Guidelines for Accessing, Sharing, and Transferring Clinical Data for Research Purposes

1. Introduction

This document provides guidelines for accessing and sharing clinical data generated or held by the University of Chicago Medicine or the University of Chicago Biological Sciences Division (collectively "UCM") for secondary research purposes. UCM faculty use and share many types of data for research purposes, but it is the sharing and especially the transfer of clinical data to third parties for secondary use that tends to raise the greatest number of questions. The breadth of research utilizing clinical data, the complex regulatory environment, and rapidly changing technology means that it is difficult for any one document to fully anticipate all scenarios.

We thus articulate general principles and guidance here to cover the most common cases, especially those instances when clinical data are shared with entities outside of UCM for secondary research purposes. This guidance is designed to be utilized by the Clinical Research Data Stewardship Committee and other relevant committees and leaders in considering the appropriateness of and mechanisms by which clinical data use, sharing and transfer for research purposes is performed.

Foremost in these considerations is the institutional obligation to patients who entrust their clinical care to UCM and have an expectation of privacy as well as an expectation that any research use of their data is for the generation of knowledge and the common good. Federal and state regulations require certain reviews and controls in connection with some data sharing arrangements (e.g., human subjects research regulations, HIPAA) and UCM also adheres to practices and procedures (e.g., security reviews, entry into data use agreements) that help ensure UCM honors these expectations.

Nevertheless, regulations may not keep pace with the rapid development of new technologies, and even if all regulatory criteria are met, UCM should consider additional factors to ensure proper stewardship and sharing of patient data.

In addition to meeting patient expectations, UCM should also consider institutional reputational, research mission, and business risks, especially for projects the Institutional Review Board (IRB) has determined do not meet criteria for human subject research OR for which the IRB has waived the requirement for individual consent. Under these circumstances it is possible that the patients involved or the public in general would find the data use or sharing inappropriate, even if legal, thus markedly impacting the trust that patients seeking treatment need, and/or the reputation of the faculty and their research.

There is therefore a need to have overarching guidelines for sharing and transferring biomedical data for secondary research purposes, which we base upon the following principles:

a) **Right to be informed.** While the federal rules for the protection of human subjects require researchers to notify participants in certain studies if commercial use of their data is possible, these regulations do not apply to all studies and also do not apply to uses of deidentified data. UCM recognizes that there may be other situations in which it is ethically appropriate to inform participants that data may be accessed or shared with other entities and that, in some cases, data may be used for commercial or financial benefit.

b) **Commitment to not selling patient data.** While it is appropriate to account for costs associated with data aggregation, de-identification, and faculty as well as staff research effort, clinical data as well as any associated clinical interpretation will not be sold to any third party.

c) **Compliance with relevant regulatory and legal requirements.** There are a large number of local, national and international requirements for accessing, sharing and transfer of clinical data. It is the expectation that any data use or sharing is fully compliant with the relevant regulatory requirements, including whether consent needs to be obtained.

d) **Balancing support of the research mission with the privacy-related and institutional reputational risks.** It is necessary to find an appropriate balance between supporting faculty in pursuit of their independent research, which can lead to important discoveries benefitting the larger community, with risks to privacy and institutional reputational and business risks.

e) **Responsibility for, and control of, data.** As the trusted steward of its patient data, as between UCM and individual investigators, data derived from patient records, whether identifiable or deidentified, belong to UCM. In the event of a disagreement between UCM and any investigator, UCM shall make any determinations regarding the proper treatment or use of any such data.

f) **Patient right to their own clinical data** Patients right to their own clinical data is recognized and will not be limited.

Of course, there are many other principles associated with ethical research, but here we are focused on those relevant to the use and sharing of clinical data and especially large biomedical datasets.

There are several important dimensions that can be used for reducing institutional risk when accessing or sharing biomedical datasets.

a) **Who is using the data.** Use of data by UCM faculty and their research teams to further their own research interests, especially in controlled environments and in a regulatory compliant manner is encouraged and generally less problematic than sharing of data with external entities

b) **Who controls the data: UCM or a third party?** It is important to distinguish between data transfers (data is transferred to a third party) vs sharing via in-place queries in which the data remains within UCM controlled systems. In the latter scenario, UCM controls who accesses the data, what data is accessed, the access mechanism, and the results or derived data that is then transferred.

c) **From where is the data shared?** For data that leaves UCM it is important to consider in which UCM system the data currently resides and the available controls to assure that only data specified in a protocol and agreement is shared.

d) **What is the relative size of the data?** The larger the number of patients affected, the greater the chance that certain attempts to de-identify patients will succeed on at least some of the patients. Large-scale access and larger datasets can also increase institutional reputational risks, even for participants who may have previously consented to broad reuse.

e) **With whom is the data being shared?** As an academic medical center, UCM's goals include promoting greater understanding of disease, improving the delivery of clinical care, and advancing public health. These goals can be facilitated through relationships with external entities, including other academic medical centers, universities, not-for-profit foundations, government agencies, and commercial entities. In any data sharing arrangements, UCM should consider the nature of the recipient, but relationships with commercial entities require special scrutiny, especially if any of the involved UCM individuals have a financial interest in the applicable entity.

