The Lumbar Imaging with Reporting of Epidemiology (LIRE) Trial- Data Sharing

Jerry Jarvik, M.D., M.P.H.
University of Washington
May 2016

UNIVERSITY of WASHINGTON

Data Sharing UH3

- *Have your research partners expressed concerns about how the data will be shared (enclave, repository, etc.)?*
- *How will individual health systems be identified in shared data sets?*
- Are there legal/regulatory obstacles to sharing your data sets?
- *How/Where will you will be sharing your results?*
- *Can the analysis be replicated using the shared data sets?*

Have your research partners expressed concerns about how the data will be shared (enclave, repository, etc.)?

- *YES, everyone's concerned*
  - *All realize data sharing is happening*
  - *Difficult to know what to be concerned about without details*
  - *Casting of system/clinic negatively concerning*
  - *One suggested approach: have rep from each site review pubs for misrepresentations/errors*

Have your research partners expressed concerns about how the data will be shared (enclave, repository, etc.)?

- *One major concern → PHI*
  - *We should not be getting it at DCC*
  - *Shouldn't ≠ Won't*

How will individual health systems be identified in shared data sets?

- *Individual health systems not explicitly identified*
- *Easy to identify 2 systems based on size and demographics*

Are there legal/regulatory obstacles to sharing your data sets?

- *Existing DUAs would likely need to be amended with details of proposed data sharing*
- *Not clear at sites who is in charge of these policies*

How/Where will you will be sharing your results?

Resources Sharing Plan: Data and Software
LIRE Project

I. Types of data

Data captured as a product of this research project will be discrete data in alphanumeric formats and summarized from each partner sites' electronic medical record (EMR). Data will consist of Limited Data Set (de-identified according to HIPAA, other than dates of service) index files and longitudinal records of healthcare utilization abstracted from each partner EMR on eligible patients. We will generate records of provider-level healthcare delivery summary data for a 12-month period prior to the randomization date as well as for a 12-month period after exposure to the LIRE intervention. In addition, we will collect provider-level data for a 24-month period after exposure to the intervention for the 75-80% of enrolled clinics for which data after exposure are available for a 24-month period. Finally, we will derive an overall provider-level measure of total and spine-related relative value units (RVUs) defined by healthcare utilized over 12 months for eligible patients.

We have defined the data dictionary for abstraction from partner EMR systems (Appendix 8).

II. Data and metadata standards

We will transfer EMR data and metadata at six month intervals from partnering institutions to the University of Washington through the PopMedNet System, and stored within a secure data environment housed within the [Department of Biostatistics]. Each of the partner sites will manage the common development of EMR data extraction algorithms based on the requirements of the LIRE project, and in coordination with the HCS Collaborative DCC. Through Quality validation processes the first cross-site data request and through verification in the second request, we will use terminology range checking and expert review to ensure that semantically and syntactically identical data is being received All data requests, schemas, and resulting data deliveries will be version controlled for a consistent provence of the data workflows. We will capture all XML schema formats, field definitiitions, and in use SAS code applied at each of the sites.

Each of the HMO Research Network (HMORN) partners will ensure that the common VDW-based data dictionary is being appropriately used. The Mayo Clinic Health System will develop a direct mapping from this schema to it's research data warehouse. We will additionally capture metadata to validate institutional-level terminology versions, and systems implementation versions. The UW DCC will manage all definitions of metadata and data dictionaries.

III. Policies for access and sharing and provisions for appropriate protection/privacy

The primary outcomes from the project will be new knowledge regarding the intervention, and best practices for implementing and managing a pragmatic clinical trial. The former will be made available through peer-reviewed publications, and de-identified data will be made available to interested parties (in XML, SAS XPORT, and R formats) through a web-based request form hosted on a study-specific secured web page. We will solicit guidance from NIH Health Care Systems Research Collaboratory as well as partner health care systems regarding the level of aggregation of the de-identified data made available to third parties. We will contribute protocols, data dictionaries, and QA models and outcomes to the HCS DCC.

