The Clinical Research Environment

4

Philip R. O. Payne

Abstract

The conduct of clinical research is a data- and information-intensive endeavor, involving a variety of stakeholders spanning a spectrum from patients to providers to private sector entities to governmental policymakers. Increasingly, the modern clinical research environment relies on the use of informatics tools and methods, in order to address such diverse and challenging needs. In this chapter, we introduce the major stakeholders, activities, and use cases for informatics tools and methods that characterize the clinical research environment. This includes an overview of the ways in which informatics-based approaches influence the design of clinical studies, ensuing clinical research workflow, and the dissemination of evidence and knowledge generated during such activities. Throughout this review, we will provide a number of exemplary linkages to core biomedical informatics challenges and opportunities and the foundational theories and frameworks underlying such issues. Finally, this chapter places the preceding review in the context of a number of national-scale initiatives and trends that seek to address such needs and requirements while advancing the frontiers of discovery science and precision medicine.

Keywords

Clinical research funding · Clinical research design · Clinical research workflow · Clinical research data management · Discovery science · Precision medicine · Real World Data · Real World Evidence · COVID-19

Learning Objectives

1. List and describe the eight general classes of processes and activities for clinical research studies and provide two examples of how informatics theory or tools can be applied within each class.
2. List the different settings where clinical research can be conducted and various actors and stakeholders. Describe how these actors interact with information and communication technology to form a sociotechnical clinical research environment.
3. Discuss the data and information management needs of various clinical research actors and describe how informatics tools are used to support the clinical research workflow and communications patterns.

P. R. O. Payne (✉)
Institute for Informatics (I2), Washington University in St. Louis School of Medicine, St. Louis, MO, USA
e-mail: prpayne@wustl.edu

R. L. Richesson et al. (eds.), *Clinical Research Informatics*, Health Informatics,
https://doi.org/10.1007/978-3-031-27173-1_4

4. List at least five barriers to common research tasks that threaten the successful completion of research studies and describe how informatics tools and approaches can help overcome these barriers.
5. Define the terms "real-world data" (RWD) and "real-world evidence" and describe how research data management and informatics activities are changing to support the generation of new clinical knowledge from RWD sources.

Overview

We describe here the clinical research environment, including an overview of common activities and processes, as well as the roles played by various stakeholders involved throughout the life cycle of clinical studies, including both interventional and observational study designs. This discussion summarizes data and information management requirements incumbent to the clinical research domain. This chapter concludes with a review of the state of knowledge concerning clinical research workflow and communication patterns as well as emergent trends in the design and conduct clinical research. In addition, the chapter includes an introduction to the relationship between clinical research and the pursuit of both discovery science and precision medicine paradigms.

This chapter is organized into three general sections describing the following:

1. The basic processes, actors, settings, and goals that serve to characterize the physical and sociotechnical clinical research environment.
2. A framework of clinical research data and information management needs.
3. The current understanding of the evolving body of research that seeks to characterize clinical research workflow and communications patterns. This understanding can be used to support the optimal design and implementation of informatics platforms for use in the clinical research environment.

Clinical Research Processes, Actors, and Goals

In the following section, we introduce the major processes, stakeholders, and goals that serve to characterize the modern clinical research environment. Taken as a whole, these components represent a complex, data- and information-intensive enterprise that involves the collaboration of numerous professionals and participants in order to satisfy a set of tightly interrelated goals and objectives. Given this complex environment and the role of informatics theories and methods in terms of addressing potential barriers to the efficient, effective, high-quality, and timely conduct of clinical research, this remains an area of intensive research interest for the biomedical informatics community [1–6].

Common Clinical Research Processes

At a high level, the processes and activities of the life cycle of a clinical research program can be divided into eight general classes, as summarized below. Of note, we will place particular emphasis in this section on describing those processes relative to the conduct of interventional clinical studies (e.g., studies where a novel treatment strategy is being evaluated for safety, efficacy, and comparative effectiveness if an alternative treatment strategy exists). However, similar processes generally apply to observational or retrospective studies, with the exception of processes related to the tracking and execution of study-related participant encounters and interventions. An example of this clinical research life cycle, its major phases, and constituent processes and activities, relative to the context of an interventional clinical trial, is illustrated in Fig. 4.1. Key processes and activities include the following.

