Ontario Best Practices Research Initiative (OBRI)

University Health Network

Research Practice in the Digital Age:
A new model for consent and governance
in longitudinal cohort and population based
observational studies

*Address for Correspondence:

Claire Bombardier
claire.bombardier@utoronto.ca
Ontario Best Practices Research
Initiative (OBRI)

University Health Network
Toronto General Research Institute

200 Elizabeth Street
Eaton North - 13th Floor, Room 225
Toronto, Ontario M5G 2C4
Phone: 416-927-2027 x2132
Fax: 416-927-4167

www.obri.ca

Working Report 2014-10-30

Table of Contents

Acknowledgements

The following members of the Ontario Best Practices Research Initiative (OBRI) and the University Health Network contributed to the development of this report:

Allan Harold Ph.D.
Josh Leslie
Annette Wilkins
Sharin Collins
Heather Begley
Ron Heslegrave Ph.D.
Dr. Claire Bombardier

Executive Summary

Although the adoption of information technology (IT) in clinical research is well underway, the realization of its capability for transformative change has lagged in two important areas of the research process; consent and governance. What little progress there has been in developing electronic consent and governance tools has only been incremental in nature, e.g. electronic consent being no more than a literal transcription of a paper based consent form into an electronic form. The use of IT in longitudinal cohort and population based research may turn out to be the driving force to move the concepts and implementation of electronic consent and governance forward.

An analysis of consent and governance for longitudinal cohort and population based research suggests that a new model for this type of research design is needed. This report is an overview of this new model which proposes significant changes to the nature of consent and governance.

The model demonstrates:

a) how to design IT so that key aspects of privacy, confidentiality, and governance are built in,
b) how to increase the research participant's involvement in the research process through dynamic consent, and
c) how to improve governance transparency to show compliance with legislation, current guidelines, and institutional policies.

Four themes have emerged which requires further work in the following areas:

1. Facilitate productivity and efficiency of research process
2. Design and build privacy, confidentiality, and governance features as dynamic processes
3. Design research platforms to enhance protection of participant's rights and their ability to exercise those rights
4. Increase the transparency and auditability of research process

Introduction

The objective of this document is to present an overview and analysis of the consent, governance, and compliance issues and challenges faced by researchers, research participants, and research ethics boards (REBs) once research data gathering and studies are hosted on electronic platforms. We focus on longitudinal cohort and population based research because of the importance of real world research for disease prevention, surveillance, treatment and outcomes, all of which may take place outside of the clinical setting. Out of this analysis, a recommendation for a new model for longitudinal cohort and population based research hosted on an electronic platform will be made.

To illustrate how consent, governance, and compliance are impacted by a research design that makes extensive use of information technology, a longitudinal cohort or population based observational (LCPO) study design will be used as a case study.

It has been estimated that for approximately 1000 ethics submissions in one year at an established Canadian Academic Health Sciences Centre, approximately one third would be interventional clinical trials, one third retrospective chart and database reviews, and the remaining third would comprise everything else (including observational research, qualitative research, personal communications, etc). This would imply that LCPO submissions are not the dominant form of ethics submission, and that in some cases, due to lack of familiarity, REBs may apply constraints to protocols that are appropriate for RCTs and retrospective studies but aren't appropriate for LCPOs. This suggests the opportunity to inform and present options and recommendations that can be used uniformly by REBs whether they are familiar with these other types of research or not.

Randomized Controlled Trials (RCT) are considered the gold standard in research design, however, they are not the only or best design for all research questions (Cartwright, 2007; Grossman & Mackenzie, 2005). One challenge to RCT design comes from its strength in internal validity, but at the cost of external validity. That is, in the search for rigor, the scope is narrowed such that generalizability is made more difficult. As an alternate, to answer research questions not amenable to an RCT design, longitudinal cohort or population based observational studies can be considered. Examples of population based studies include: Frequency of infection in patients with rheumatoid arthritis compared with controls: a population-based study (Doran, Crowson, Pond, O'Fallon, & Gabriel, 2002), Increased unrecognized coronary heart disease and sudden deaths in rheumatoid arthritis: a population-based cohort study (Maradit-Kremers et

While a longitudinal cohort or population based study is a well-accepted research design, it has been our experience that there are several challenges in implementation. Gaining ethics approval increases in difficulty the more the study design departs from a single protocol, single question, and single site fixed term study. A common feature of LCPO studies is that they are often multi-site. This may be the result of low prevalence of a measured variable or a need for a large sample size. Multi-site studies can also be used improve the generalizability of study findings. In any case, gaining ethics approval for multi-site studies can be difficult. In their analysis of how ethics reviews for multi-centre clinical trials are performed in Canada, Enzle and Schmaltz identify two common complaints: inconsistency of decisions between REBs, and many demands for revisions to a protocol, some of which also produce inconsistencies between REBs. This is perceived to make research more difficult or discouraged, resulting in a negative public health consequence, i.e. fewer population based studies are conducted (Enzle & Schmaltz, 2005). Watson et al. recount what may be an extreme case for a multi-site observational study in Australia (Watson, Rayner, & Lumley, 2007). In attempting to mount a study across 85 hospitals, the ethics approval process took 16 months, used 26,000 sheets of paper, and required 258 copies of the ethics submission, with their conclusion that it was an inconsistent and uncoordinated process. But as a counter point, Enzle and Schmalz also point out that centralization of ethics review is not necessarily without problems of its own (Enzle & Schmaltz, 2005).

