Toward leveraging publicly accessible clinical trials data-sharing, dissemination, and repurposing

Sanchita Bhattacharya
Bakar Institute of Computational Health Sciences
University of California, San Francisco
ASCPT, 03.16.2019
Clinical Trial Life Cycle: When to Share Data

1. Trial Design & Registration
2. Participant Enrollment
3. Study Completion or Termination
4. Publication
5. Regulatory Application?

Key:
- Metadata
- Individual Participant Data
- Summary Data

- At trial registration
- 12 months after study completion
- 6 Months after publication*
- 18 months after study completion
- 18 months after product abandonment OR 30 days after regulatory approval**

Data Sharing Plan
Summary Level Results
Post-Publication Data Package
Full Data Package
Post-Regulatory Data Package

Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk.
http://www.iom.edu/Reports/2015/
The Landscape of Clinical Trial Data Sharing Initiatives

- Pharmaceutical Companies
- Government Health Agencies
- Non-Profit Organizations

ClinicalTrials.gov
ImmPort.org

ImmPort data portal was developed to collect and share research and clinical trials data from NIAID/DAIT funded researchers.

FAIR Principles

Make your data:
- **Findable**
- **Accessible**
- **Interoperable**
- **Reusable**

Findable
- Descriptive metadata
- Persistent

Accessible
- Determining what to share
- Participant consent and risk management
- Access status

Interoperable
- XML standards
- Data documentation initiative
- CD-ISC

Reusable
- Rights and licence models
- Permitted and non-permitted use

http://dataairport.org/
ImmPort redistributes data from major NIAID-funded programs and more

Data from 300+ trials and studies already released, involving:

- **Human Immunology Project Consortium (HIPC)**
- **Accelerating Medicines Partnership (AMP) in Rheumatoid Arthritis and Lupus (AMP)**
- **National Cancer Institute, Oncology Models Forum**
- **March of Dimes: Preterm Birth Research**
  - Immune Tolerance Network (ITN)
  - Atopic Dermatitis Research Network (ADRN)
  - Clinical Trials in Organ Transplantation (CTOT) and in Children (CTOT-C)
- **Population Genetics Analysis Program**
  - Protective Immunity for Special Populations
  - HLA Region Genomics in Immune-mediated Diseases
  - Modeling Immunity for Biodefense
  - Reagent Development for Innate Immune Receptors

- **Adjuvant Development Program**
- Immunity in Neonates and Infants Asthma and Allergic Diseases
  - Cooperative Research Centers
  - HLA and KIR Region Genomics in Immune-Mediated Diseases
- **Cooperative Study Group for Autoimmune Disease Prevention**
- Immunobiology of Xenotransplantation
- **Centers for Medical Countermeasures against Radiation Consortium**
- **Inner City Asthma Consortium**
- Systems Approach to Immunity and Inflammation
- **Innate Immune Receptors and Adjuvant Discovery Program**
- **Maintenance of Macaque Specific Pathogen-Free Breeding Colonies**
- **Non-human Primate Transplantation Tolerance Cooperative Study Group**
- **Consortium for Food Allergy Research**
- Development of Sample Sparing Assays for Monitoring Immune Responses (U24)
### Focus Areas and Data Types

<table>
<thead>
<tr>
<th>Study Area</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atopic Dermatitis</td>
<td></td>
</tr>
<tr>
<td>Transplantation</td>
<td></td>
</tr>
<tr>
<td>Infection</td>
<td></td>
</tr>
<tr>
<td>Vaccination</td>
<td></td>
</tr>
<tr>
<td>Immune</td>
<td></td>
</tr>
<tr>
<td>Influenza</td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td></td>
</tr>
<tr>
<td>Aging</td>
<td></td>
</tr>
<tr>
<td>Renal</td>
<td></td>
</tr>
<tr>
<td>Vaccination</td>
<td></td>
</tr>
<tr>
<td>HAI</td>
<td>22592</td>
</tr>
<tr>
<td>HLA Typing</td>
<td>71294</td>
</tr>
<tr>
<td>Multiplex</td>
<td>337184</td>
</tr>
<tr>
<td>ELISA Results</td>
<td>247278</td>
</tr>
<tr>
<td>ELISPOT Results</td>
<td>69709</td>
</tr>
<tr>
<td>Flow Cytometry Results</td>
<td>933214</td>
</tr>
<tr>
<td>Gene Expression Results</td>
<td>125302</td>
</tr>
<tr>
<td>Neutralizing Antibody</td>
<td>21102</td>
</tr>
</tbody>
</table>

109 registered clinical trials shared
Data Model
Leveraging of existing minimum information guidelines and data standards

CONSORT Statement
FDA and NIH ‘Demographics’ Policies
COMET Initiative
CDISC Clinical Trial Data Standards
Biomedical Controlled Vocabularies

Clinical Trial Summary
Subject Demographics
Subject Outcomes

Clinical Trial Methods
Subject Clinical Data
Data Accessibility

RImmPort package

Foundational R data model is based on existing data standards

- Study Summary (study id, title, investigators,...)
- Study Protocol (schedule of events,...)
- Subject Demographics (age, gender, ...)
- Subject Clinical Data (assessments,...)
- Subject Adverse Events

- Flow Data (flowCore)
- PCR Data (EasyqpcR)
- Elisa Data
- HLA Data
- Elispot Data
- Microarray (GEO)

Generic methods to slice and dice data across different dimensions

Custom methods to get specific types of data

sdy212 = getStudy('SDY212',...)