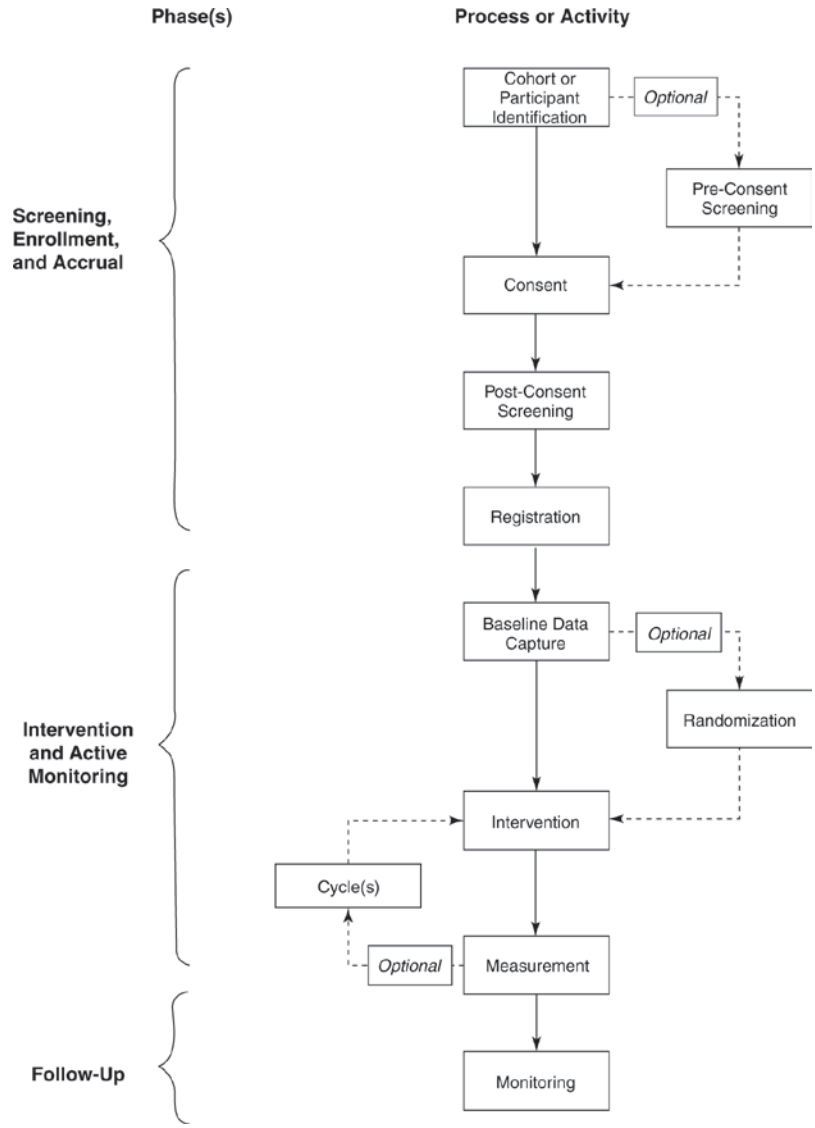

Fig. 4.1 Interventional clinical trial phases and associated execution-oriented processes

Identifying Potential Study Participants

This process usually involves either (1) the pre-encounter and/or point-of-care review of an individual's demographics and clinical phenotype in order to determine if they are potentially eligible for a given research study, given a prescribed set of eligibility criteria concerned with those same variables (also referred to as inclusion and exclusion criteria), or (2) the identification of a cohort of potential study participants from whom data can be derived, via a retrospective review of available data sources in the context of a set of defining parameters. In many cases, the data elements required for such activities are either incomplete or exist in unstructured formats, thus complicating such activities. This usually makes

it necessary for potential participants to be identified via automated methods, such as natural language processing [7], that provide a partial answer as to whether an individual is or is not eligible for a trial, which is then further explored via screening activities such as physical examinations, interviews, medical record reviews, or other similar labor-intensive mechanisms (see section "Screening and Enrolling Participants in a Clinical Study" for more details). Due to prevailing confidentiality and privacy laws and regulations, if the individual performing such eligibility screening is not directly involved in the clinical care of a potential study participant and eligibility is determined through secondary use of primarily clinical data, then the individual performing such screening must work in coordination with an individual who is involved in such clinical care in order to appropriately communicate that information to a potential study participant.

Screening and Enrolling Participants in a Clinical Study

Once a potential participant is identified, they are often subjected to evaluation, as introduced above, in order to satisfy all applicable study eligibility criteria. If they do so successfully, the participant is "enrolled" or "registered" in a study (note that both of these activities depend upon a documented informed consent process or equivalent mechanism for human subject protection). During this process, it is common for a study-specific enrollment identifier to be assigned to the participant. Of note, study staff usually maintain a set of records (often known as a "screening log") that summarize numbers of potential participants who were identified via such screening processes and how many of those individuals were successfully enrolled in a given study. Such screening logs may also include de-identified or abstracted data that details the reasons why some individuals were not successfully enrolled in a study, which can be used to help inform the pursuit of recruitment efforts for the investigation in process as well as the design of future studies.

Scheduling and Tracking Study-Related Participant Events

Once participants have been identified, screened, and enrolled in a study, they are usually scheduled for a series of encounters as defined by a study-specific calendar of events, which is also referred to as the study protocol. Sometimes, the scheduling of such events is sufficiently flexible (allowing for windows of time within which a given task or event is required to take place) that individuals may voluntarily adjust or modify their study calendar. In other cases, the temporal windows between study-related tasks or events are very strict and therefore require strict adherence by investigators and participants to the requirements defined by said calendars. Such participant- and study-specific calendars of events are tracked at multiple levels of granularity (e.g., from individual participants to large cohorts of participants enrolled in multiple studies) in order to detect individuals or studies that are "off schedule" (e.g., late or otherwise non-compliant with the required study events or activities specified in the research protocol).

Executing Study Encounters and Associated Data Collection Tasks

For each task or activity specified in a study protocol, there is almost always a corresponding study encounter (e.g., visit or phone call), during which the required study activities will be executed and the resulting data collected using either paper forms (i.e., case report forms or CRFs) or electronic data capture (EDC) instruments that replicate such CRFs in a computable format. While EDC tools are preferable for a number of reasons (e.g., quality, completeness, and auditability of data capture and management, as well as maintaining the security and confidentiality of study data) and access to computational resources has become commonplace in many study environments, there still remain large numbers of studies that are conducted using paper CRFs.

Ensuring the Quality of Study Data

Throughout a given study, study investigators and staff will usually engage in a continuous cycle of reviewing and checking the quality of study-related data. Such quality assurance (QA) usually includes reconciling the contents of CRFs or EDC instruments with the contents of supporting source documentation (e.g., electronic health records or other legally binding record-keeping instruments). These QA methods are typically time consuming and expensive, which creates a need and opportunity for information technology-based methods to improve the efficiency of data quality assessment. It is now quite common for such QA checks to be triggered via automated or semiautomated reports or "queries" regarding inconsistent or incomplete data that are generated by the study sponsor or other responsible regulatory bodies. (A more thorough characterization of data quality and quality assurance activities specific to clinical research is presented in Chapter 10).

Regulatory and Sponsor Reporting and Administrative Tracking/ Compliance

Throughout the course of a study, there are often prescribed reports concerning study enrollment, data capture, and trends in study-generated data that must be submitted to regulatory agencies, study-specific and/or institutional monitoring bodies, and/or the study sponsor. As was the case with study-encounter-related data capture, such reports can be submitted on paper or electronically. In addition, for studies regulated by government agencies (such as the FDA) or local institutional review boards (IRBs), further study-related reporting requirements must be tracked and complied with, often using proprietary or locally developed reporting instruments or tools. A primary example of such tracking/compliance is the preparation, submission, and approval of institutional review board (IRB) protocols that define how participants will be recruited and enrolled in studies and subsequently how data will be collected from them and how any physical or other risks (such as those related to security and confidentiality) are to be identified, reported, and mitigated. Additional activities included in this particular class of processes include seeking and retrieving information related to study protocols and any changes (or amendments) made to those documents throughout the course of their execution.

Budgeting and Fiscal Reconciliation

At the outset of a study, throughout its execution, and after its completion, an ongoing process of budgeting and fiscal reconciliation is conducted. The goal of these processes is to ensure the fiscal stability and performance of the study, thus making it possible to maintain necessary overhead and support structures in what is ideally a revenue or cost neutral manner.