For the primary data, in project years 2 through 4, only investigators affiliated with the project or sponsored by the DCC will have permission to access the data as it accumulates. In order to allow adequate time to prepare and submit manuscripts to peer-reviewed journals that address the primary and secondary aims, we will make available de-identified data to other interested parties beginning in Year 5 with a signed data use agreement. Access to data and metadata will be granted and managed through a secure password protected data portal using the University of Washington's Catalyst ShareSpaces (http://www.washington.edu/lst/web_tools/).

Resources Sharing Plan: Data and Software
LIRE Project

While software development is not a primary aim of the project, we will share all coding, patient identification algorithms, and RVU coding scripts either through published manuscripts or through open-access downloadable files on the study-specific web page.

The institutional review board at the University of Washington will manage the oversight of the data and software sharing policies in concordance with institutional review boards of the partnering health care delivery systems. Since we will not obtain patient consent as a part of the project, we will aggregate any data shared with internal or external investigators. We will generate random provider codes to prevent re-identification of individual providers by investigators at partnering institutions. We will also share data in accordance to policies set forth by the HCS Collaborative DCC.

IV. Policies and provisions for re-use and re-distribution

The data generated in this research proposal will be of interest to health services researchers, primary care providers, both partner and non-partner health care delivery institutions, and health care community stakeholders across the United States.

In Year 5 of the project, we will open a restricted-access data portal to interested parties who sign a data use agreement. The data use agreement in Year 5 aims to protect the intellectual property rights of the project investigative team by restricting external development of manuscripts using LIRE data that substantially overlap with those that are already in development by study investigators. We will form a publications committee, with investigator representatives from the University of Washington and the health care delivery partner institutions to establish manuscript development and publication guidelines. These guidelines will use as a model pre-exisitng publication guidelines for the Back pain Outcomes using Longitudinal Data project.

After Year 5, we will continue to restrict and manage data access through a centralized web-based portal. However, we will relax the data use agreement to allow open development of manuscripts provided that others will not use the data in an attempt to identify individual primary care providers or patients.

V. Plans for archiving and preservation of access

We will manage all project documentation, algorithms, and software stored in a version control system. We will physically archive all data, metadata, and code on a semi-annual basis and store it on a secured server at a separate location for documentary purposes. Institutional guidelines require that we keep project data and metadata for for 6 years after project conclusion. However, the preservation plan for LIRE will be to preserve anonymized data and metadata indefinitely. We will accomplish this through both physical media and storage on an existing secure NetApp file server that is maintained by the Center for Biomedical Statistics at the University of Washington and has automatic back-up redundancy.

How/Where will you will be sharing your results?

- Peer-reviewed publications
- De-identified data via web-based request form hosted on study-specific web page
- Since we will not obtain pt consent, will aggregate shared data, generating random provider codes to prevent re-identification.
- We will share all coding, patient identification algorithms, and RVU coding scripts

How/Where will you will be sharing your results?

- In yrs 2 through 4, only investigators affiliated with the project will have data access
- We will make available de-identified data to other parties beginning in Year 5 with a signed DUA

