Model Formulation ■

The Primary Care Research Object Model (PCROM): A Computable Information Model for Practice-based Primary Care Research

STUART M. SPEEDIE, PHD, ADEL TAWEEL, PHD, IDA SIM, MD, PHD, THEODOROS N. ARVANITIS, DPHIL, BRENDAN DELANEY, MD, KEVIN A. PETERSON, MD, MPH

A b s t r a c t   **Objectives:** Chronic disease prevalence and burden is growing, as is the need for applicable large community-based clinical trials of potential interventions. To support the development of clinical trial management systems for such trials, a community-based primary care research information model is needed. We analyzed the requirements of trials in this environment, and constructed an information model to drive development of systems supporting trial design, execution, and analysis. We anticipate that this model will contribute to a deeper understanding of all the dimensions of clinical research and that it will be integrated with other clinical research modeling efforts, such as the Biomedical Research Integrated Domain Group (BRIDG) model, to complement and expand on current domain models.

Design: We used unified modeling language modeling to develop use cases, activity diagrams, and a class (object) model to capture components of research in this setting. The initial primary care research object model (PCROM) scope was the performance of a randomized clinical trial (RCT). It was validated by domain experts worldwide, and underwent a detailed comparison with the BRIDG clinical research reference model.

Results: We present a class diagram and associated definitions that capture the components of a primary care RCT. Forty-five percent of PCROM objects were mapped to BRIDG, 37% differed in class and/or subclass assignment, and 18% did not map.

Conclusion: The PCROM represents an important link between existing research reference models and the real-world design and implementation of systems for managing practice-based primary care clinical trials. Although the high degree of correspondence between PCROM and existing research reference models provides evidence for validity and comprehensiveness, existing models require object extensions and modifications to serve primary care research.

■ **J Am Med Inform Assoc.** 2008;15:661–670. DOI 10.1197/jamia.M2745.

Introduction

Research findings are most valid and generalizable when they arise from an appropriately designed randomized

clinical trial (RCT), performed in the setting in which the research is to be applied. Over the course of 1 year, more medical office visits are performed in primary care practices than in all other medical specialties combined.[1] Although primary care provides the "final common pathway" for the introduction of new findings into the community, primary care practices have historically been isolated from participation in RCTs. The development of practice-based research networks (PBRNs) has introduced new methods to facilitate clinical research in community primary care practices by enabling collaboration, supplying necessary expertise, and providing additional financial and technical resources. Defined as a group of ambulatory practices, devoted principally to the primary care of patients and affiliated in their mission to promote research and improve quality, PBRNs have been described as "new clinical laboratories for primary care research and dissemination."[2]

The organization and management of clinical research performed in a primary care practice differs substantially from that in other clinical research settings. This difference is exemplified by the fact that even the most common conditions represent only a few percent of primary care visits. Consequently the number of qualified subjects recruited

from a single clinic for a given study will likely represent a small percentage of the total number of patients seen. This can make it expensive and logistically difficult to recruit patients for primary care studies compared with a specialist environment. Although not a unique feature, many more trials conducted in primary care settings are dependent on large numbers of clinical sites with small numbers of subjects, introducing new challenges in recruitment, training, and protocol compliance. The range of relevant and important research questions is considerably broader and primary care interventions tend to be more complex, often involving health services or social care approaches. Research outcomes may include resource utilization, symptoms score, satisfaction scales, quality of life measures, or other complex outcomes infrequently found in other clinical settings.

Because large community-based studies are currently difficult to do, there is a relative paucity of primary care trials compared with their need. Yet as the population ages and the chronic care disease burden increases, enhancing the RCT evidence base for primary care will become increasingly important. Therefore systems that can support the process of clinical research in this setting will need to be designed and implemented.

The work described in this article grew out of work on the electronic Primary Care Research Network (ePCRN), a National Institutes of Health–funded Roadmap Initiative project. The ePCRN is building a highly secure, grid-based information system infrastructure that, among several goals, facilitates the conduct of RCTs in primary care.[3] It is being designed to enable primary care practices, anywhere in the United States, to link with researchers in academic centers or to the National Institutes of Health to facilitate trial planning, recruitment, entry, participation, and follow-up of subjects in multidisciplinary RCTs. The network accomplishes this goal by using a collection of systems that: (1) provide a highly secure communications network combining Citrix servers and clients (a thin client model) using 3-factor identity management and Kerberos-based,[4] system-to-system communications layered over Open Grid Services Architecture—Database Access and Integration (OGSA-DAI) grid technologies,[5] (2) create a collection of grid nodes located in primary care practice environments that consist of local security and a standardized registry of patients using the OGSA-DAI distributed database technology to facilitate clinical trials recruitment, and (3) use a primary care trials management system that facilitates the conduct of trials once the eligible patients have been recruited. The overall goals of the ePCRN are to provide the ability to perform multiple, large-scale collaborative studies in primary care settings throughout the United States, improve efficiency and reduce costs for individual trials, and provide easier access for clinical trial data retrieval and analysis.

This article describes an effort to create a standard, computable representation of an RCT that meets the needs of primary care research. It describes the Primary Care Research Object Model (PCROM), an information object model that meets these needs, and compares it with a major existing standardization effort in the area of RCTs.