Human Subjects Protection Reporting and Monitoring

As mentioned previously, compliance with human subject-related reporting and the monitoring of such compliance are a central part of the conduct of clinical research. This type of compliance can include obtaining IRB or equivalent approval for a study protocol and its associated practices and the execution of informed consent (a process by which potential participants are informed of the nature of a study, its risks, and benefits, in a way that allows them to weigh such factors before voluntarily engaging in a study). In addition, suspected adverse events must be collected and reported periodically to the institutional, sponsor, and regulatory organizations. The definition of "reportable" adverse events can vary by protocol, sponsor, and institution and can include local events (called internal AEs) and those occurring at other research sites (called external AEs). Similarly, actions taken in response to an AE (e.g., an amendment to a protocol reflecting changes or elimination of study procedures, adding new risks to informed con-

sent documents) must be communicated, documented, and tracked for compliance.

Common Tasks and Barriers to Successful Study Completion

According to several recent studies concerned with clinical research workflow and the tasks executed by investigators and study staff, the most common tasks performed by those individuals relative to the preceding activity areas include [2, 8–12] (1) completing paper or electronic case report forms; (2) seeking source documentation to validate the contents of such case report forms; (3) identifying, screening, and registering new study participants; and (4) responding to various reporting and monitoring requirements. In an analogous group of studies, the most common barriers encountered by investigators and study staff to the successful completion of clinical research program include [3, 11, 13, 14] (1) an inability to identify and recruit a sufficient number of study participants, (2) the attrition of participants in a study due to noncompliance with the study calendar or protocol, and (3) missing, incomplete, or insufficient high-quality data being collected such that planned study analyses cannot be performed using such data.

Clinical Research Stakeholders

As was noted previously, the clinical research environment involves the collaboration of abroad variety of stakeholders fulfilling multiple roles. Such stakeholders can be classified into six major categories, which apply across a spectrum from community practice sites to private sector sponsors to academic health centers (AHCs) and ultimately to governmental and other regulatory bodies. In the following discussion, we will briefly review the roles and activities of such actors, relative to the following six categories [3, 10, 15, 16]. *It is important to note that much of the data and information intensity of modern clinical research is a function of the need for these diverse stakeholders to interact and coordinate their activities in near real time, often in settings that span organizational, geographic, and temporal boundaries*.

Patients and Advocacy Organizations

The first and perhaps most important stakeholder in the clinical research domain is the patient, also known as a study participant, and as an extension, advocacy organizations focusing upon specific disease or health states. Study participants are the individuals who either (1) receive a study intervention or therapy or (2) from whom study-related data are collected. Participants most often engage in studies due to a combination of factors, including:

- The availability of novel therapies as a result of participation, which may provide better clinical or quality of life outcomes and that are not available via standard-of-care models.
- The exhaustion of standard-of-care options for a given disease state, thus leaving interventional clinical studies as the only viable treatment modality.
- A desire to support the advancement of the understanding of a specific uncharacterized or *under*-characterized disease or condition via an observational or natural history study or the advancement of understanding of biological processes, life sciences more generally, or public health.

Unfortunately, identifying participants who are motivated by one or more of the preceding factors and that meet appropriate demographic or clinical criteria for enrollment in a study (e.g., eligibility or inclusion/exclusion criteria) is a difficult task. In fact, in a recent report, it was found that less than 4% of the adult US population who could have participated in a clinical research study actually did so. Such low participation is a significant impediment to our collective ability to advance the state of human health and disease treatments. It is also important to note in any dis-

cussion of clinical research participants that family and friends play an equally important role as the participants themselves, providing the encouragement, information, support, and environment that may lead to or support such individual's participation in a given study [3, 16–18].

As mentioned previously, patient advocacy organizations also play a major role in clinical research, largely through a combination of (1) promoting policy and funding initiatives intended to motivate and support clinical research efforts in targeted disease states and (2) providing a medium by which potentially large cohorts of study participants may be recruited. In recent years, patient advocacy organizations have been taking increasingly active roles in shaping the agenda of the clinical research community, especially in rare and genetic diseases [6, 12, 15, 16, 19].

Academic Health Centers

Any number of sites can serve as the host for a given clinical research program, including individual physician practices, for-profit or not-for-profit clinics and hospitals, academic health centers (AHCs), colleges or universities, or community-based institutions such as schools and churches (to name a few of many examples). However, by far, the most common site for the conduct of clinical research in the United States is the AHC [3, 5, 16, 20]. During the conduct of clinical studies, AHCs or equivalent entities may take on any number or combination of the following responsibilities:

- Obtaining local regulatory and human subjects protection approval for a research study (e.g., IRB approval).
- Identifying, screening, and enrolling or registering study participants.
- Delivery of study-specific interventions.
- Collection of study-specific data.
- Required or voluntary reporting of study outcomes and adverse events.

As part of these responsibilities, study sites such as AHCs take on significant fiscal and ethical liabilities and risks related to a study's aim and objectives. Such fiscal risks are most often times shared with study sponsors, while ethical liabilities must be mitigated through the provision and maintenance of appropriate training and oversight structures for site-specific investigators or research staff.

Within an AHC, it is common for clinical studies to be motivated by a champion, who most often serves as the study investigator. Such investigators take primary responsibility for the clinical, scientific, and ethical design and conduct of a study within their immediate or otherwise defined scope of control and influence (e.g., at a site or across a network of sites in the cases of a study site and sponsor-affiliated investigator, respectively). Study investigators may be engaged in a number of study-related activities for a given clinical research program, including the following:

- Development of preclinical or other pilot data as required to support study objectives and design.
- Authoring and approval of study protocol documents.
- Securing local or broader-scale regulatory and ethical approval.
- Interactions with study participants in order to either deliver study-based interventions or collect study-related data elements.
- Analysis and reporting of study outcomes and adverse events.
- Analysis and reporting of data and knowledge generated during the course of a study (both regulatory reporting and scholarly communication, such as articles or presentations).

In addition to these activities, investigators are also responsible for overseeing the activities of research staff involved in a study and ensuring that the actions of those staff comply with applicable best practices and regulatory or ethical frameworks. In some studies, investigators may also serve as a type of study sponsor, usually when the hypotheses or interventions being evaluated are the result of the investigator's own

scientific discoveries or research questions. We refer to such studies as being "investigator initiated." Most investigator-initiated studies are of a small scale and are funded using a combination of institutional and grant-related resources [10, 14–16].