The second set of challenges relates to how the use of Information Technology or the internet introduces a new set of privacy, confidentiality, governance and compliance issues. For example, it is expected that LCPO studies will be providing data to more than one researcher in more than one study. With different researchers with differing data requirements, but potentially accessing one data base, it raises questions of how privacy and confidentiality are maintained, what governance processes and structures are put in place to ensure compliance, and indeed how to demonstrate that compliance has been maintained.

Recognizing that this document will discuss and propose significant changes to consent documents and the consenting process, how might REBs respond? In the past, health researchers have reported that they were frustrated by how REBs handled consent forms, which often reached the point of line editing with editorial comments (Burris & Moss, 2006).

This suggests that REBs may approach enforcement by way of procedural diligence and paperwork (Burris & Welsh, 2007).

The purpose of this article is not to create a laundry list of complaints, but instead to propose some ideas and possible solutions that could improve the ethics review process for LCPO studies, the operation of these studies, and the experience of the study participants.

Problem statement: Longitudinal cohort/population based observational studies (LCPO) do not fit easily into the standard single protocol, single question, and fixed endpoint study design. Forcing LCPO based studies into the standard mold creates operational issues for both researchers and REBs. Using IT to alleviate some of these issues will have consequences, some not yet identified, potentially creating additional compliance, governance, and participant rights issues.

Proposal: As research and ethics practice evolves, so does the use of IT in research. Existing practice has significant issues, and incremental change coupled with naïve IT implementations are likely to create new problems rather than solve existing ones. Therefore a new model for LCPO studies that optimizes (i.e. minimizes harm and maximizes benefit) for the researcher, participant, REB, and the public is required. The model will focus on the consent, governance, and compliance issues that arise from LCPO studies hosted on an electronic platform.

The development of the model will proceed along these steps: identify the objectives of the model; determine the source of consent, governance, and compliance issues through a discussion of the differences a clinical study and the LCPO research. Inherent to this approach is a discussion of IT and procedural options that would address the identified consent, compliance and governance issues.

Ethical research involving humans is not a fixed or static endeavor. When IT brings new capabilities to improve research it can also have a positive or negative impact on the protection of the research participant's rights. Researchers seek more efficient and productive use of resources, hence the use of technology. But that cannot come at the cost of reduced protection of participant rights. The new model should avoid unintended consequences from the use of technology. e.g., technology should not make it more difficult to track data access while at the same time making it easier to access data.

Differences and Issues

A clinical trial is likely the most common type of research design conducted by institutions that have an established REB. Longitudinal, cohort or population observational research, while common, is often forced to fit into the paradigm of the more common research designs. It is important to recognize that these differences have implications for the researcher, participant and authorizing REB. These differences have an impact on study design, privacy, confidentiality, governance and ultimately compliance.

The following are some of the implications for consent, compliance, and governance resulting from differences between the two study design types (clinical trial and LCPO):

1. Clinical trials and LCPO research are significantly different in terms of what information is determined or known at the initiation of activities. In a clinical trial key defining information such as the study PI and team, governance structure, research questions, interventions and tools for measurement and method for the determination of the study end point must be specified at the outset. Cohort/population based research on the other hand must take into account that data collection and storage are separate and distinct from data extraction, analysis, and reporting. Cohort/population based research can at the outset specify data collection and storage methodologies, with the appropriate governance structures. But the specification of researchers, research questions, data requirements and analysis methods can't be made because they will not be known until sometime in the future.

   The information provided to a study participant in the consenting process for a clinical trial is all pre-determined that is not the case for LCPO research. Specific information to describe the *research* is not known and as a result can only be communicated in general terms.

   The consent questions pertaining to the use of research data must also be non-specific reflecting a lack of knowledge of what the actual research questions will be.

   Governance in LCPO research must respond to the two different activities (data collection, and question or hypothesis driven data analysis).

2. Clinical trials often have an all or nothing consent. If the potential participant can't meet the qualifying conditions, or agree to all of the consent questions, then they are not admitted to the study. LCPO research on the other hand may have multiple questions and qualifying requirements, not all of which need to be answered in the affirmative or

fulfilled for the participant to enter the cohort or population. This is distinct from the requirements of any one study within that cohort or population. The participant may be accepted into the cohort, while being excluded from some specific studies but not others. Conceptually then a clinical trial could be said to have in effect only one question, where LCPO research may require multiple consent questions.

With the recognition that the LCPO consent document may have more than one question, leading to a categorization of the types of questions that may be posed to the participant: 1) questions pertaining to participation, and by implication agreement to interventions if any, 2) questions pertaining to initial and possibly secondary use of data, and 3) questions pertaining to re-contact and communication.

The consent document has two significant roles: providing 1) information to the participant so that they may make informed decisions, and 2) one or more questions that define the agreement or contract between the study (and PI) and the participant. This means that for each question there must be supporting information, which brings into question the usual structure of the consent document. Typically the question(s) are grouped together in one section at the end of the consent document along with space for the required signature(s). It may be more appropriate to place the questions with the information pertaining to that question, e.g., if there is a consent question pertaining to linking data gathered in a study to an administrative data base, then the question asking for that permission would be with the information about linking to an external database and not at the end of the consent document. New technologies also have the ability to provide additional information according to the needs to the participant (i.e. drill down menus for additional explanations or information for each particular concept).