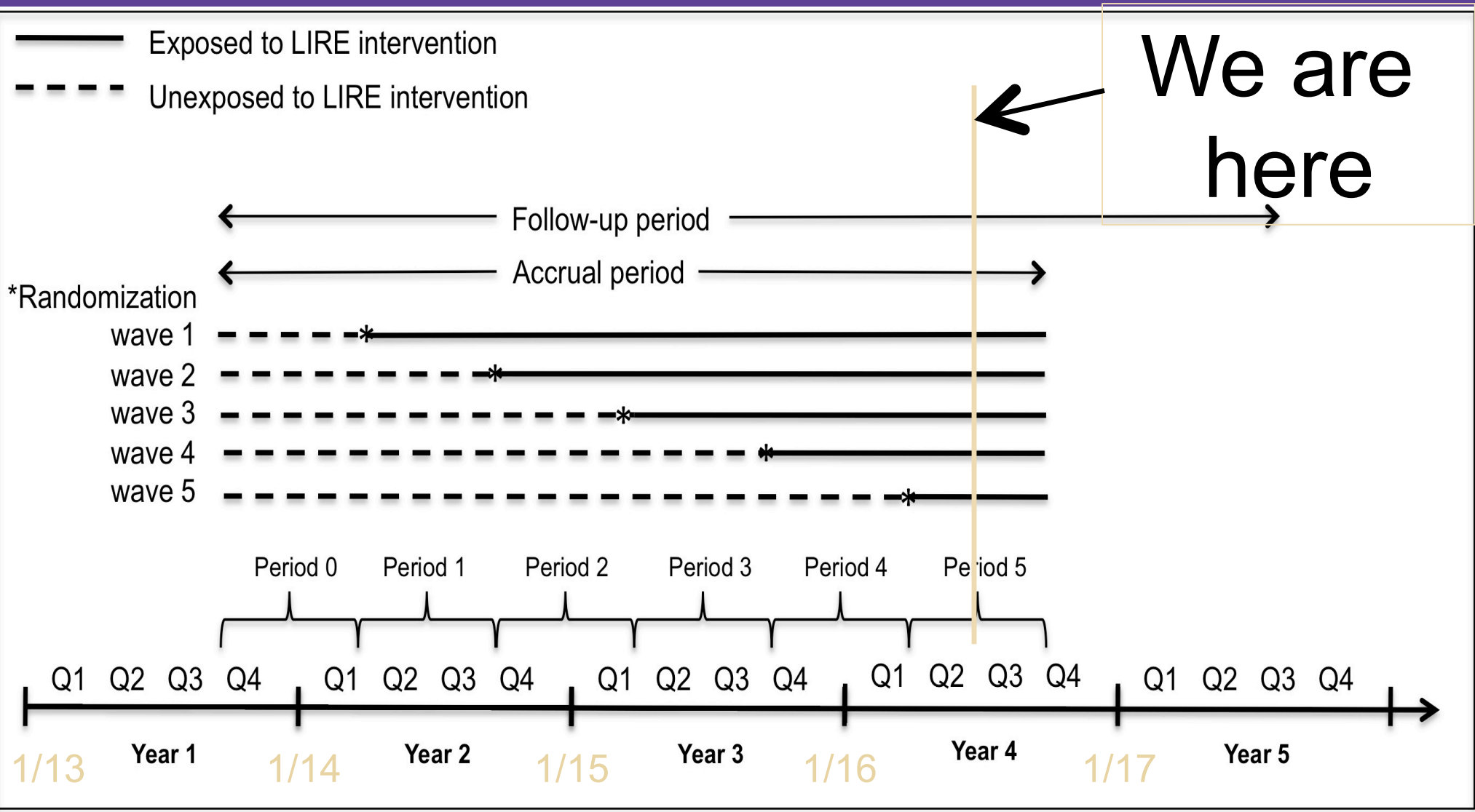
How/Where will you will be sharing your results?

- "We will solicit guidance from NIH Health Care Systems Research Collaboratory regarding the level of aggregation of the de-identified data made available to third parties."
- "Will also share data according to HCS Collaboratory policies."

Can the analysis be replicated using the shared data sets?

- Since the shared data will be aggregated, it will not be possible for 3rd parties to replicate all the analyses done by the DCC.

Timeline- On Track



Key People

- Katie James, PA-C, MPH- PD
- Brian Bresnahan, PhD- Hlth Econ
- Bryan Comstock, MS- Biostats
- Janna Friedly, MD- Rehab
- Patrick Heagerty, PhD- Biostats
- Larry Kessler, PhD- HSR
- Danielle Lavallee, Pharm D, PhD
- Zach Marcum, PhD BCPS
- Eric Meier, MS- Biostats
- Nancy Organ, MS- Biostats
- Sean Rundell, PT- Rehab
- Kari Stephens, PhD- Informatics
- Pradeep Suri, MD- Rehab
- Judy Turner, PhD- Psychol/Psych
- Rick Deyo, MD, MPH-OHSU
- Dan Cherkin, PhD-GHRI
- David Carrell, PhD- GHRI
- Brent Griffith, MD-HFHS
- Dave Nerenz, PhD- HFHS
- Dave Kallmes, MD- Mayo
- Patrick Luetmer, MD- Mayo
- Andy Avins, MD, MPH-KPNC
- Curt Langlotz, MD, PhD-Stanford