Need for Standardized, Computable Trial Descriptions

The RCT is the gold standard by which the effectiveness of health care interventions is determined. The defining feature of an RCT is the use of randomization as an attempt to ensure that subjects have an equal probability of assignment to experimental groups, and hence to reduce the likelihood of known and unknown confounders affecting the results. The Food and Drug Administration, and other regulatory authorities worldwide, prefer RCTs for demonstrating the safety and efficacy of drug therapies, and the Centers for Medicare and Medicaid Services seldom will not provide reimbursement for new medical procedures and diagnostic methods without evidence provided by RCTs. The published medical literature contains tens of thousands of such randomized clinical trials.[6] The evidence-based medicine approach to clinical practice relies on the results of published RCTs as the primary criteria for making medical decisions about therapies in individual patients and gives additional credence to the evidence if a result is demonstrated in meta-analyses of randomized clinical trials. The Cochrane Collaboration[7] maintains an on-line database of meta-analyses, widely used by agencies with an interest in the promulgation of clinical guidelines, such as the National Institute for Clinical Effectiveness in the UK and the Agency for Health Care Research and Quality in the United States. The health care community anticipates that the conclusions from such work will soon be used as the basis for widespread reimbursement for medical services in the form of value-based purchasing.[8] In fact, a number of efforts are already underway to do so.[9,10]

The RCTs are recognized as providing high value for evidence-based clinical practice. As a result, a standardized, computable representation of trials is needed to support activities throughout a trial's lifecycle, from a trial's execution, to comparing and combining trial results for meta-analysis, to applying trial results to clinical care. As reviewed herein, many of the existing research modeling efforts have focused on representing trials mostly for execution, wherein the needs of evidence-based medicine focus more on comparing, combining, and applying trial results. We summarize the needs for a computable representation of RCTs with particular attention to the needs for practice-based primary care research.

In primary care research, a standard, computable representation of RCTs is required to uniformly identify and recruit potential trial participants from multiple, geographically dispersed sites of care that are often small physician practices. The general need to uniformly identify individuals who are suitable candidates for trials is a pressing one. Trial recruitment accounts for about 30% to 40% of trial costs, only 1 of 20 patients approached for potential enrollment is eventually enrolled, and a majority of trials have recruitment delays.[11] Numerous publications have pointed out the difficulties of identifying and recruiting trial participants.[12–15] One very promising solution to the problem is to identify such persons through information from electronic medical records systems created as part of the medical care process. However, to do so, trial eligibility criteria and other relevant aspects must be expressed in standard terms that can facilitate automated searching for such patients within a

system of electronic records. Because of the wide range of conditions studied in primary care research, efforts at standardized representation (e.g., for eligibility rules) need to go beyond enumerating standards for a particular condition (e.g., eligibility for breast cancer trials) to a more generic approach.

A standard representation of RCTs is necessary to clearly and accurately communicate the structure of a trial for uniform implementation at multiple sites. One of the challenges in such multisite trials is consistent implementation, when numerous individuals at the different sites are charged with executing the trial. Inconsistencies can arise, not necessarily from deliberate deviations from the trial's protocol but from different understandings of the protocol's elements. Consistency is supported by a common understanding of the relevant aspects of the trial. Such a common understanding is facilitated by communication of a shared standard representation of the trial's elements, and is especially important in practice-based trials where site investigators are often less familiar with conducting clinical research.

Aside from needing a standard representation of RCTs to help run a trial, such a representation is essential for combining results from multiple heterogeneous RCTs in a meta-analysis, where small differences in trial design and outcome measures may lead to inaccuracy in the overall effect estimate.[16] The ability to determine which elements of 2 or more trials are similar and which are different is critical to detecting such differences. Without a standard method of representing the components of a trial, it is necessary to depend entirely on the interpretations of readers regarding the comparability of trial elements. There is an overlapping and equally important issue of the standard representation and reporting of clinical data for the purposes of comparing the results of multiple clinical studies.[17,18] This latter issue has been and continues to be extensively addressed in the literature and will not be the subject of this article, although the investigators do acknowledge its importance.

Integral to the task of conducting a systematic review of RCTs is the need to objectively evaluate the quality of the trials. For this task, it is important to be able to understand the design elements of a given trial and be able to compare it with others of known quality. These comparisons require identification and description of trial components such as treatment allocation strategies, in clear and unambiguous terms, to make valid judgments about the overall trial quality. The lack of a standard representation of trial design features impedes this process by making it more difficult to locate and characterize the important elements of a trial that are used in critical appraisals of trial evidence. A standard, computable representation would improve the ability to evaluate the quality of RCTs and provide a basis for doing so in an automated fashion.

A standardized representation of a trial promotes the ability to determine the applicability of a trial result in the treatment of an individual patient. The ultimate purpose of a clinical trial is to discover a method to improve the health or quality of life for an individual patient. One of the major challenges that repeatedly arises deals with the determination of whether or not the outcome of a particular trial or set of trials is applicable to a given patient. The decision often involves a review and evaluation of the structure of the trial with particular attention to the eligibility criteria, treatment regimen, and observed results. The lack of a standard trial description makes this task more difficult both in terms of simply locating the necessary information within a published trial description, but also in the interpretation and understanding of that information.[19]

The need for automated search and retrieval of clinical trial descriptions and their results spans trial design, execution, and matching of trial evidence to patients. Fundamental to the formulation of clinical trials and the use of trial outcomes to patient treatment decisions is the ability to locate and retrieve relevant existing clinical trials. In the modern electronic world, search and retrieval strategies are based on the computable characteristics (metadata) of published trials, such as Medical Subject Headings classifications and terms embedded in the text (among many such methods of tagging or categorizing a given published work). These classification schemes assume that there is some underlying, commonly understood and accepted description that characterizes all such clinical trials. Although numerous efforts are addressing this issue, a commonly accepted and widely used representation is still largely lacking. As a consequence, the process of searching out applicable RCTs continues to be inhibited both in terms of accuracy and efficiency.

Finally, fundamentally, a standard, computable representation of RCTs is a prerequisite for interoperability among clinical trial management systems,[20] clinical information systems, and decision support systems. The need for interoperability arises from the development of electronic clinical trials management and trial registration systems[21] that require a structured representation of a clinical trial to accomplish their functions. The issue of interoperability arises when it is desired to move information from one system to another. For example, it is desirable to be able to register a trial in one system and be able to then disseminate that trial across multiple registration systems. The originating system must be able to construct a message containing a description of a trial that is understood in a computational sense by the receiving system. This requires both syntactic interoperability in which the structure of the message itself is understood and semantic interoperability in which there is a commonality of meaning in the contents of the message structure.[22] Similarly, both semantic and syntactic interoperation are required for exchanging information among clinical trial managements systems, as may be needed when combining systems from several vendors to run a single trial, for exchanging information with electronic medical record systems, and with decision support systems that may be separate from the electronic medical record. The challenges of interoperation are heightened in practice-based research where electronic health record penetration is still relatively low and where the data exchange infrastructure is nascent (e.g., regional health information organizations).