If the consent document has more than one question, then it can be assumed that not all questions must be answered in the affirmative, the situation now arises that the participant will only participate in some aspects of the cohort or study. Whether it is intervention, data capture, or use of data, the platform must control the participant's actions based on the answers to the questions. The platform must also control how data is used based on the answers to the questions. The platform must therefore be responsive, and in doing so it ensures

compliance both with the agreed upon privacy protection and the participant's preferences.

3. A clinical study has a fixed end point and, once a participant is enrolled in that study, typically the only option offered is withdrawal. Since the data collection portion of a LCPO is longitudinal, a LCPO may be able to offer revision of consent.

> When considered in conjunction with point 1 above, the option for the participant to stop participation (i.e. data collection), but still have the option of allowing or disallowing use of data becomes possible, along with the option at some later date of resuming participation.

> This requires that the research platform hosting LCPO research respond immediately to changes in the consent questions by altering the participant's status and options/actions.

4. In LCPO research a participant may be in multiple studies, or even multiple cohorts all hosted on the same platform. While this may also be the case on a research platform hosting clinical studies, it is not done with the expectation that data may be used in multiple studies, or shared between studies, as is more likely the case with a LCPO research. An example of where data sharing may be appropriate is when different studies may use the same measures or questionnaires, then with participants consenting to two or more studies on the platform they may be in the situation of completing the same measures twice or more within the same window of time, which if data sharing were permitted would avoid participant fatigue.

> There is the potential that a platform designed to release data to multiple studies or share data between studies may have more privacy and confidentiality risks than a platform designed to strictly segregate each study's data. In which case, both the REB and the participant must be informed, and safeguards must be put in place to ensure privacy and confidentiality. Failure to do so may lead the REB and participants to under-assess privacy risks, and may threaten the enlightened nature of consent.

> If sharing data is contemplated, then the risks and benefits of that must be presented to the participant, and the explicit question(s) seeking agreement for sharing must be asked.

From the researcher's perspective sharing of data between cohorts must also include some method of ensuring that all of the responsible researchers agree to that sharing, and that they are able to see and confirm that any claims or assertions they may have made relating to privacy and confidentiality are maintained. The platform must provide tools so that the researchers and ultimately REBs can see and show that compliance has been maintained, and if it has failed, to what extent. The platform has to have built in governance and provide tools to show compliance.

A potential benefit to having a population on a research platform is the possibility of recruiting from that population. Which again ties into point 1 above, that the participant should be given the choice of being contacted or not about future studies. Optimally, the question of contact for the purpose of recruiting is not tied to a specific study, but to a level of consent that is at the research platform level. i.e., when a participant first starts use of the research platform, they are asked consent questions pertaining to contact and communication before being enrolled in any specific study that uses data. By extension, they would also be able to modify this permission, supporting point 2 above.

5. LCPO research platform may have multiple unrelated researchers, in multiple studies accessing the same underlying data (same cohort). A researcher may have different roles or access to different types of data in different studies.

From the researcher's perspective, the platform must have some method of controlling access to data that is study specific unless it is not required because a broad consent was obtained. See point 8 for discussion of acceptability of broad consent. Roles and permissions would have the default functionality of not allowing access unless permission was specifically given.

Roles should not be research platform wide. Having access to data in one study, that is, being a researcher on the research platform should not then give the researcher access to data in all studies. Similarly, there should not be a super user able to access all data in all studies. This suggests that policies and procedures are needed to control data access; in addition the research platform needs to be designed such that those policies and procedures are built in and enforced.

6. More personal information and personal health information is likely to be present on a platform hosting LCPO research than a single study system, even though any single study may only need access to some of this information. A study will only have access to a subset of the total data that is available. A clinical trial will usually only make available the data that is specifically required to answer its study questions. In a clinical trial, personal data (i.e. contact information) will often not be part of the research data, but stored separately. LCPO research on the other hand will often need to have access to personal information and personal health information as part of the ongoing data collection activities. In general, LCPO research will have real world data that is not for research purposes (Terry & Terry, 2011).

> The issue here is that a research platform may well store much more identifying and personal health information (PHI) than would normally be kept in a database for a clinical trial type of study. Like a clinical system, access to this information may be needed for the operation of the data collection activities, but when needed, only necessary data should be accessed. More extensive collection and storage of personal data such as contact information and socioeconomic status, along with PHI increases the severity of the risk to the participant and increases potential for breach of compliance events.

> The data gathering and consent required for the cohort may well result in information being collected that would not ordinarily be in a research platform, so the criteria for evaluation of what may be collected should be based on the needs of the cohort and not only the strict research need.

> This shows another way in which the research platform must be able to control access to data, not just by study, but also data that is not research data. For example, some studies may employ telephone interviewers who in the course of their work would need to know a phone number and name to address the participant, none of which is likely needed as research data. The consent document would need to communicate this to the participant along with the safeguards to this data.

> Having more data available makes data analysis techniques such as data mining possible. Instead of only collecting the data required to answer a specific research question, the whole of the data base could be mined for relationships that were otherwise hidden or not known. This powerful technique requires greater access to data and therefore increases the risk of de-anonymizing the

data. Testing of data sets by trained personnel is therefore required before they are released to the researcher (Hertzman, Meagher, & McGrail, 2013).