Another recurring feature of AHCs is the engagement of research staff in the conduct of studies. Such research staff can be either fully focused upon research activities or only partially focused on such efforts, depending on their organization and role. Examples of research staff members include research coordinators/associates/assistants, data managers, statisticians, nurses, allied healthcare professionals, and information technology professionals. Such individuals usually serve as investigator "extenders," performing the detailed and day-to-day work required to satisfy the range of study-related tasks and activities attributed to investigators in the preceding discussion. There are numerous professional groups and certifications for such individuals, who normally serve as the true implementers of the vast majority of clinical research projects.

Clinical or Contract Research Organizations

Clinical or contract research organizations (CROs) are agencies that administer and facilitate clinical research processes and activities, most often on a contract basis that is funded by the study sponsor. Such CROs often provide study monitoring or regulatory support (acting as a proxy for sponsors and/or regulatory bodies) as well as study-specific research staffing relative to conduct research encounters and/or manage study-related data sets. The use of CROs is most prevalent in studies involving multiple sites that must adhere to and administer a common research protocol across those sites. In this role, the CRO can ensure consistency of study processes and procedures and support participating sites, such as community-based practices, that may not nominally have the research experience or staff usually seen in AHCs.

Sponsoring Organization

Sponsoring organizations are primarily responsible for the origination and funding of clinical research programs (except in the case of investigator-initiated clinical trials, as discussed earlier). Examples of sponsors include pharmaceutical and biotechnology companies, nonprofit organizations, as well as government agencies, such as the National Institutes of Health. Sponsors may be responsible for some combination of the following tasks or activities during the clinical research life cycle:

- Conducting preclinical studies (e.g., animal models, in silico evaluations) of therapeutic interventions.
- Developing or securing therapeutic agents or devices that are appropriate for use in human subjects.
- Preparing a study protocol and informed consent documents and obtaining necessary regulatory approvals.
- Identifying and engaging sites and/or investigators to execute a trial.
- Negotiation and funding of protocol contracts, grants, or other fiscal and operational agreements as required to scope, inform, and fund a given study.
- Training investigators concerning study procedures and activities.
- Coordinating and monitoring data collection, including the performance of data quality assurance checking (often referred to as monitoring).
- Preparation and submission of required or otherwise necessary reports concerning trial activities, outcomes, and adverse events.
- Aggregation, analysis, and dissemination of study data, outcomes, and findings.

As can be surmised from the preceding exemplary list of sponsor tasks and activities, the

nature of such items is broadly variable given the type of clinical research program being executed. For example, in the case of a trial intended to evaluate a novel therapy for a specified disease state, a private sector sponsor could be responsible for all of the preceding tasks (any of which could theoretically be outsourced to a CRO). In contrast, in the case of an epidemiological study being conducted by a government agency, such a sponsor may only be engaged in a few of these types of tasks and activities (e.g., preparing a protocol, identifying and engaging sites, funding participation, and aggregating or analyzing study results or findings). Ultimately and in the vast majority of clinical research programs, the sponsor possesses the greatest fiscal or intellectual property "stake" in the design, conduct, and outcomes of a study [10, 14–16].

Federal Regulatory Agencies

Federal regulators are primarily responsible for overseeing the safety and appropriateness of clinical research programs, given applicable legal frameworks, community-accepted best practices, and other regulatory responsibilities or requirements. Examples of federally charged regulators can include institutional review boards (IRBs, who act as designated proxies for the US Department of Health and Human Services (DHHS) relative to the application and monitoring of human subjects protection laws) as well as agencies such as the Food and Drug Administration (FDA). Such regulators can be responsible for numerous tasks and activities throughout the clinical research life cycle, including the following:

- Approving clinical research studies in light of applicable legal, ethical, and best practice frameworks or requirements.
- Performing periodic audits or reviews of study data sets to ensure the safety and legality of interventions or other research activities being undertaken.
- Collecting, aggregating, and analyzing voluntary and required reports concerning the outcomes of or adverse events associated with clinical research activities.

Broadly characterized, the overriding responsibility of regulators is to ensure the safety of study participants as well as monitor the adherence of study investigators and staff with often times complex regulatory and ethical requirements that define the responsible and appropriate conduct of a given research model or approach [4, 6].

Healthcare and Clinical Research Information Systems Vendors

Software developers and vendors play a number of roles in the clinical research environment, including (1) designing, implementing, deploying, and supporting clinical trial management systems and/or research-centric data warehouses that can be used to collect, aggregate, analyze, and disseminate research-oriented data sets; (2) providing the technical mechanisms and support for the exchange of data between information systems and/or sites involved in a given clinical research program; and (3) facilitating the secondary use of primarily clinical data in support of research (e.g., developing and supporting research-centric reporting tools that can be applied against operational clinical data repositories associated with electronic health record systems) [1, 9, 11, 21, 22]. Given the ever-increasing adoption of healthcare information technology (HIT) platforms in the clinical research domain and the corresponding benefits of reduced data entry, increased data quality and study protocol compliance, and increased depth or breadth of study data sets, the role of such healthcare and clinical research information systems vendors in the clinical research setting is likely to increase at a rapid rate over the coming decades. Further, with the advent of open standards for the interoperability of data across and between such HIT platforms, entirely new modalities for the capture, integration, QA, and reporting of data relevant to the conduct of clinical research are becoming possible and helping to overcome

numerous resource barriers that may have otherwise impeded the conduct of large-scale and/or complex studies [22–24].

Other Clinical Research Actors

Additional actors who play roles in the clinical research setting include the following [10, 16]:

- Administrative managers/coordinators: Administrative managers and coordinators are often responsible for multiple aspects of regulatory or sponsor reporting, administrative tracking/compliance, budgeting and fiscal reconciliation, and human subjects protection reporting and monitoring.
- Data safety and monitoring boards (DSMBs): DSMBs are usually comprised of individuals without a direct role in a given study and who are charged with overseeing the safety and efficacy of study-related interventions. The members of a DSMB are usually empowered to halt or otherwise modify a study if such factors are not satisfied in a positive manner. A related mechanism for patient safety oversight in observational research studies is the Observational Study Monitoring Board (OSMB).