f) **What is the nature of the data and the proposed analysis?** Is the data deidentified, a limited data set, fully identifiable, or does the data include particularly sensitive information that requires special handling and destruction? Certain data may be considered particularly sensitive and patients and research subjects may be less comfortable if it is widely shared or if it becomes deidentified. Depending on the context, such data may include, but is not limited to, psychotherapy notes, substance abuse data, violent trauma data, inclusion of full text clinical notes, billing data, or a data cohort that includes a significant proportion of vulnerable subjects. Data analysis for which a specific hypothesis is not being tested and that is more open-ended can also be problematic. We use the term "narrow focus" or "broad focus" for these two types of research questions.

2. Guidelines

Since the details of a specific project define the critical institutional risk versus research mission benefit considerations, we have created a Clinical Research Data Stewardship Committee that amongst other duties, has the responsibility to:

- Review and determine appropriateness of complex requests to use or share clinical, imaging, or related data; and
- Provide guidance to the IRB on complex Privacy and HIPAA-Security issues

The committee will not review every clinical research data use or sharing proposal, but should consider whether proposals that are potentially problematic from an institutional risk perspective are justifiable from a research mission benefit perspective. This is especially relevant for cases in which individual consent is not required or waived.

The committee can and should invite ad hoc members with specific expertise in the subject matter at hand to any review. In general, the committee's recommendations will be:

- Acceptable clinical data sharing proposal
- Unacceptable clinical data sharing proposal
- Conditional acceptance based on additional criteria, including but not limited to relevant community based input
- No consensus of acceptability for data sharing

The committee's recommendations are to be shared with the Dean, who has the authority to accept or reject any recommendations. A risk matrix incorporating the above considerations, as well as the following guidelines, will be used as the basis for when committee review should be conducted and will also guide the committee's recommendations. The general approach will be:

a) Data use, sharing or transfer conducted under individual informed consent is ***generally acceptable.*** To this end, novel approaches for obtaining patient consent can and should be considered including electronic consenting and provision of patient controlled/provided research data through, for example Fast Healthcare Interoperability Resources (FHIR) interfaces. It will be important to determine whether the proposed secondary use of data is compatible with the original consent or whether patient choice for secondary use of data will be offered through a prospective, explicit consent process.

b) Data use and sharing that is summary level, fully deidentified and aggregated is ***generally acceptable*** but the risks of reidentification should be considered. If the data contains individual-level, identifiable private information from patients or study

participants, then additional analysis will be needed to determine acceptability of data sharing.

c) Data use, sharing or transfer of small cohorts or small slices of data for narrowly focused research projects are ***generally acceptable***.

d) Data sharing with or transfer to non-UCM University entities, not-for-profit entities, especially other academic institutions, and for projects conducted in the context of external peer reviewed funding is ***generally acceptable***.

e) Whenever possible and technically feasible, in-place queries of cohorts or slices, with appropriate protections to prevent data leakage, should be considered and are ***generally acceptable***, especially for narrowly focused research questions. "In place" in this context can also refer to architectures that are cloud based but for which UCM controls the data and its access.

f) Data sharing with commercial entities with whom the PI has a conflict of interest needs to be carefully considered and may require additional COI management considerations

g) Data transfers of large cohorts or data slices for research projects with a broad focus and with a commercial entity are ***generally unacceptable***.

h) Data sharing of potentially sensitive information with commercial entities is ***generally unacceptable***.

i) Data sharing with and especially transfer to commercial entities for which the only or principle benefit is financial remuneration is ***generally unacceptable***.

j) Data sharing with entities that restrict freedom of use of the data for other purposes and/or is exclusive is ***generally unacceptable***.

Note that these guidelines are not designed to address the reasonable and appropriate IP and financial requirements for contracts with third parties.

3. Definitions

Broadly focused research – when data is shared but there is little, if any, restrictions on the research that can be done. See narrowly focused research.

Data leakage – information obtained by aggregating data from multiple queries that can be used to reidentify subjects.

Data slices – a dataset formed using specific query criteria, but not necessarily related to participants or patients. For example, all electronic medical records for the period 2015-2018.

Generally acceptable/unacceptable – these terms are not "rules" but are used to guide committee discussions focused on balancing institutional risk and research mission benefit.

In-place queries – when queries are shipped to the data, the queries are executed, and the answers returned. In-place queries do not require data transfers of datasets to third party. Of course protections must be in place so that a *sequence* of queries to an in-place dataset does not "leak" data that is contrary to the data sharing policies.

Narrowly focused research – when data is shared to answer a specific question or explore a particular hypothesis that is specified beforehand. See broadly focused research.

Secondary use research - secondary analysis of existing data, including, without limitation, medical records and data collected from previous studies that were initially collected for a purpose other than the current research purpose.

Sharing of data – when a third party has access to UCM's biomedical data

Transfer of data – when datasets are provided to third parties