7. In a LCPO study data cleaning and maintenance are ongoing activities, since the data collection phase is ongoing.

> This suggests that there are operational reasons to access data that are different than the research needs. Access to specific records and contact information may be needed to address queries. This is not the same type of information required for operational status reports or dashboards which don't require any identifying or personal information at all.

> Governance practices and structures for cohort operation will of necessity be different than for specific studies. Reports and data access for maintenance and cleaning will need to be specified, with appropriate governance. Operational data access must therefore be specified separately, and included in the consent documents, likely at the research platform level of consent, as well as specifying operational data access in ethics submission documents.

8. In a LCPO research, contact and communication with study participants can be more extensive and ongoing, even if the participant is not participating in a study at the current time. This may take the form of recruiting for new studies, communicating status and results from studies, uses of previously collected data, security updates, returning individual research results, and incidental findings. In some studies, this type of communication may be considered an intervention and so questioned; in an observational study it may be less of an issue.

> A major benefit of a platform hosting multiple studies and or cohorts or populations is that there is a large population to recruit from. But consumer legislation constrains the contact and communication that is possible, meaning that consent questions pertaining to contact and communication must be asked and permission gained. Potential breech of compliance events are possible where contact or communication is made against the participant's direction.

> Communication that may not be typically approved in the context of a study may be appropriate for communication to members of a cohort. For example, results from a study that has been published and is in the public domain may be of interest to a cohort and may help retention, but the same communication may be considered an intervention in a study that includes hypothesis testing (e.g.

comparative effectiveness research which uses real world cohorts for hypothesis driven studies).

Ongoing contact with study participants has an impact on broad consent and re-consent practices. If one of the justifications for broad consent was that it was impossible or impractical to obtain consent for a new use of the research data in the future, then ongoing contact with the participant for communication purposes removes that justification. Research participants are now used to being able to control their ongoing communication with other entities such as banking institutions, universities, or in social medial.

9. Clinical trials will have a definition of end of study, whereas longitudinal cohorts or populations do not typically have an end point, or at least lack a well-defined end point. Meaning that the term of consent for participation in the cohort/population can be open ended. Longitudinal studies will ideally follow participants as long as possible to examine longer term effects, latent side effect, emerging co-morbidities, etc.

> This suggests another reason to have separate consent questions for participation in the research platform, the cohort and the specific study.

> The lack of an end point in data collection and storage raises the question of what should be the appropriate criteria that would trigger the need to re-consent a participant. Should the criteria be the simple passage of time (i.e. every 5 years) or when significant or meaningful changes are contemplated for the research platform or the cohort's data gathering, or other criteria?

10. Clinical studies, typically require a formal written signature indicating agreement to participate in the study. The requirement for a written signature comes both from convention and the potential for risk faced by the participant by participation in the study. In other fields of research where the risk to the participant is lower, such as with a mailed survey, there is not a requirement for a written signature, and in some cases implied consent is sufficient. If revisions to consent, multiple levels of consent, and re-consent are implemented as discussed above, then it becomes increasingly difficult if not impossible to manage a written signature paper-based consent system.

> The question therefore is what if any signature is required in the consent process since there is not a universal directive that can be applied to all studies?

If a signature is required, then the platform will need to implement a form of digital signature that fulfills the four requirements of authentication, authority, integrity, and non-repudiation.

- Authentication is to show that the person who signed is who they say they are. Fraud avoidance.
- Authority shows that the person signing has the ability to agree. The lack of authority may come from age, intellectual capacity, or other limiting attribute.
- Integrity ensures that the artifact specifying the agreement can be shown to not have been modified, and that it reflects what was agreed.
- Non repudiation means that the signatories are not able to deny their signature. It is in essence the compliment of authentication.

11. With respect to LCPO governance and compliance, the requirements of the data collection and storage activities have some differences from those of the research activities.

At a minimum any ethics submission at the cohort/population or research platform level must demonstrate appropriate governance, and with the use of an electronic platform, how compliance and compliance issues are addressed. For any LCPO study, protecting privacy and confidentiality and secure storage of data must include descriptions of how that is to be accomplished and validated.

It follows then that the ethics submission for a specific study need not cover the functionality put in place at the research platform and cohort/population level, but need only address the requirements of the specific study.

With an ongoing relationship, the research platform and cohort/population level governance will need to provide for questions, complaints, and requests from study participants. In addition, a review board if constituted should likely include participant representation, and possibly an ombudsman or other person with increased responsibility and powers.

The above discussion makes clear the differences between a clinical trial type of study and a LCPO, and therefore justifies a new conceptualization or model for LCPO studies.

IT options

IT plays two roles, one of enabling new functionality (which may well have unintended consequences), and a second whereby it can by careful design and implementation prevent those unintended consequences. This section will review IT options that will provide functionality in terms of consent and governance, and also prevent non-compliance events.

Options for consent

One of the most basic functions of an online system is to be able to provide an online form that the user is able to respond to and submit, at which time the system then stores the results in a database. In this way the most basic form of electronic consent takes place. An electronic form that is a direct translation of a paper based consent document captures the responses of the participant. But electronic consent has much more potential and can bring significant changes to consent content, presentation, structure and functionality. Some examples are:

IT is able to transform consent **content** with multi-media in addition to static text, and allow for finer grained consent questions. When coupled with changes in structure and functionality it is possible to formulate consent questions that permit consent revision rather than forcing outright withdrawal from a study.