Common Clinical Research Settings

As was noted in the earlier sections of this chapter, clinical research programs are most commonly situated in AHCs. However, such institutions are not the sole environment in which clinical research occurs. In fact, as will be discussed in greater detail in section "Identifying Potential Study Participants," there are significant trends in the clinical research community toward the conduct of studies in community practice and practice-based network (e.g., organized networks of community practice sites with shared administrative coordinating processes and agents) settings as well as global-scale networks. The primary motivations for such evolution in the practice of clinical research include (1) an access to sufficiently large participant populations, particularly in rare diseases or studies requiring large-scale and diverse patient populations, (2) reduced costs or regulatory overhead, and (3) increasing access to study-related therapies in underserved or difficult to access communities or geographic environments [1, 17, 25, 26].

Common Clinical Research Goals

In a broad sense, the objectives or goals of most clinical research programs can be stratified into one or more of the design patterns summarized in Table 4.1. These patterns serve to define the intent and methodological approach of a given study or program of research.

Table 4.1 Summary of clinical research design patterns

Pattern description	Goals/objectives	Exemplary methodological approaches
Evaluation of the safety of a new or modified therapy	Establish safety of therapy as prerequisite for efficacy testing	Phase I clinical trial[a]
Evaluation of the efficacy (ability to positively effect a targeted disease state) of a new or modified therapy	Establish efficacy of therapy relative to targeted disease state as prerequisite for comparison to existing therapies	Phase II clinical trial[a]
Comparison of new of modified therapy to existing therapies	Establish benefits or equivalency of new or modified therapy relative to existing therapies	Phase III clinical trial[a]
Observation of the longitudinal effects of a new, modified, or existent therapy	Identify long-term effects of therapies and population level	Phase IV clinical trial[a]
Collection of observational data to identify clinical, behavioral, or other manifested phenomena of interest	Identify phenomena of interest that serve to inform basic science, clinical, or population-level studies and interventions	Observational study
		Ethnography
		Surveys
		Interviews RWD analyses

Table 4.1 (continued)

Pattern description	Goals/objectives	Exemplary methodological approaches
Collection of biospecimens and/or correlative clinical data	Identify and collect biospecimens and data that can support retrospective studies and/or hypothesis generation activities	Biospecimen banking
		Remnant tissue capture

[a]The gold standard for such methodological approaches is the randomized controlled trial (*RCT*)

A Framework for Data and Information Management Requirements in Clinical Research

In order to better understand the relationships between the information needs of clinical researchers and available data management and informatics tools or platforms, it is helpful to conceptualize the conduct of clinical research programs as a multiple-stage sequential model [27]. At each stage in this model, a combination of general purpose, clinical, and research-specific HIT systems may be utilized. Examples of general purpose and clinical systems that are able to support the conduct of clinical research include the following:

- Literature search tools such as the National Library of Medicine's PubMed can be used to assist in conducting the background research necessary for the preparation of protocol documents.
- Electronic health records (EHRs) can be utilized to collect clinical data on research participants in a structured form that can reduce redundant data entry.
- Data mining tools can be used in multiple capacities, including (1) determining if participant cohorts meeting the study inclusion or exclusion criteria can be practically recruited given historical trends and (2) identifying specific participants and related data within existing databases (also see Chapter 21).
- Clinical decision support systems (CDSS) can be used to alert providers at the point of care that an individual may be eligible for a clinical trial.
- Computerized physician order entry (CPOE) systems, which collect data describing the therapies delivered to research participants, can be used in both participant tracking and study analyses.

In addition to the preceding general purpose and clinical systems, research-specific IT systems have been developed that include the following:

- Simulation and visualization tools can streamline the preclinical research process (e.g., disease models) and assist in the analysis of complex data sets.
- Protocol authoring tools can allow geographically distributed authors to collaborate on complex protocol documents.
- Participant screening tools can assist in the identification and registration of research participants.
- Research-specific web portals provide researchers with a single point of access to research-specific documents and information.
- Electronic data collection or capture tools (EDC) can be used to collect research-specific data in a structured form and reduce the need for redundant and potentially error-prone paper-based data collection techniques.
- Research-specific decision support systems provide protocol-specific guidelines and alerts to researchers, for example, tracking the status of participants to ensure protocol compliance.

Clinical Research Workflow and Communications

Despite the critical role of workflow in determining both operational efficiencies and effective tactics for the deployment and adoption of information technology in the biomedical domain,

there is a paucity of literature describing systematic clinical research workflow paradigms. However, a small body of literature does provide some insight into the basic workflows engaged in or experienced by clinical research investigators and staff and associated challenges and opportunities. In the following section, we will highlight a number of salient features of such findings, in order to provide a general overview of prevailing clinical research workflow characteristics.

Workflow Challenges

There are a number of workflow challenges that serve to characterize the clinical research environment [5, 11, 16, 22], including the four broad categories of such issues as summarized below:

Paper-Based Information Management Practices

As was noted previously, a majority of clinical research tasks and activities are completed or otherwise executed using some combination of paper-based information management practices. As with all such scenarios involving the use of paper-based information management, inherent limitations associated with paper, including its ability to only be accessed by one individual at one time in one location, severely limit the scalability and flexibility of such approaches. Furthermore, in many clinical research settings, with the number of ongoing studies that regularly co-occur, the proliferation of multiple paper-based information management schemes (e.g., study charts, binders, copies of source documentation, faxes, printouts) leads to significant space and organizational challenges and inefficiencies.

Complex Technical and Communications Processes

In recent studies of clinical research workflow, it has been observed that most research staff conduct their activities and processes using a mixture of tools and methods, including the aforementioned paper-based information management schemas, as well as telephones, computers, and other electronic mediums, and interpersonal (e.g., face-to-face) communications. The combined effects of such complex combinations of tools and methods are an undesirable increase in cognitive complexity and corresponding decreases in productivity, accuracy, and efficiency, as described later in this chapter.

Interruptions

Again, as has been reported in recent studies, upward of 18% of clinical research tasks and activities are interrupted, usually by operational workflow requirements (e.g., associated with the environment in which a study is occurring, such as a hospital or clinic) or other study-related activities. Much as was the case with the preceding issues surrounding complex technical and communication processes, such interruptions significantly increase cognitive complexity, with all of the associated negative workflow and efficiency implications.