Existing Efforts to Create Standard RCT Models

Several organizations, including Clinical Data Interchange Standards Consortium (CDISC), Health Level Seven (HL7), the National Cancer Institute, and the World Health Organization (WHO), have recognized the need for and are developing standard trial representations or models. The

CDISC aims "... to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of health care."[23] Their efforts focus on clinical trials that support submissions to regulatory agencies such as the Food and Drug Administration and hence tend to be focused on pharmaceutical interventions and on meeting regulatory requirements as expressed in the rules and regulations of the U.S. Food and Drug Administration.[24] In partnership with HL7, CDISC has undertaken a similar effort through its Regulated Clinical Research Information Management committee that has attempted to characterize clinical trials and the information they generate in terms of the Reference Implementation Model Version 3.0 (RIM 3.0).[25] At the same time the Cancer Biomedical Informatics Grid (caBIG) of the National Cancer Institute commissioned the development of the Clinical Trials Object Data System to address the problems of combining the results of multiple clinical trials of cancer treatments to increase the speed of discovering of new methods of treatment.[26]

Recognizing the commonality of these efforts, the involved principals are working to harmonize their various efforts into a single standard domain model of regulated clinical research that shares a common set of terms spanning CDISC, HL7, and all caBIG, among others. This ongoing work is incorporated into the Biomedical Research Integrated Domain Group (BRIDG) model that represents the process and results of protocol-driven biomedical/clinical research.[27] That model is serving as a focus for trial model standardization worldwide.

Another set of development efforts involves the establishment and use of clinical trials registries. One of the earliest was the National Library of Medicine's ClinicalTrials.gov database,[28] which laid the groundwork for implementation of the International Committee of Medical Journal Editors' policy requiring the public registration of clinical trials to be considered for publication.[29] Subsequently the WHO established a global network of international, national, and regional trial registries. These registers collaborate in capturing the 20-item Trial Registration Data Set of design and administrative information to be publicly registered before the enrollment of the first participant for all trials worldwide.[30] To minimize the need for multiple data entry and to facilitate data reuse for multiple purposes, such as health systems planning, meta-analysis, and patient recruitment, the WHO is working with CDISC to define a standard XML model for interchanging the WHO Trial Registration Data Set among the Register's Network, and this model is being incorporated into BRIDG.

Finally, efforts are ongoing to define an ontology of clinical trials and clinical research. The Ontology of Clinical Investigations (OCI) is part of a broader effort to define a suite of Ontologies for Biomedical Investigations (OBI) based on the Basic Formal Ontology (BFO) upper-level ontology.[31] The Immune Tolerance Network in conjunction with Stanford Biomedical Informatics Group has reported to the development of the Epoch framework for clinical trials management.[32] Another effort, the Ontology of Clinical Research (OCRe), is a collaboration of UK CancerGrid and The Trial Bank Project.[33] The precise relationship of these ontologies to domain models such as BRIDG is as yet unclear.

The development of the PCROM has taken place largely at the same time as the work described above. It has borrowed where possible from these efforts, but has been driven by a different, though similar, set of needs focused on the development and implementation of a system to support RCTs in the primary care setting.

Methods

To meet the ePCRN's goal of facilitating primary care RCTs, we are developing a collection of information systems that we label a clinical trial management system (CTMS) to support the needs and requirements of primary care research.[34] Compared with the majority of commercially available CTMS, which are targeted to larger multitrial research-intensive medical centers and are often of more limited scope in terms of a primary focus on data collection and reporting, ours addresses the specific research environment of primary care, as previously described. It incorporates the functionality of many of the currently available CTMSs but is designed to perform a broader range of functions specifically targeted to primary care research. We did limit the scope of the initial system design to randomized clinical trials because of their prevalence and desirability. Future plans call for the incorporation of other types of research designs, including observational studies and quasi-experimental designs. For this first implementation we made the simplifying assumption that the logic of the research protocol had been fully designed and finalized before use of the system, and that all steps after the conclusion of data analysis such as manuscript preparation, regulatory submissions, etc., would be outside the scope of the ePCRN's remit.

In designing the CTMS it quickly became evident that a standard, computable representation of an RCT would be required to build a system to efficiently and effectively support multiple trials. Unfortunately, the standard approach to creating a clinical trial information system in the past has too often been the development of a custom-coded software package for each trial, an activity that is inefficient and potentially counterproductive to creating sharable clinical trial descriptions and data. An approach that is built on a commonly accepted model of an RCT should both be more efficient and extensible as the volume of RCTs increase and the system is required to support larger numbers of trials. Accordingly we set 4 goals for our design. The CTMS should:

1. Be based on a standard representation of a primary care RCT that separates domain semantics from the system implementation to increase the expressive power and flexibility of the system.
2. Provide assistance in the design and implementation of clinical trial protocols across multiple sites.
3. Facilitate identification and recruitment of candidates for existing clinical trials within Health Insurance Portability and Accountability Act and internal review board restrictions.
4. Support and manage the execution of clinical trials conducted in the multiple, primary care sites.