Consent **presentation** can be responsive to the participants' needs, e.g. formatting for desktop monitor, tablets, and smartphones. User interface elements (buttons, font sizes, text to speech…) can be adapted for people with various disabilities.

Consent **structure** may be tiered: via a research platform level consent (agreement to be registered with the research platform), a program consent (agreement to be registered with a program or disease group), and a study specific consent (hypothesis driven protocols). The consent documents need no longer be a linear text document but may have hyperlinks, outline views, navigation aids.

IT will give the consent process **functionality** through direct and dynamic consent, digital signatures, and when coupled with tiered consent, the ability of the participant to have fine grained control over the research experience.

Electronic consent functionality requires further elaboration. Direct consent when the participants complete the consent document themselves in itself doesn't seem like much of an advance. The implications of direct consent open up the discussion to broader functionality however. One implication is that withdrawal from a study may also be done directly rather than having to submit a request or form to the study coordinator (this assumes that there are no

participant safety reasons to consider). This leads to the concept of dynamic consent, where once the request is made, the research platform then acts upon the information provided by the participant. In our example, once the participant signifies that they have withdrawn from the study, the research platform takes the appropriate actions to make that so. This may be an issue with some researchers who need to ensure they meet appropriate sample sizes etc. Studies with an inability to remove data or linkages can still exist; this condition just needs to be explicit in the consent content

Dynamic consent can also be used to implement the finer grained questions that may be required in a study. If the study permits it, a participant could now start, pause, and resume their participation in a study. For example, if the participant were asked for permission for contact and communication, they could directly turn that on and off depending on their needs i.e. block recruiting requests for a period of time, then granting permission for them at a later date. Another example is participants who paused participation in an observational study to participate in a proprietary RCT that precludes other study involvement.

Compliance Issues

As the discussion above showed, the introduction of IT increases the potential for privacy and confidentiality compliance issues. One of the ways IT can help prevent those issues is by using principles of privacy by design such as:

- **Segregation of data:** using database technology it is possible to segregate data between studies, and to have segregation of the data gathered (Personal Data Store) from that accessible by the researcher (Researcher Data Store). Segregation of data does not preclude combining or federating data across studies as part of a specific study protocol, it just does not enable it by default. A less useful segregation is between studies.
- **Encryption of data:** All or some of the data in the Personal Data Store can be encrypted preventing even database technicians from viewing that data without taking deliberate steps contrary to their terms of employment and their organizations research policies and procedures. The data would only be un-encrypted when made available to the researcher in anonymous form.

It is also possible to use technology to show that compliance has been maintained, and if not, the extent of the compliance breach by use of:

- An **Audit Log** of all data entry and data access transactions is possible. In addition it is also possible to track creation and modification of data access queries by the database technicians.

Governance processes can also be built into the research platform, such as:

- A researcher's activities within a study would be constrained by **Roles and Permissions** defined within the research platform. These roles and permissions would be study specific so that higher level permissions in one study would not carry over to another study that had lower permissions for that researcher. The creation and modification of the roles and permissions would also be an auditable activity.

Governance Options

Principles and Policies

Turning now to governance, it is worthwhile defining the relationship between consent and governance as it is used in this document. A conceptualization of the research process can take the view that the consent document and the ethics submission are artifacts of a contract the researcher has with the participant and a contract the researcher has with an REB (and their host institution) respectively. Governance can then be seen to be the processes and structures put into place to ensure that when the study is made operational that those commitments are fulfilled. In other words governance is research management (Shaw, Boynton, & Greenhalgh, 2005). Compliance is therefore the state of the study with respect to the commitments made in the consent and ethics submission documents. This of course ignores that research management is also an aid to ensure that the research study itself is carried out correctly, but that is not within the scope of this document.

It is not expected, nor is it feasible to implement all governance functionality within the research platform; a large part of it will take the form of policies, procedures, and organizational structures. One reason that it is not feasible is that it will have the same limitation as regulations have in not being able to handle all of the edge or corner cases, i.e. it is not possible to enumerate all possible conditions and the actions required. In addition from the security field is the concept that nothing is 100% secure. To even approach high levels of security incurs high cost for diminishing returns for what may be a low risk scenario. Common practice is to build into the software what is reasonable and implement policies and a process to handle changes and exceptions.

A principle to help determine the extent of privacy, confidentiality safeguards and security protection to put into place is to design the research platform to prevent accidental or casual compliance breaches. Then consider everything else as deliberate, which should be traceable and found during audit.

Data Ownership

An interesting principle that will impact the design of governance, is the question of who owns the research data. An older view holds that the researcher or institution owns the data (Culliton, 1988). Culliton further describes multiple meanings of ownership, namely control of access, responsibility for maintenance and security, and possession. In view of the "colonial behavior" of researchers Schnarch suggests that aboriginal groups take ownership, control, access and possession (OCAP) of research data (Schnarch, 2004). Taking this further is the argument for public ownership of secondary health data (Rodwin, 2009). Rosenbaum breaks the problem into parts by defining a data steward as an entity that can acquire, hold, aggregate data, with a separate method to release it for use in research (Rosenbaum, 2010). This option implies that the data ownership is maintained by the participant and that a form of dynamic consent is used to give researchers access to the data on a study by study basis (Anderson, Bragg, Hartzler, & Edwards, 2012; Rosenbaum, 2010; Terry & Terry, 2011).