Single Point of Information Exchange

One of the most problematic workflow challenges in the clinical research environment is the fact that, in many instances, a single staff member (most often a Clinical Research Coordinator, CRC) is the single point of research-related information management and exchange. In such instances, the physical and cognitive capacities, as well as availability of such individuals, serve as a primary rate-limiting component of overall research productivity and workflow. This phenomenon is most often associated with the scarcity of individuals with the necessary training to conduct clinical research activities and/or the availability of funding and resources to support such positions.

Cognitive Complexity

As was briefly introduced in the preceding discussion, many of the characteristics of the current clinical research environment lend themselves to increased cognitive complexity. At a high level, the concept of cognitive complexity refers to scenarios in which the frequent use of multiple methods and artifacts to accomplish a given task exceeds inherent human cognitive capacities for information retention and recall. In such instances, increased errors and reduced efficiencies are usually observed. Ideally, such cognitive complexity is alleviated through the implementation or optimization of workflows and tools that minimize the need to switch between modalities and artifacts in order to accomplish a task [28, 29]. A small number of studies in the clinical research setting, including efforts focusing on clinical trial management systems and, in particular, clinical trial participant calendaring applications, have demonstrated that the use of rigorous, human-centered design principles can reduce cognitive complexity and increase the speed and accuracy of task completion in commonly occurring clinical study tasks and events (such as scheduling and/or rescheduling protocol-related events) [30]. However, the proliferation of paper-based information management and manually oriented workflows in the modern research environment, largely as a result of slow or incomplete information technology adoption, continues to preclude large-scale reengineering efforts intended to tackle the important problem of cognitive complexity.

Emergent Trends in Clinical Research

In the preceding sections of this chapter, we have outlined the basic theories and methods that serve to inform the design and conduct of clinical research programs, as well as the stakeholders and their workflow characteristics that define the domain and current state of clinical research practice. Throughout these discussions, we have described the ways in which informatics theories and methods can enable or enhance such processes and activities. Building on this background, in the following section, we will explore some of the emergent trends in clinical research that will serve to drive future innovation in healthcare, the life sciences, and the role of informatics as it relates to the research activities needed to support and enable such innovation.

Precision or Personalized Medicine

The advent of national-scale research programs focusing on precision or personalized medicine has served to draw increased attention to the critical role of data and computation in terms of pursuing some of the most complex research questions in the health and life science domains. At its most basic level, precision (or personalized) medicine involves:

> *…the tailoring of medical treatment to the individual characteristics of each patient. It does not literally mean the creation of drugs or medical devices that are unique to a patient, but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease, in the biology and/or prognosis of those diseases they may develop, or in their response to a specific treatment.* (National Academies of Medicine, 'Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease', 2011)

As can be seen from this definition, being able to achieve the vision of precision medicine requires that we establish an evidence base that can link a deep understanding of a patient's individual biomolecular, clinical, environmental, behavioral, and social phenotypes with the best available scientific evidence that may in turn inform an optimal therapeutic strategy given those characteristics. Building this knowledge base requires the design and execution of relevant clinical research programs, through which large numbers of research participants will need to be recruited to participate in studies where such data and outcomes will be collected and analyzed either retrospectively or prospectively. Doing so

introduces numerous challenges relative to the design and execution of such studies, including being able to recruit sufficient numbers of participants or finding alternative strategies for the design of studies that can overcome the need to recruit large numbers of individuals but instead focus on generating more targeted data that can quickly prove or disprove a hypothesized connection between phenotype and treatment outcomes [12, 19, 31–33]. Programs such as the "All of Us" initiative, sponsored by the US National Institute of Health (NIH), serve as prime examples of this emergent area of activity [34, 35].

Learning Healthcare Systems and Evidence Generating Medicine

In a manner that is closely aligned with the emergence of precision and personalized medicine as a national and international research priority, there is also an increasing awareness of the need to instrument the healthcare delivery environment such that every patient encounter becomes an opportunity to learn and improve the collective biomedical knowledge base. Such activities are often referred to as the creation of "learning healthcare systems" that can support or enable "evidence generating medicine." In this context, we can define a learning healthcare system as a system in which:

> *science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience.* (National Academies of Medicine, 'The Learning Healthcare System' 2015)

In such a model, we move beyond a unidirectional relationship between evidence generation (e.g., clinical research) and practice, toward a model in which there is a continuous cycle of learning that feeds data from the point of care to researchers for analysis, with ensuing knowledge products being delivered for clinical decision-making via rapid-cycle innovation [21]. Achieving this type of outcome requires a number of clinical research innovations, including (1) the creation of data capture instruments within EHRs and other clinical systems that are compatible with both standard of care and research activities (e.g., delivering sufficient data while not impeding clinical workflow), (2) the establishment of pragmatic clinical research designs that can produce empirically defensible results with incomplete or otherwise "messy" or incomplete data resulting from clinical care activities, and (3) the implementation of mechanisms for returning actionable knowledge generated via the analysis of such data to the point of care in short time frames, often via computable guidelines and/or decision support rules [8, 14, 17, 21, 22, 25, 33].

Real-World Data and Real-World Evidence (RWD and RWE)

Similarly, and in a manner that is synergistic with the two preceding themes (e.g., precision or personalized medicine and learning healthcare systems or evidence generating medicine), there is an increasing focus being placed by the life science, biotechnology, and pharmaceutical industries on the pursuit of research using real-world data (RWD) , to produce what is known as real-world evidence (RWE). In this context, RWD can be defined as those data that are produced during the course of standard-of-care activities that can be accessed and analyzed retrospectively for research purposes. Such RWE-focused research extends beyond traditional post-market surveillance of drug safety and efficacy, toward the identification of new uses for existing therapeutics, the identification of potential toxicities and adverse events associated with emergent used of medications, or the use of predictive modeling methods to anticipate the outcomes, safety, and value of therapeutic interventions. In a formal sense, RWE is the product of analyses applied to RWD, which can be defined as:

the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. RWD can come from a number of sources, for example: (1) Electronic health records (EHRs); (2) Claims and billing

activities; (3) Product and disease registries; (4) Patient-related activities in out-patient or in-home use settings; and (5) Health-monitoring devices. https://www.fda.gov/scienceresearch/specialtopics/realworldevidence/default.htmOne of the most common examples of leveraging RWD to generate RWE is the retrospective analysis of collections of disease-specific registries generated during the course of either prospective trials or observational studies [6, 13, 36]. In such instances, informaticians, data scientists, and statisticians find ways to link and integrate such data so that longitudinal or outcome-oriented hypotheses can be tested with large amounts of data within short time frames. Such study designs represent new models for defining and conducting clinical studies, particularly when the therapeutic agent of interest is already FDA approved and in widespread use or when seeking to conduct the sorts of analyses needed to establish a precision medicine knowledge base. However, it should be noted that these types of research paradigms require particular attention to the provenance, granularity, and domain coverage of source data, such as that derived from EHRs, as well as the challenges of integrating and harmonizing such data across and between a variety of sources, all of which represent open areas of informatics research and practice [37–39].