Designing such a system required a thorough analysis of RCT characteristics and how they are conducted in primary care settings involving multiple sites. To develop this understanding, we applied the principles and procedures of

software engineering using the Unified Modeling Language (UML)[35] to identify the actors involved, define use cases, describe the relevant activities, and create the supporting class model of primary care RCTs. Along with other representational technologies, such as frame-based models and description logics, UML provides the capability of building precise, unambiguous, and complete models, which is an essential feature for capturing real-life domain concepts and their relationships. We chose to use UML over these other technologies because it offers a widely accepted set of standardized representations[36] that are readable and easy to understand, are applicable and familiar to multidisciplinary teams, and extensible to different domains. Use of a UML also provides great flexibility through a large set of useful and predefined constructs for domain analysis (e.g., through modeling at the Domain Analysis Model level of abstraction), yet it also provides a semiformal definition of syntax and semantics for software development (e.g., through modeling at the Data Model level) that are independent of the implementation language. The UML provides 9 different representation tools or diagrams covering the different stages of a system development life cycle: use case, activity, class, state, sequence, collaboration, object, component, and deployment.

Figure 1 shows the various actors that were identified as being involved in primary care research and includes both individuals who undertake research-specific roles as well as systems and organizations that support them. The most abstract level is the StudyActor, describing any entity that plays a role in the research process. Three major subcategories of StudyActor are the PersonActor, which represents individual persons who either participate in or are involved in execution of the study. The OrganizationActors are entities that are involved in carrying out the research study and include sponsoring organizations and research sites. The StudyStaff actor is further delineated into the various administrative, support, and execution roles required to successfully execute the study.
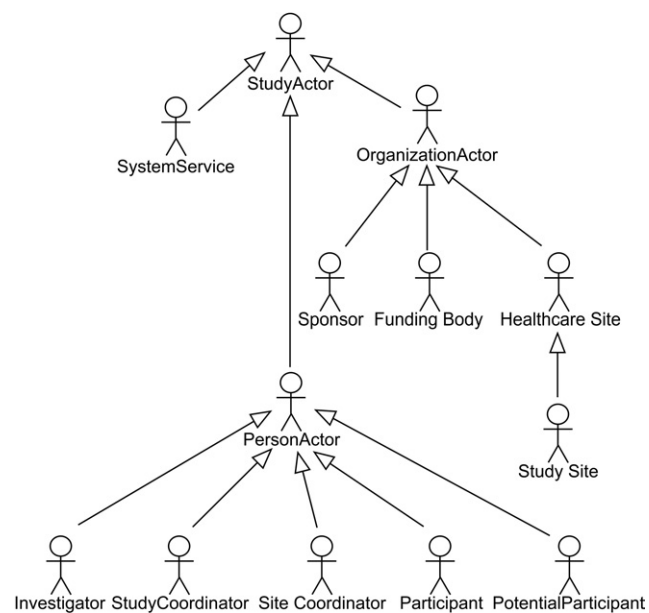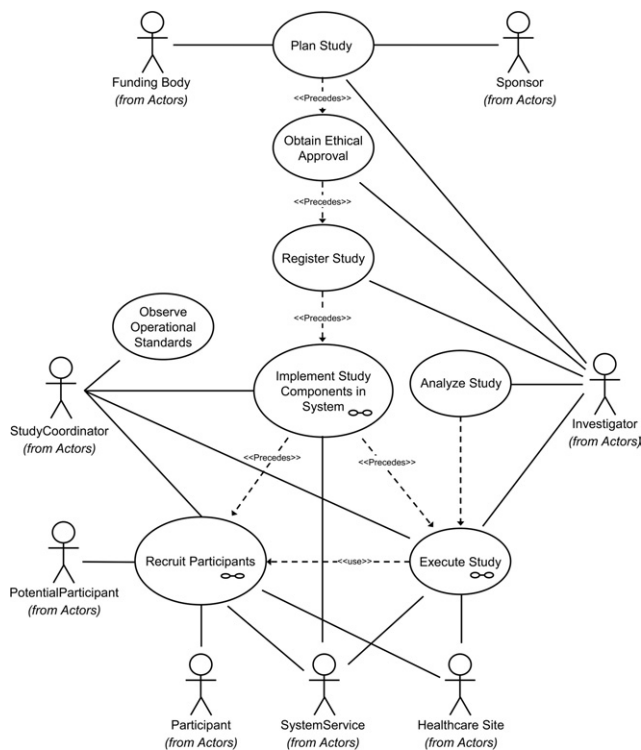

F i g u r e 2.   The primary care RCT use case and its components.

Following the UML methodology, we began by defining a set of use cases based on examples of conducting RCTs in primary care and attempted to identify their essential components, including conditions before and after study. The use case descriptions were derived from studies in which the clinician-authors of this article had participated either as primary investigators or as co-investigators. Figure 2 is a visualization of the high-level use case, its components, and interactions with the identified actors. The use case consists of several subcases that detail different components of the RCT. The Plan Study use case incorporates all activities carried out in preparation for obtaining sponsorship and funding and submission for ethics approval and registration in a trials database. The Implement Study Components in System subcase focuses on designing the details of a specific clinical trial and incorporating the resulting description into an operational form, including such aspects as case report forms, explicit eligibility criteria, and security protocols. The Obtain Ethical Approval use case detailing the internal review board approval process and the Register Study use case focusing on formal trial registration with entities such as ClinicalTrials.gov are included because they are important administrative actions that must take place before the trial begins. Once the trial is initiated, it involves use cases whose goals are to recruit subjects for the trial. Once a subject is recruited, the Execute Study use case is applied to deliver or administer the experimental condition (drug treatment, procedure, control, etc.) and observe its effects. Analysis and summary of the trial's results are detailed in the Data Analysis use case.

Based on these use cases, we established the scope of our modeling effort and proceeded to develop a draft set of

F i g u r e 1.   Primary care research RCT actors and their relationships.

activity models that corresponded with our use cases. These models defined the steps in the process of carrying out the tasks specified in the use cases, including descriptions of discrete steps and their order, whether deterministic or conditional. Making use of the activity models and use cases, we then proceeded to develop a class model consisting of objects or classes and their static associations that became the Primary Care Research Object Model. These classes were labeled and defined to reflect their purpose and function. The purpose of the resulting model is to specify the necessary classes and their relationships that would be required to fully support the development of the CTMS that would implement the primary care RCT use case.