For the purposes of analysis later in this document, data ownership will be categorized into the following: researcher, third party, and participant ownership. Third party ownership could include aboriginal groups and public as mentioned above, as well as disease, cohort, or advocacy groups.

Governance Entities

In addition to researchers, technical and research staff, participants, and REBs, entities such as various forms of LPCO specific governance or oversight boards can be considered. When there are multiple researchers, the responsibility of governance can move from a single PI to a review board, which would oversee the implementation and operation of an online research platform. This would require that researchers accede responsibility and authority. Membership on this board could be expanded to include the participant and ethics communities. The specific responsibilities and actions of a governance board are discussed in the next section.

Processes and procedures

A thorough analysis and discussion of the range of processes and procedures that are possible is beyond the scope of this document since the processes and procedures used will be study dependent. Rather a select number of common options will be covered.

As per CIHR privacy best practice recommendations, and following from the discussion above, it will be assumed that some form of LPCO specific governance board will be in place (Canadian Institutes of Health Research, 2005). The makeup and scope of this board will need to be consistent with the principles chosen as well as the research platform used.

The determination of ownership of data can lead to quite different options. If researcher ownership is selected then the PI directly controlling data access and use may be sufficient. If third party or participant ownership is chosen, then the governance board would need to consider policies, procedures, and processes for data acquisition, maintenance, retention-disposal, security, data access and use in their role as data stewards. Oversight and auditing would be required on an ongoing basis.

To expand on the governance boards options, consider the following scenario. The board may consider splitting data collection from data use, resulting in a participant data store/repository and researcher data store. The researcher data store could then become a data warehouse.

In this example, data moves from the participant data store to the data warehouse via a process known as Extract, Transform, and Load (ETL). The creation and execution of ETLs is typically done by technical staff, but under whose direction? One option is that an ETL process cannot be created or modified without explicit approval of the responsible body. Any activity in the ETL software should be auditable. Meaning that technical staff working on queries must have clear authority to perform the task, and it should be able to be shown that they performed only that task and what the outcome was. Therefore the governance board would receive data access or use requests, have those evaluated to determine compliance with the boards charter and privacy and confidentiality requirements derived from the consent documents permitting the gathering and use of the data. The board will also need procedures to determine that the ETL activity doesn't create any new compliance issue such as making de-anonymization of the data possible, followed by auditing and corrective action procedures. A mechanism will also be required for the actual owners of the data to approach the board, submit grievances, request information, etc..

If study specific as opposed to research platform wide roles and permissions are contemplated, then appropriate policies and procedures are required. This will include the specification of what data access and operations are permitted in each role in each study.

Auditing procedures and reporting. Because it is not assumed that the research platform will be able to enforce all aspects of governance, some level of auditing is required to ensure that the residual procedural governance procedures have been followed and to determine if there had been any compliance breeches.

The governance board may need to have a role in the management of researchers, and the resolution of disputes between researchers.

The next section will summarize the discussion and options presented above into a model for consent and governance recommendations for LCPO studies.

Discussion and Recommendations

As has been shown, a multitude of options are available when implementing an LCPO study on an electronic platform. And it is also clear that the manner in which some options are implemented may have both a benefit as well as harm. Therefore the proposed model seeks to provide a new optimal balance for all the stakeholders in the research enterprise. That is, providing improvements for all stakeholders without introducing new negative aspects to the research activity.

Key Design Principles

The first principle is to recognize that LCPO studies are comprised of two distinct activities, one data gathering, and the other *defined* research activities. The distinction is that research activities are defined within the Tri-council policy statement and relevant institutional policies, whereas data gathering until it is used in research is not. This is not to say that data gathering does not have to be done in an ethical manner or that ethics approval is not required, only that if looked at as two distinct activities, then new insights into the design of the model are possible.

The second principle is to consider that the data collected from study participants is considered to be under the control of the participant with the oversight board having possession and stewardship of that data. The oversight board will then permit access to extracts of that data for research studies. This does not answer who owns the data; it only demonstrates a design principle that aligns with the principle of a data collection activity separate from the research activities.

Proposed model for Longitudinal Cohort/Population Based Observational Studies

The model is presented in layers to show the separation or segregation of data and functionality. It is beyond the scope of this document to detail the implementation of the components of this model, since most components require significant expertise for their implementation. Only recommendations common to all LCPO studies will be made, those that are study or context specific will be left to the reader. Discussions for each of the recommendations is not made in this section, readers can refer to the appropriate sections above.

