: http://obi-ontology.org/page/Main_Page