The resulting information model was subject to validation through review by groups of experts in 3 settings. Firstly the model was discussed in detail at a meeting with 5 experienced trialists, members of the Department of Primary Care Clinical Sciences at the University of Birmingham. Secondly the model was presented and discussed at a workshop at World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA) 2007, Singapore, where 16 participants from around the world were present. Finally, the model was discussed at a meeting of the Federation of Practice-Based Research Network Directors in May 2007. The final version of the model was defined on the basis of the comments received.

The modeling effort reported in this article was initially independent of the work being conducted by caBIG, CDISC, and HL7 to create the BRIDG model. Because the work on both models was taking place at the same time, we undertook a comparison of our model with the BRIDG Release 1.0 model on a class by class basis once our work was completed.[37] Each of the PCROM classes was directly mapped to a corresponding BRIDG class where there appeared to be a high degree of correspondence between the two. If there was no direct correspondence, we next explored the possibility that our remaining classes were either superclasses or subclasses of existing BRIDG classes. With any of the remaining classes or objects that were not mapped to BRIDG by 1 of these 2 methods, we then evaluated whether they were missing from BRIDG because of some unique characteristic of primary care research RCTs or were more general concepts related to RCTs that had not yet been formally incorporated into the BRIDG model.

Results and Discussion

Figure 3 diagrammatically portrays the class model that supports the primary care RCT use case by specifying the labels of the concepts involved and exhibiting their relationships through linkages. Table 1 (available as a JAMIA online-only data supplement at www.jamia.org) lists each of the classes (class, concept, and object are used interchangeably) in the model, provides a definition, indicates which terms are linked to which others and the nature of that link, and which concepts make use of objects from other models such as BRIDG, CDISC, or WHO and describes their relationship to the BRIDG model. Together Figure 3 and Table 1 constitute the Primary Care Research Object Model in its current form. It is important to note that attributes of these classes are not listed in this article because they are still in a draft form that continues to evolve as we build the CTMS. The discerning reader will also notice that the level of granularity varies considerably within the model. Again this reflects the state of development of the various components of our CTMS, and as with the BRIDG model, the PCROM presented here should be considered to be the initial formal release of a model that is subject to ongoing development and elaboration. In the following discussion, unusually capitalized words and concatenated terms are class labels taken directly from the PCROM.

The model is organized into 3 interconnected submodels as listed in Table 1. The first of these is the Trial Process submodel, which represents the information used by and/or generated by the individual steps or activities carried out in an RCT. The primary class in this submodel is an Activity such as administering an experimental treatment. Instances of Activities are related to each other by ActivityActivityRelationships, which can include time precedence (e.g., Activity B follows Activity A) and conditions (e.g. perform Activity B only if Activity A results in a positive finding, else perform Activity C). The balance of the objects in the submodel describes the principal types of RCT activities, including interventions or experimental treatments, assessments of patient status and condition, and observations of data results from measurement procedures.

The Organizations, People, and Systems submodel focuses on the entities involved in a clinical trial. The core concept is that of a Study Actor, which is an entity that has some relationship to a study. Organizations are entities that are groups of other entities (organizations or people) that are brought together for some common purpose. A Person is the other core concept that specifies an individual human being. Investigators and study participants are examples of Persons in this model. The rest of the submodel specifies the various types of organizations and persons that are involved in a clinical study. It is linked to the Trial Process submodel through associations with the Study object and to the Trial Information submodel through associations with Activity.

The Trial Information submodel focuses on classes that describe the nature of the trial itself with the core concept being that of a Study. For example, one class models the concept of the intervention being tested in the study, with class distinctions characteristic of primary care research. To illustrate, one subclass of Intervention commonly tested in primary care research is CognitiveIntervention, which is characterized as some form of human communication that attempts to influence the participant's thoughts, beliefs, and behaviors. An example of such an intervention might be smoking cessation intervention through a support group process. The Trial Information submodel includes classes describing important processes in study development and execution, including ethics approval, trial registration, operational monitoring, and analysis of study outcomes. The model also specifies other important descriptive components of the study, including study outcomes, eligibility criteria, the methods for allocating subjects to treatments, and the timeline of study events. It is linked to both the Organizations, People, and Systems submodel as described previously and to the Trial Process model through the
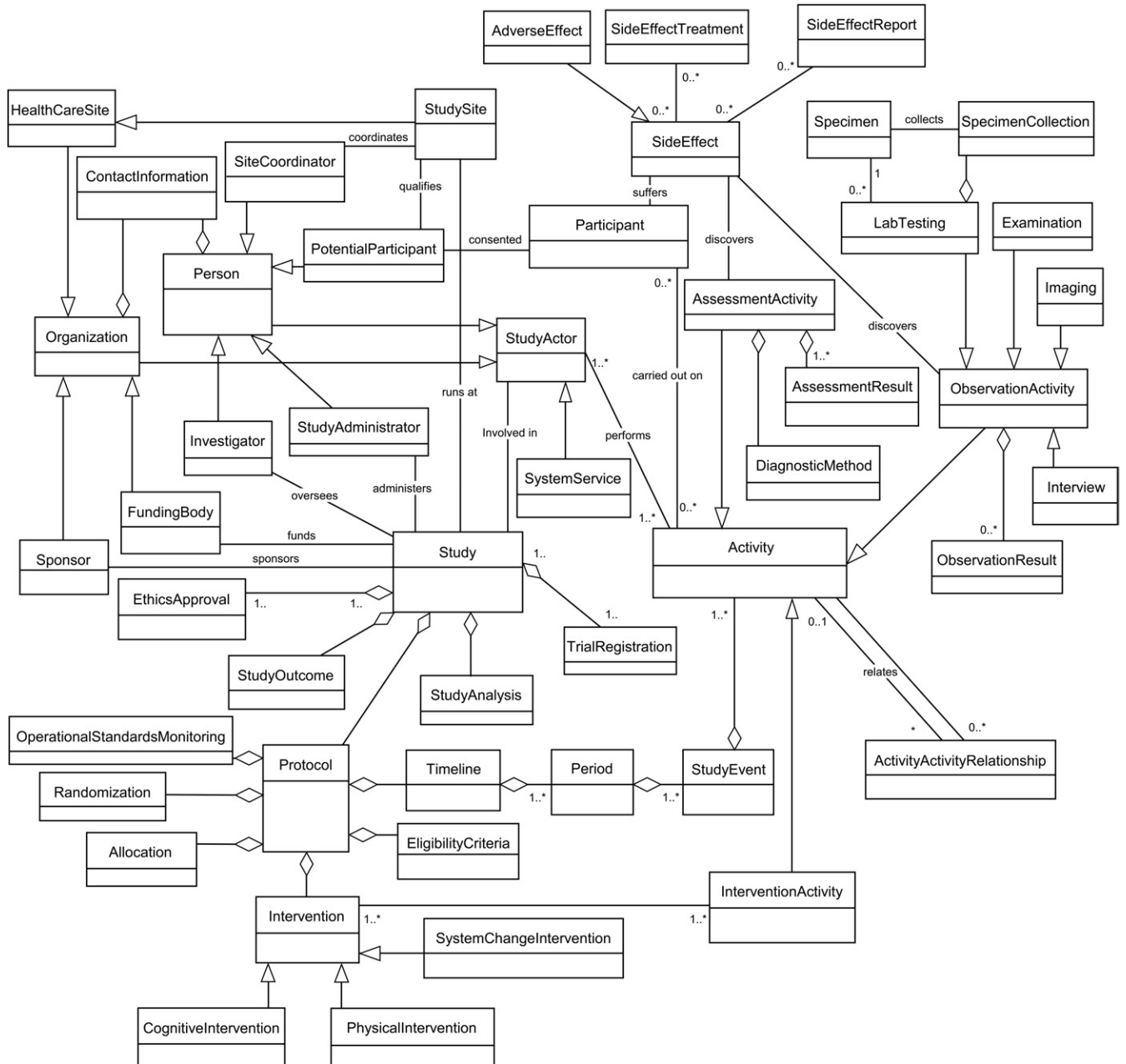
Figure 3.   Primary Care Research Object Model (PCROM).