_____

Figure 1. LCPO Model
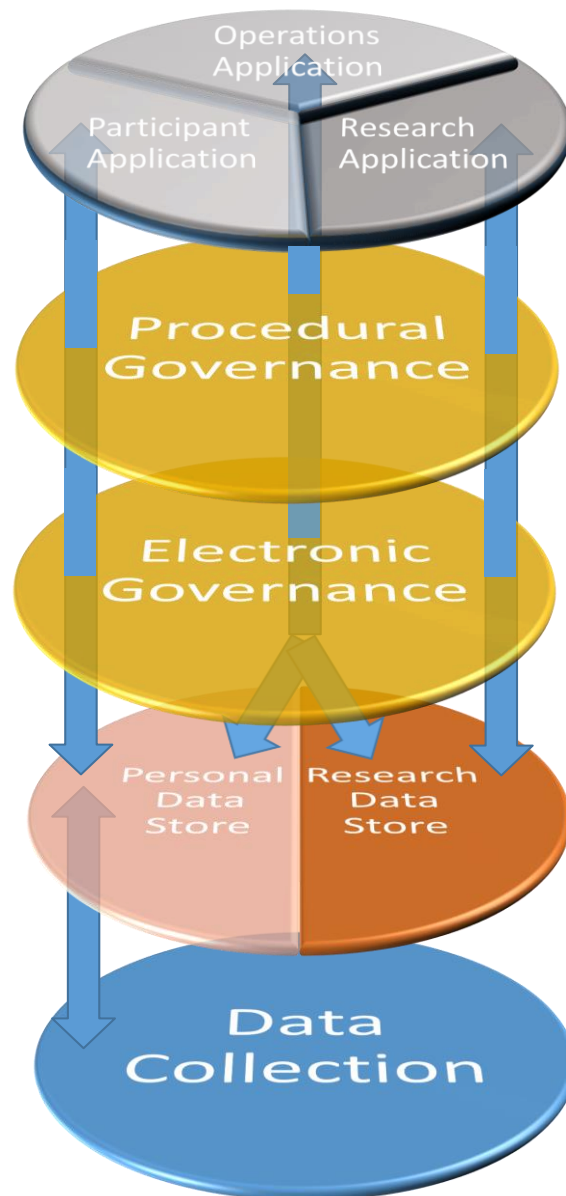
Data Collection and Data Stores Layers

At the most basic level of functionality, the role of the research platform is to collect and store data, then to subsequently make it available to researchers for analysis. Since the data collected from the participants is under their control, the repository for this data will be termed the personal data store, and the repository used by the researchers will be termed the research data store.

The two design principles lead to the recommendation that the personal data store and the researcher data store be separate and distinct repositories, with a defined process for selecting, filtering, and transforming the data, which is then loaded it into the researcher data store. As a consequence each study will have its own unique data set and will not have direct access to the underlying data in the personal data store. Data collected by whatever means can only go the Personal Data Store.

By default data will be stored as anonymously as possible. Identifying information will be delinked from the research data, along with appropriate use of data encryption.

Consent

An outcome of the consent process is the capture of a study participant's response to consent related questions, and therefore that aspect of consent is part of this layer. The logic that uses those responses to direct the behavior of the research software is housed in the application layer.

Consent will be tiered with at least two levels, one to permit data collection and the second for study specific access to the data. A third research platform level of consent maybe used to permit contact for recruiting.

Options such as revision of consent or withdrawal, will be study specific. The consent for data collection may be designed to have study specific access, or broad consent to access with opt out. In either case the participant is able to deny access to data on a study specific basis

Governance Layers

To the extent reasonable, governance will be built into the research platform software, and can be termed electronic governance. There are many aspects of governance that are not amenable to being coded into software, such as decisions that require expertise or that require discretion, these will be in the procedural level.

The **electronic governance layer** will be built upon the principles of privacy by design, which are then implemented through technology such as segregation of data, delinking identifying information, encryption, role based permissions for data access, and logging of data access.

Within the **procedural governance layer** will be an oversight board, which is required for data stewardship. The oversight board has direct responsibility of the operational applications and procedures.

Policies and procedures that will be required:

- Data gathering and storage
- Data maintenance
- Data access (who, where, time limits, etc.)
- Data destruction
- Extraction, Transformation, and Loading of the data from the personal data store to the research data store. This process is then completed with testing for the maintenance of anonymity.

Application Layer

The research platform software will implement logic such that consent is dynamic in terms of its response to the participant's response, and the interaction if any between the levels of consent.

Participant Application

Provides functionality for:

- Completing data collection instruments, review and revision if appropriate and permitted
- Review and revision of consent documents at all levels if appropriate and permitted
- Review and complete recruiting requests if appropriate

Researcher Application

Provides functionality for:

- The design of data collection instruments, including consent
- Configuring study logic
- Access study data from the researcher data store
- Implement study specific roles and permissions with logging of data access for audit purposes.

- Switch between studies and roles within studies as permitted
- Researcher, interviewer, or third party entered data

Operational Application

Provides functionality for:

- Functionality for extraction and transformation of data from the personal data store, then to load that data into researcher data store
- Testing of anonymity
- Operational data reporting, including audit reporting

Conclusion

Summary of Consent implications

Consent will be impacted in structure, content and functionality.

- Consent will not be reduced to a single question, but will have finer grained questions pertaining to participation, data collection, data use, communication and contact.
- Consent questions need not be grouped at the end of the consent document, but can be placed with the content that informs the participant.
- There will be more than one level of consent for a LCPO study that requires placing the questions at the level most appropriate.
- Consent becomes more explicitly a process that when considered along with other functionality allows for revision of consent in addition to withdrawal.
- Consent needs to be a dynamic and responsive process so that changes or revisions are reflected in the protocol operation.
- The functions of a digital signature: authenticity, authorization, integrity, and nonrepudiation must be addressed at each level of consent, but only to the extent required.
- Consent must speak to the new risks when using a shared platform hosting other studies and other researchers.
- When the participant has an ongoing relationship the consent document needs to indicate how the participant can seek information, get answers to complaints or queries, or other actions as needed.
- Dynamic consent and an ongoing relationship with the participant will help increase the participant's involvement or role in the research enterprise.

