MHRN Mental Health Research Network

Outline:

- Goals
- Pre-existing requirements
- Special considerations with health system data
- Technical options
- Recommendations
- Illustrative examples from three demonstration projects

MHRN Mental Health Research Network

NIH data sharing policy & guidance:

- Data should be made as widely and freely available as possible.
- Data should be shared no later than the acceptance for publication of the main study findings.
- Initial investigators may benefit from first and continuing use of data, but not from prolonged exclusive use.

MHRN Mental Health Research Network

What HIPAA requires:

- A formal data use agreement is required if:
  - Data to be shared include any of 17 specified identifiers
  - OR other elements create more than “very small” risk of re-identification
- Investigators are responsible for determining risk of re-identification
- This has implications for data sharing method (more in a minute).

MHRN Mental Health Research Network

Special issues with health system data:

- Clear distinction between health system data and research data
  - Technical access does not equal permission to use or keep
  - Retention and use of data specifically limited by participant consent and/or specific waiver granted by IRB

- Protecting health system privacy
  - Data sharing could inadvertently disclose proprietary business information (e.g. costs or prices for specific services)
  - Data could be misused to attempt to evaluate provider or health system performance.
  - Protection from misuse is essential if healthcare systems and providers are to continue to welcome research.

MHRN Mental Health Research Network

Technical options for data sharing (in ascending order of control):

- Unsupervised data archive: Release appropriately de-identified data to any potential users
  Control of dataset contents only
- Unsupervised public data enclave: Allow any user to send any question to the data
  Control of dataset contents, query logic and return of results
- Unsupervised private data enclave: Allow specific users to send any question to the data
  Control of dataset contents, query logic, return of results, and user qualifications
- Supervised data archive: Release specific datasets to specific users
  Control of dataset contents, user qualifications and specific authorized use (e.g. DUA)
- Supervised private data enclave: Specific users may ask to send specific questions to data
  Control of dataset contents, user qualifications, query logic, return of results and topic

More control = more expense for infrastructure and governance.
(e.g. supervised means live people are involved)

MHRN Mental Health Research Network

Recommendations:

- REQUIRED: All Collaboratory trials are expected to share one or more public use datasets through an unsupervised data archive. These data should be structured to maximize future scientific value while protecting patient and health system privacy.

- OPTIONAL: Collaboratory trials may also choose to make more detailed data available through one of the more restricted options described above. Sharing additional data through one of these more restricted mechanisms is appropriate when sharing such data would have scientific or public health value but also increase risk of re-identification or other misuse.

MHRN Mental Health Research Network

Three levels of data

WHAT WE CAN TOUCH	WHAT WE CAN KEEP	WHAT WE CAN SHARE
AGE		
SEX		
VARIABLE 1		
VARIABLE 2		
VARIABLE 3		
VARIABLE 4		
VARIABLE 5		
VARIABLE 6		
VARIABLE 7		
.....		
.....		
.....		
.....		
.....		
.....		
.....		
.....		
.....		
.....		
.....		
.....		
.....		
.....		

MHRN Mental Health Research Network

What we can touch vs. What we can keep

- What data (what variables at what level of detail) are necessary to address the scientific question?
- Controlled either by language in consent process OR by specific waiver or alteration of consent

MHRN Mental Health Research Network

What we can keep vs. What we can share

- Starting position is sharing complete analytic dataset
- Restrict based on:
  - Threats to patient/participant privacy
  - Threats to health system privacy
  - ? Threats to scientific integrity
- Using a more controlled data sharing structure may reduce these threats

MHRN Mental Health Research Network

Suicide Prevention Trial: What we can touch

- For recruitment:
  - EHR data regarding PHQ9 depression questionnaires, visit types
  - Demographic data regarding age, sex, race/ethnicity
  - Membership data regarding enrollment status
- For outcome ascertainment
  - Claims data regarding outpatient, ED, and inpatient encounter diagnoses
  - Mortality files regarding date and cause of death
  - EHR data regarding telephone encounters for specific complaints
- For censoring of time at risk
  - Membership data regarding disenrollment date
  - Mortality files regarding date of death from other cause

MHRN Mental Health Research Network

Suicide Prevention Trial: What we can keep (actual analytic dataset):

- Eligibility date
- Treatment assignment
- Age (grouped), sex, race/ethnicity
- Site
- Visit type where PHQ was completed (MH specialty, primary care, other)
- Time to disenrollment from health system
- Time to death from other cause
- Time to death by suicide
- Time to first encounter with diagnosis of probable suicide attempt

MHRN Mental Health Research Network

Suicide Prevention Trial: What we can share (via unsupervised data archive):

- Eligibility date (grouped to year)
- Treatment assignment
- Age (grouped), sex, race/ethnicity
- Site (blinded)
- Visit type where PHQ was completed (MH specialty, primary care, other)
- Time to disenrollment
- Time to death from other cause
- Time to death by suicide
- Time to first encounter with diagnosis of probable suicide attempt

BUT – We can't include site and race/ethnicity in same dataset.

Questions for discussion:

- Do we concur with the recommendation that all Collaboratory trials be expected to produce and make available a public-use dataset?
- If yes, what is the appropriate deadline (relative to end of data collection or acceptance of main study findings)?
- Will the Collaboratory host a common unsupervised data archive for current and future Collaboratory trials?
- Will the Collaboratory support the infrastructure and governance for a data enclave or more supervised data archive?