aggregation of Activities that is associated with a StudyEvent.

A Comparison of PCROM and BRIDG

A comparison between the BRIDG and PCROM models was conduct and is also detailed in Table 1 (available as a JAMIA online-only data supplement at www.jamia.org). We noted that for a number of classes, there is a reasonably direct mapping of PCROM to BRIDG. For example, the concept of Activity is defined similarly in both models. These similarities account for 19 (39%) of the 49 PCROM classes. For 16 (33%) of the PCROM objects, those objects elaborate a BRIDG class by adding either a superclass or subclass. For example, PCROM maps to and elaborates the BRIDG PerformedActivity class by adding additional subclasses of CognitiveInterventions, SystemChangeInterventions, and

PhysicalInterventions. With respect to interventions, the BRIDG model focuses on the objects of a regulated clinical trial such as a pharmaceutical entity, radiotherapy, or procedure. The PCROM considers these to be subclasses of a PhysicalIntervention and thus a superclass in the BRIDG model. All such relationships are identified as types of elaborations.

For the PCROM object, ContactInformation, which describes information used to describe a means of delivering messages to individuals or organizations, BRIDG lists similar information as an attribute of several classes including Person and Organization. Although this represents a difference in formulations of the 2 models, both models accomplish essentially the same end by representing the necessary information to make contact with individuals or organizations.

In the formulation of the PCROM model, the developers determined that a person becomes a potential participant after it has been determined that he or she meets the eligibility criteria for the study and subsequently becomes a participant in the study on formal consent. Thus in PCROM, a Participant is a generalization of a PotentialParticipant via consent. The BRIDG model seems to reverse this relationship in that it defines a Participant as an individual who participates in a clinical trial and is associated via participation as a StudySubject, who in turn is defined as a potential participant is a trial. This represents a conflict between the 2 models that requires further exploration.

The BRIDG distinguishes between planned and performed characteristics of a study and represents these aspects as separate collections of classes. Thus it defines a Planned-Study as a collection of activities that is described before the beginning of the study and a Performed Study as a corresponding collection of executed study activities, which by implication may be the same or somewhat different from the PlannedStudy. The PCROM follows a more software engineering modelling approach in which states of entities (Planned versus Performed) are captured by the system in the form of attribute values rather than explicitly described as separate concepts.

There is one component of the regulated clinical research domain portrayed in the BRIDG model that is not included in this version of the PCROM. The component is comprised of the classes related to an application and its associated documents required for regulatory approval of a particular device or biopharmaceutical substance. This was not included in PCROM for the principal reason that it was considered out of scope for the first formulation of the model. The apparent thinking in developing these BRIDG concepts is consistent with the PCROM model and could well be added as an extension of the PCROM in future versions.

The BRIDG makes more frequent use of the concept of a relationship class (from the HL7 RIM Act_relationship) to describe and define the relationship between or among multiple instances of a particular class. Those classes include documents, activities, observation results, and assessments. The PCROM has placed this type of relationship at the level of an Activity with the assumption that relationships among research activities can be specified at that level and then inherited by more specific types of activities at the observation or assessment level.

The PCROM takes a much simpler view of the event flow within a study than does the BRIDG model in its current draft proposal for a study calendar. The PCROM represents these events as a timeline consisting of a collection of 1 or more activities occurring within a specified time in which where those periods are related in a chronological sequence. Such a timeline, although adhering to the requirements of the study protocol, is potentially different for each patient–intervention combination. The draft approach in BRIDG Release 1.0 takes a much more structured approach by defining the flow of events as a study calendar that is a 2-dimensional matrix of study activities described by time specifications crossed with study subjects. Study activities are further organized into epochs, arms, and elements. This

latter approach conforms more closely to the usual formulation of highly structured drug trials and the vocabulary that has developed to describe the sequencing of events within those trials. The PCROM approach may have some greater flexibility in portraying sequences of events, but at the cost of deviating from the more conventional vocabulary used in regulated clinical research. However, comparison between the 2 models does not reveal any significant conflicts, and it should be quite possible to harmonize this aspect of the 2 models.