Summary of Governance Implications

- Governance structures and processes will be required for both the data gathering activity and the research activities, and for the movement of data from the one to the other.
- The manner in which ethics approval of the data gathering activity is obtained is left for future research.
- Governance processes will have to explicitly include the operational aspects of the research process, and be responsible for the actions of the operations staff.
- Qualified staff will be required to test and document that anonymity is maintained in the research data store.
- A well designed and documented research platform with built in governance and auditable logging functions will increase the transparency of the research process to participants, REB's, and the researcher's institution.
- The governance entities will need to include representation from all of the stakeholders in the research enterprise.

Further work is required to implement the recommendations within an existing research platform, Consent and Data Management (CDM) system at UHN. Results will be disseminated in a future document.

References

1. Anderson, N., Bragg, C., Hartzler, A., & Edwards, K. (2012). Participant-centric initiatives: Tools to facilitate engagement in research. Applied & Translational Genomics, 1, 25–29. doi:10.1016/j.atg.2012.07.001

2. Burris, S., & Moss, K. (2006). US health researchers review their ethics review boards: A qualitative study. Journal of Empirical Research on Human Research Ethics, 1(2), 39–58.

3. Burris, S., & Welsh, J. (2007). Regulatory paradox: A review of enforcement letters issued by the Office for Human Research Protection. Nw. UL Rev., 101, 643.

4. Canadian Institutes of Health Research. (2005). CIHR best practices for protecting privacy in health research. Canadian Institutes of Health Research.

5. Cartwright, N. (2007). Are RCTs the gold standard? BioSocieties, 2(1), 11–20.

6. Culliton, B. J. (1988). Authorship, Data Ownership. Science, 242, 658.

7. Doran, M. F., Crowson, C. S., Pond, G. R., O'Fallon, W. M., & Gabriel, S. E. (2002). Frequency of infection in patients with rheumatoid arthritis compared with controls: a population-based study. Arthritis & Rheumatism, 46(9), 2287–2293.

8. Enzle, M. E., & Schmaltz, R. (2005). Ethics review of multi-centre clinical trials in Canada. Health Law Rev, 13(2-3), 51–7.

9. Gabriel, S. E., Crowson, C. S., Kremers, H. M., Doran, M. F., Turesson, C., O'Fallon, W. M., & Matteson, E. L. (2003). Survival in rheumatoid arthritis: a population-based analysis of trends over 40 years. Arthritis & Rheumatism, 48(1), 54–58.

10. Grossman, J., & Mackenzie, F. J. (2005). The randomized controlled trial: gold standard, or merely standard? Perspectives in Biology and Medicine, 48(4), 516–534.

11. Hertzman, C. P., Meagher, N., & McGrail, K. M. (2013). Privacy by Design at Population Data BC: a case study describing the technical, administrative, and physical controls for privacy-sensitive secondary use of personal information for research in the public interest. Journal of the American Medical Informatics Association, 20(1), 25–28. doi:10.1136/amiajnl-2012-001011

12. Maradit-Kremers, H., Crowson, C. S., Nicola, P. J., Ballman, K. V., Roger, V. L., Jacobsen, S. J., & Gabriel, S. E. (2005). Increased unrecognized coronary heart disease and sudden deaths in rheumatoid arthritis: a population-based cohort study. Arthritis & Rheumatism, 52(2), 402–411.

13. Rodwin, M. (2009). THe case for public ownership of patient data. JAMA, 302(1), 86–88. doi:10.1001/jama.2009.965

14. Rosenbaum, S. (2010). Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access. Health Services Research, 45(5p2), 1442–1455. doi:10.1111/j.1475-6773.2010.01140.x

15. Schnarch, B. (2004). Ownership, control, access, and possession (OCAP) or self-determination applied to research. Journal of Aboriginal Health, 1(1), 80–95.

16. Shaw, S., Boynton, P. M., & Greenhalgh, T. (2005). Research governance: where did it come from, what does it mean? Journal of the Royal Society of Medicine, 98(11), 496–502.

17. Terry, S. F., & Terry, P. F. (2011). Power to the people: participant ownership of clinical trial data. Sci Transl Med, 3(69), 69cm3.

18. Watson, L. F., Rayner, J.-A., & Lumley, J. M. (2007). Hospital ethics approval for a population-based case–control study of very preterm birth. Australian Health Review, 31(4), 514–522.

Glossary of Terms

IT   Information Technology

LCPO   Longitudinal cohort or population based Observational research

PHI   Personal Health Information

REB   Research Ethics Board

RCT   Randomized Controlled Trial

PI   Principle Investigator

Research Platform:
Computer hardware, software, and network infrastructure designed to host research activities. May include recruiting, communication, data capture, data storage, data manipulation, data extraction and reporting. May include governance functionality such as roles and permissions, etc. All with appropriate controls for security, protection of privacy, confidentiality.

Consent and Data Management (CDM):
CDM is a software platform for Electronic Data Collection of clinical and self-reported research data. With unique functionality for longitudinal cohort/population based studies. CDM provides for the hosting of multiple studies, each with their own protocol, study population, and research personnel.