Bridging Public Health, Epidemiology, and Clinical Research

The experiences of the public health, epidemiology, clinical research, and informatics communities during the COVID-19 pandemic have illustrated emergent opportunities to bridge such domains to support and enable rapid cycle research programs. In particular, given the exigencies of the pandemic, it was realized that there was a need to accelerate both traditional research designs (e.g., evaluating novel diagnostics or therapeutics for COVID-19), as well as RWD/RWE-related efforts to ascertain the etiologic basis, trajectory, and public health impact of the novel coronavirus. These drivers resulted in important innovations across a spectrum including (but not limited to):

- The recruitment of research participants using a combination of CDSS and direct-to-patient electronic communications [40].
- The conduct of study encounters and data collection activities via telemedicine and other virtual care paradigms [40, 41].
- The phenotyping of patients and cohorts using harmonized data derived from multiple institutions and EHRs, integrated into common data models [42–44].
- The application of hypothesis-generating analytical methods, such as machine learning (ML),to identify complex patterns within study data sets and inform evidence-generation efforts [45–47].

A common thread across all of these developments as being the accelerated implementation and use of contemporary informatics and data science methods, as well as highly scalable computational infrastructures, such as those afforded through the use of cloud computing platforms, to enhance and increase the speed of data collection, analysis, as well as the dissemination of ensuing evidence, resulting from large-scale, RWD assets, all in order to respond to a public health emergency while maintaining rigorous and appropriate evidentiary standards. The existence of bias in RWD such as EHRs is well-known and will be a continued challenge for research and the application of evidence-based interventions. The informatics innovations and developments mentioned above can and should play an important role in the future of the fields of clinical research and clinical research informatics, as they illustrate ways in which the timeliness, efficiency, and impact of such efforts can be improved upon, with ensuing benefits to patients and populations. Further, improvements in data quality and efficiency of data collection can support the generation of new evidence as well as its tailored application to patients in real-world settings [48]. The development of informatics-enabled clinical research infrastructure and

environments therefore is fundamental to advancing the visions of personalized health care and learning health systems, ultimately improving population health and quality of life for patients and families.

Conclusion

As stated in the introduction to this chapter, the primary learning objectives to be addressed were associated with following three aims:

1. To describe the basic processes, activities, stakeholders, environments, and goals that serve to characterize the modern physical and sociotechnical clinical research environment.
2. To introduce a framework of clinical research information management needs.
3. To summarize the current state of an evolving body of research and knowledge that seeks to characterize clinical research workflow and communications patterns, in order to support the optimal design and implementation of informatics platforms in the clinical research environment.

We have addressed these objectives and aims by reviewing common processes, activities, stakeholders, environmental settings, and goals that characterize the contemporary clinical research environment. We have also introduced a conceptual model by which the information needs incumbent to the clinical research domain can be satisfied by a combination of general purpose and research-specific information systems. Finally, we have introduced the major workflow activities and challenges that exist in the clinical research setting, as well as emerging trends in the broad health and life sciences research domain that are helping to advance the state of clinical research design and practice. Taken as a whole, this overview should equip readers with a solid grounding by which they can place the content in the remainder of this text in context. Furthermore, this background should serve as the basis for educating clinical research informatics researchers and professionals about the basics of clinical research design and practice, thus catalyzing their acculturation to this critical and rapidly evolving domain.

References

1. Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. J Am Med Inform Assoc. 2009;16(3):316–27.
2. Embi PJ, Payne PR. Advancing methodologies in Clinical Research Informatics (CRI). J Biomed Inform. 2014;52(C):1–3.
3. Johnson SB, Farach FJ, Pelphrey K, Rozenblit L. Data management in clinical research: synthesizing stakeholder perspectives. J Biomed Inform. 2016;60:286–93.
4. Kahn MG, Weng C. Clinical research informatics: a conceptual perspective. J Am Med Inform Assoc. 2012;19(e1):e36–42.
5. Payne PR, Pressler TR, Sarkar IN, Lussier Y. People, organizational, and leadership factors impacting informatics support for clinical and translational research. BMC Med Inform Decis Mak. 2013;13(1):20.
6. Weng C, Kahn M. Clinical research informatics for big data and precision medicine. IMIA Yearb. 2016;1:211–8.
7. Idnay B, Dreisbach C, Weng C, Schnall R. A systematic review on natural language processing systems for eligibility prescreening in clinical research. J Am Med Inform Assoc. 2022;29(1):197–206.
8. Embi PJ, Kaufman SE, Payne PR. Biomedical informatics and outcomes research. Circulation. 2009;120(23):2393–9.
9. Goldenberg NA, Daniels SR, Mourani PM, Hamblin F, Stowe A, Powell S, et al. Enhanced infrastructure for optimizing the design and execution of clinical trials and longitudinal cohort studies in the era of precision medicine. J Pediatr. 2016;171:300–6.e2.
10. Prokscha S. Practical guide to clinical data management. Boca Raton, FL: CRC Press; 2011.
11. Richesson R, Horvath M, Rusincovitch S. Clinical research informatics and electronic health record data. Yearb Med Inform. 2014;9(1):215.
12. Saad ED, Paoletti X, Burzykowski T, Buyse M. Precision medicine needs randomized clinical trials. Nat Rev Clin Oncol. 2017;14(5):317–23.
13. Nelson EC, Dixon-Woods M, Batalden PB, Homa K, Van Citters AD, Morgan TS, et al. Patient focused registries can improve health, care, and science. BMJ. 2016;354:i3319.
14. Pencina MJ, Peterson ED. Moving from clinical trials to precision medicine: the role for predictive modeling. JAMA. 2016;315(16):1713–4.
15. Friedman LM, Furberg C, DeMets DL, Reboussin DM, Granger CB. Fundamentals of clinical trials. Cham: Springer; 1998.

16. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing clinical research. Philadelphia, PA: Lippincott Williams & Wilkins; 2013.
17. Brightling CE. Clinical trial research in focus: do trials prepare us to deliver precision medicine in those with severe asthma? Lancet Respir Med. 2017;5(2):92–5.
18. Browner WS. Publishing and presenting clinical research. Philadelphia, PA: Lippincott Williams & Wilkins; 2012.
19. Vicini P, Fields O, Lai E, Litwack E, Martin AM, Morgan T, et al. Precision medicine in the age of big data: the present and future role of large-scale unbiased sequencing in drug discovery and development. Clin Pharmacol Ther. 2016;99(2):198–207.
20. Korn EL, Freidlin B. Adaptive clinical trials: advantages and disadvantages of various adaptive design elements. J Natl Cancer Inst. 2017;109(6):djx013.
21. Embi PJ, Payne PR. Evidence generating medicine: redefining the research-practice relationship to complete the evidence cycle. Med Care. 2013;51:S87–91.
22. Murphy SN, Dubey A, Embi PJ, Harris PA, Richter BG, Turisco F, et al. Current state of information technologies for the clinical research enterprise across academic medical centers. Clin Transl Sci. 2012;5(3):281–4.
23. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc. 2016;23(5):899–908.
24. Mandl KD, Mandel JC, Kohane IS. Driving innovation in health systems through an apps-based information economy. Cell Syst. 2015;1(1):8–13.
25. Chambers DA, Feero WG, Khoury MJ. Convergence of implementation science, precision medicine, and the learning health care system: a new model for biomedical research. JAMA. 2016;315(18):1941–2.
26. Embi PJ. Future directions in clinical research informatics. Clinical research informatics. New York: Springer; 2012. p. 409–16.
27. Payne PR, Johnson SB, Starren JB, Tilson HH, Dowdy D. Breaking the translational barriers: the value of integrating biomedical informatics and translational research. J Investig Med. 2005;53(4):192–201.
28. Patel VL, Arocha JF, Kaufman DR. A primer on aspects of cognition for medical informatics. J Am Med Inform Assoc. 2001;8(4):324–43.
29. Zhang J, Patel VL. Distributed cognition, representation, and affordance. Pragmat Cogn. 2006;14(2):333–41.
30. Payne PR. Advancing user experience research to facilitate and enable patient-centered research: current state and future directions. eGEMs. 2013;1(1):1026.
31. Ashley EA. Towards precision medicine. Nat Rev Genet. 2016;17(9):507–22.
32. Hunter DJ. Uncertainty in the era of precision medicine. N Engl J Med. 2016;375(8):711–3.
33. Tenenbaum JD, Avillach P, Benham-Hutchins M, Breitenstein MK, Crowgey EL, Hoffman MA, et al. An informatics research agenda to support precision medicine: seven key areas. J Am Med Inform Assoc. 2016;23(4):791–5.
34. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372(9):793–5.
35. Sankar PL, Parker LS. The precision medicine initiative's all of us research program: an agenda for research on its ethical, legal, and social issues. Genet Med. 2017;19(7):743.
36. Ekins S. Pharmaceutical and biomedical project management in a changing global environment. Hoboken, NJ: Wiley; 2011.
37. Kohn MS, Topaloglu U, Kirkendall ES, Dharod A, Wells BJ, Gurcan M. Creating learning health systems and the emerging role of biomedical informatics. Learn Health Syst. 2022;6(1):e10259.
38. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. NPJ Dig Med. 2021;4(1):1–1.
39. Bastarache L, Brown JS, Cimino JJ, Dorr DA, Embi PJ, Payne PR, Wilcox AB, Weiner MG. Developing real-world evidence from real-world data: Transforming raw data into analytical datasets. Learn Health Syst. 2022;6(1):e10293.
40. Coller BS, Buse JB, Kimberly RP, Powderly WG, Zand MS. Re-engineering the clinical research enterprise in response to COVID-19: the Clinical Translational Science Award (CTSA) experience and proposed playbook for future pandemics. J Clin Transl Sci. 2021;5(1):e96.
41. Loucks TL, Tyson C, Dorr D, Garovic VD, Hill J, McSwain SD, Radovick S, Sonnenberg FA, Weis JA, Brady KT. Clinical research during the COVID-19 pandemic: the role of virtual visits and digital approaches. Journal of Clinical and Translational. Science. 2021;5(1):e102.
42. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, Payne PR, Pfaff ER, Robinson PN, Saltz JH, Spratt H. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021;28(3):427–43.
43. Moal B, Orieux A, Ferté T, Neuraz A, Brat GA, Avillach P, Bonzel CL, Cai T, Cho K, Cossin S, Griffier R. Acute respiratory distress syndrome after SARS-CoV-2 infection on young adult population: international observational federated study based on electronic health records through the 4CE consortium. PLoS One. 2023;18(1):e0266985.
44. Brandt PS, Pacheco JA, Adekkanattu P, Sholle ET, Abedian S, Stone DJ, Knaack DM, Xu J, Xu Z, Peng Y, Benda NC. Design and validation of a FHIR-based EHR-driven phenotyping toolbox. J Am Med Inform Assoc. 2022;29(9):1449–60.
45. Jing X, Patel VL, Cimino JJ, Shubrook JH, Zhou Y, Liu C, De Lacalle S. The roles of a secondary data analytic tool and experience in scientific hypothesis generation in clinical research: a study design. JMIR Res Protoc. 2022;11(7):e39414.
46. Jungkunz M, Köngeter A, Mehlis K, Winkler EC, Schickhardt C. Secondary use of clinical data in

data-gathering, non-interventional research or learning activities: definition, types, and a framework for risk assessment. J Med Internet Res. 2021;23(6):e26631.
47. Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, Freitag DF, Benoit J, Hughes MC, Khan F, Slater P. The role of machine learning in clinical research: transforming the future of evidence generation. Trials. 2021;22(1):1–5.
48. Kang T, Perotte A, Tang Y, Ta C, Weng C. UMLS-based data augmentation for natural language processing of clinical research literature. J Am Med Inform Assoc. 2021;28(4):812–23.