Nine (18%) concepts were identified in the PCROM model that do not seem to correspond to the classes listed in the current version of BRIDG. Two of these are related to the way the 2 models portray adverse events. However, BRIDG Release 1.0 explicitly labels these classes as placeholders for future development. As such, a meaningful comparison of the 2 models regarding adverse effects must await completion of these classes by the BRIDG team. The PCROM concepts of TrialRegistration and EthicsApproval that represent critical steps that must precede RCT execution also do not seem to be represented as BRIDG classes. These 2 activities may be considered outside the current scope of BRIDG because it seems to take as its starting point the planned study at the point it is ready to be executed and makes no provision for steps preliminary to that point in the research design and implementation process. Similarly missing is the PCROM concept of StudyAnalysis, which is a generally acknowledged fundamental activity in clinical trial research. However statements in the BRIDG 1.0 Release Notes[38] indicate that work to be incorporated in future releases that focus on trial design may address all of these concerns.

The PCROM concept of monitoring the execution of a clinical trial to ensure that it conforms to standard operating procedures and good clinical practices is embodied in the class OperationalStandardsMonitoring. This is an important administrative aspect of conducting a clinical trial, and in particular regulated clinical trials, that does not appear as a separate class in BRIDG. Similarly the PCROM class EligibilityCriteria representing the criteria and decision rules used to determine whether a patient is eligible for a study is also not defined in BRIDG as a separate class. Explicit descriptions of eligibility criteria are fundamental to characterizing any clinical trial, including all regulated clinical trials. Investigation of the attributes of BRIDG classes does reveal that the class labeled StudyProtocol has an attribute of "monitor" with a definition that specifies a monitoring organization. Another attribute, PopulationDescription, includes a text description of inclusion and exclusion criteria. We would argue that it is more appropriate to consider both monitoring and eligibility criteria as explicit classes necessary to fully represent a clinical trial.

Conclusion

The UML methodology provides a feasible and useful mechanism and approach to creating a standard, computable description of randomized controlled trials in primary care settings. Through use cases and activity models, it promoted a thorough and comprehensive examination of primary care RCT methods and allowed us to create the PCROM, which incorporates the information and concepts

necessary to represent those types of trials. The PCROM provides a basis for characterizing and classifying these trials for several purposes. It contributes to descriptions of such trials in a standard form that can be used to determine whether it is methodologically defensible to combine the results of multiple trials for the purpose of establishing causal relationships. These standard descriptions also provide a means of classifying primary care RCTs that facilitates automated searching and promotes the creation and use of electronic databases of such trials.

The PCROM represents an important and vital link between the reference model of clinical research being defined by BRIDG and the real-world design and implementation of systems to support the design, execution, analysis, and report of clinical trials in primary care research. The avowed purpose of the BRIDG model is to create a reference model of clinical research that harmonizes the efforts undertaken by HL7, CDISC, and caBIG that is understandable to and validated by domain experts. The purpose of the PCROM is to provide a standard, computable information model of RCTs in clinical research conducted in primary care settings. That model is intended to drive the development of systems that can be used to support all aspects of clinical trial design, execution, analysis, and reporting in the complex environment of primary care. The fact that there is a high level of correspondence between the BRIDG Model Release 1.0 and the PCROM as presented here provides evidence for the validity of both models. The mapping of 82% of PCROM classes to BRIDG is evidence that the PCROM adequately represents the notion of a clinical trial as defined by the large group of domain experts who have contributed to the BRIDG model to date. The differences that do exist are primarily attributable to differing positions over whether particular components should be represented as attributes of a concept or as a distinct concept or are recognized by the BRIDG development team as areas that require further work.

Similarly, the high degree of mapping is further evidence of the validity of BRIDG as a reference model for clinical research. The fact that the PCROM was designed to drive the development and implementation of a real-world system for supporting primary care research and that the resulting model maps well to BRIDG is additional evidence for its validity. Not only does it have the endorsement of a variety of clinical research domain experts, but also BRIDG shows a high degree of correspondence with the PCROM that was designed for a different purpose in the domain of primary care research.

The proposed PCROM representation supplies a domain information framework for designing and evaluating the adequacy of clinical trials management systems in primary care. By creating a standard framework for describing these clinical trials, the model assists system developers in specifying the functions and database structures that they need to implement to manage the range of RCTs that can be expected in primary care research. It provides an essential checklist of requirements that helps to ensure that their system is reasonably complete in the functions it needs to perform and the types of information it must be capable of handling. At the same time, such a checklist can serve as an important evaluation tool for those investigators and insti-

tutions that are in the process of purchasing a vendor-supplied clinical trial management system. One can be more confident that a CTMS will be adequate to the extent that it implements the PCROM.

Finally, although some movement toward semantic interoperability has been accomplished and we believe that the PCROM represents a significant step forward in syntactic interoperability, there remains significant work to do in terms of further definition and specification, including definition of additional properties of these concepts and development methods of assigning understandable values to those concepts by creating or adopting classification and description schemes. The PCROM provides the starting point for standardized descriptions of RCTs in primary care research, but we do recognize that further work needs to be accomplished before the model can be used to achieve semantic interoperability. Specifically, the investigators will undertake an effort to harmonize the PCROM with the developing BRIDG model to resolve the areas of discrepancy reported here. Beyond that effort, we anticipate the need for an expansion to incorporate standard terminologies and values that characterize the included concepts. Such standards, although under development by groups such as CDISC and ontology projects, are still in the formative stages and not ready for full incorporation into models such as the PCROM.

References ■

1. Cherry DK, Woodwell DA, Rechtsteiner EA. National Ambulatory Medical Care Survey: 2005 Summary. Advance data from vital and health statistics; no. 387. Hyattsville, Maryland: National Center for Health Statistics, 2007.
2. Westfall JM, Mold J, Fagnan L. Practice-based research—"Blue Highways" on the NIH roadmap. JAMA 2007;297:403–6.
3. Peterson KA, Fontaine P, Speedie S. The Electronic Primary Care Research Network (ePCRN): a new era in practice-based research. J Am Board Fam Med 2006;19:93–7.
4. Neuman C, Yu T, Hartman S, Raeburn K. The Kerberos Network Authentication System (RFC4120). July 2005. Available at rfc.net/rfc4120.html. Accessed July 15, 2008.
5. The OGSA-DAI Project. Available at: www.ogsadia.org.uk. Accessed January 3, 2008.
6. Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg DAB. Issues in the registration of clinical trials. JAMA 2007;297:2112–20.
7. The Cochrane Collaboration. Available at: http://www.cochrane.org/. Accessed January 27, 2008.
8. Mayo V, Goldfarb NI, Carter C, Nash DB. Value-Based Purchasing: A Review of the Literature. New York: Commonwealth Fund, 2003.
9. Galvin RS, Delbanco S, Milstein A, Belden G. Has the Leapfrog Group had an impact on the health care market? Health Affairs 2005;24:228–33.
10. CMS Hospital Pay-for Performance Workgroup. U.S. Department of Health and Human Services Medicare Hospital Value-Based Purchasing Plan Development: Issues Paper. January 17, 2007. Washington, DC: Department of Health and Human Services.
11. Mcdonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomized controlled trials? A review of trials funded by two UK funding agencies. Trials 2006;7:9.
12. Zerhouni E. Medicine. The NIH Roadmap. Science 2003;302:63–72.

13. Warren JM, Golley RK, Collins CE, et al. Randomised controlled trials in overweight children: practicalities and realities. Int J Pediatr Obes 2007;2:73–85.

14. Kingry C, Bastien A, Booth G, et al. ACCORD study group. recruitment strategies in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Am J Cardiol 2007;99:68i–79i.

15. Fransen GA, van Marrewijk CJ, Mujakovic S, et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. BMC Med Res Methodol 2007;7:16.

16. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008 Mar 15;336:601–5. Epub Mar 3, 2008.

17. Strang WN, Cucherat M, Yzebe D, Boissel JP. Trial summary software. Comput Methods Programs Biomed 2000;61:49–60.

18. Deitzer JR, Payne PR, Starren JB Coverage of clinical trials tasks in existing ontologies. AMIA Annu Symp Proc 2006:903.

19. Zarin DA, Tse T. Medicine: moving toward transparency of clinical trials. Science 2008;319:1340–2.

20. Brandt CA, Sun K, Charpentier P, Nadkarni PM. Integration of web-based and PC-based clinical research databases. Methods Inf Med 2004;43:287–95.

21. Kuchinke W, Wiegelmann S, Verplancke P, Ohmann C. Extended cooperation in clinical studies through exchange of CDISC metadata between different study software solutions. Methods Inf Med 2006;45:441–6.

22. Komatsoulis GA, Warzel DB, Hartel FW, et al. caCORE version 3: implementation of a model driven, service-oriented architecture for semantic interoperability. J Biomed Inform 2008;41:106–23.

23. Clinical Data Interchange Standards Consortium. Available at: http://www.cdisc.org. Accessed October 6, 2007.

24. Food and Drug Administration. *Code of Federal Regulations Title 21—Food and Drugs.* Food and Drugs, 21 C.F.R. Sect. 314 (2008).

25. HL7 Version 3 Resources. Available at: http://www.hl7.org.au/HL7-V3-Resrcs.htm. Accessed October 6, 2007.

26. Getting Connected with caBIG, Clinical Trials Compatibility Framework (Brochure). Bethesda, Maryland: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services, 2007.

27. Weng C, Gennari JH, Fridsma DB. User-centered semantic harmonization: a case study. J Biomed Inform 2007;40:353–64.

28. McCray AT, Ide NC. Design and implementation of a national clinical trials registry. J Am Med Inform Assoc 2000;7:313–23.

29. Laine C, Horton R, DeAngelis CD, et al. Clinical trial registration: looking back and moving ahead. JAMA 2007;298:93–4.

30. Registration Data Set (Version 1.0), International Clinical Trials Registry Platform (ICTRP). Available at: http://www.who.int/ictrp/data_set/en/. Accessed January 28, 2008.

31. Ontology for Biomedical Investigations. Available at: http://obi.sourceforge.net/index.php. Accessed January 28, 2008.

32. Shankar RD, Martins SB, O'Conner MJ, Parrish DB, Das AK. Epoch: an ontological framework to support clinical trials management, HIKM '06. Arlington, Virginia: ACM Press, 2006.

33. Sim I, Carini S, Harris S, et al. OCRE: The Ontology of Clinical Research. AMIA Summit on Translational Bioinformatics. San Francisco, California: AMIA, 2008.

34. Arvanitis TN, Taweel A, Zhao L, et al. Supporting E-Trials Over Distributed Networks: A tool for capturing randomised control trials (RCT) eligibility criteria using the National Cancer Institute's (NCI) Enterprise Vocabulary Services (EVS). Technol Health Care 2007;15:298–9.

35. OMG Unified Modeling Language (OMG UML), Infrastructure, V2.1, November 2007. Available at: http://www.omg.org/docs/formal/07-11-04.pdf. Accessed January 28, 2008.

36. Object Management Group. Unified Modeling Language. Available at: http://www.uml.org/. Accessed October 9, 2007.

37. BRIDG 1.0. Available at: http://gforge.nci.nih.gov/frs/?group_id=342. Accessed January 27, 2008.

38. BRIDG Project Documentation and Release Notes for R1. Available at: http://gforge.nci.nih.gov/frs/download.php/2024/BRIDG_Release _1_Package.zip. Accessed January 28, 2008.