Shankar et al., Bioinformatics 2017
Reproducibility
Re-Analyze
Repurpose
RAVE Trial: Rituximab in ANCA-Associated Vasculitis

- 35% of patients treated with rituximab and 47% of patients treated with cyclophosphamide failed to achieve remission.

- In retrospect, do any measured factors predict response to therapies?

Stone et al., NEJM 2010
Proposed Personalized Treatment on the Basis of Granularity Index

What data is publicly available?

✓ Clinical Data from first 6 months
✓ Mechanistic Data
  • Flow Cytometry data
  • ANCA titers

(Open Arthritis Journal, Specks et al., 2011)

Validation against Published Data

**Manual Gating Results**
*NEJM, 2010 Stone et al.*

**Automated Gating Results**

<table>
<thead>
<tr>
<th>Subset</th>
<th>GI ≤ -9.25%</th>
<th>-9.25% &lt; GI &lt; 47.6%</th>
<th>GI ≥ 47.6%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40/187 (24%)</td>
<td>80/187 (49%)</td>
<td>57/187 (30%)</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>28/44 (63%)</td>
<td>41/80 (51%)</td>
<td>32/57 (56%)</td>
</tr>
<tr>
<td>Monocytes</td>
<td>24/44 (55%)</td>
<td>29/80 (36%)</td>
<td>15/57 (26%)</td>
</tr>
<tr>
<td>FSClSSC1 Granulocytes</td>
<td>20/44 (45%)</td>
<td>40/80 (50%)</td>
<td>25/57 (44%)</td>
</tr>
<tr>
<td>FSClSSC2 Granulocytes</td>
<td>20/44 (45%)</td>
<td>40/80 (50%)</td>
<td>25/57 (44%)</td>
</tr>
<tr>
<td>FSChiSSC1 Granulocytes</td>
<td>20/44 (45%)</td>
<td>40/80 (50%)</td>
<td>25/57 (44%)</td>
</tr>
<tr>
<td>FSChiSSC2 Granulocytes</td>
<td>20/44 (45%)</td>
<td>40/80 (50%)</td>
<td>25/57 (44%)</td>
</tr>
</tbody>
</table>

A 187 subjects

B Rituimab Treatment Group
C Cyclophosphamide Treatment Group

Failure Success Primary Endpoint Outcome

Failure Success Primary Endpoint Outcome

Failure Success Primary Endpoint Outcome

R = 0.0085

R = 0.037

P = 0.0139

P = 0.3782

P = 0.0002
Proposed Personalized Treatment on the Basis of Granularity Index

Reanalysis of the Rituximab in ANCA-Associated Vasculitis trial identifies granulocyte subsets as a novel early marker of successful treatment

Mazen Nasrallah1,2,3,4, Yannick Pouliot5, Bjoern Hartmann6, Patrick Dunn5, Elizabeth Thomson5, Jeffrey Wiser5 and Atul J. Butte17
Manual curation of studies, arms, and planned visits to filter for normal human subjects.

**Data available in the 10,000 Immunomes Project**

<table>
<thead>
<tr>
<th>Total Samples</th>
<th>42117</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Distinct Subjects</td>
<td>10344</td>
</tr>
</tbody>
</table>

**MEASUREMENT**

- **Secreted Proteins**: 4835
- **ELISA**: 4055
- **Multiplex ELISA**: 1286

- **Virus Titer**: 3609
  - **Virus Neutralization Titer**: 2265
  - **HAI Titer**: 1344

- **Clinical Lab Tests**:
  - **Complete Blood Count**: 2639
  - **Comprehensive Metabolic Panel**: 1684
  - **Fasting Lipid Profile**: 664

- **Questionnaire**: 1422

- **Cytometry**: 1415
  - **Flow Cytometry (PBMC)**: 907
  - **CyTOF (PBMC)**: 583
  - **Flow Cytometry (Whole Blood)**: 164

- **HLA Type**: 1093

- **Gene Expression Array**: 476
  - **Whole Blood**: 311
  - **PBMC**: 165
immTransplant: Visualizing Open-Access Living Donor Transplant Data

27 clinical trials
- Study Curation: clinical trials that excluded or have insufficient living donation information
- 20 Curated clinical trials
- Subject Curation: deceased donors, living donors with >95% missing data

11,263 Subjects
- Data Curation: Data compilation, Data standardization
- Curated Data
  1. Demographics (gender, ancestry, etc.)
  2. Pre-transplant (physical exam, serology, HLA, etc.)
  3. Intraoperative (surgical procedure, etc.)
  4. Post-transplant (physical exam, outcomes, etc.)
  5. Recipient relationship with donor

A) Post-donation outcomes
- Splenic Injury
- Vascular Injury
- Post-operative Hemorrhage
- Pneumothorax
- Complication of Intubation
- Post-operative Reintubation
- Rhabdomyolysis
- Post-operative ileus/Small Bowel Obstruction
- Post-operative Renal Failure
- Post-operative Deep Vein Thrombosis
- Post-operative Dialysis
- Wound Seroma
- Pancreatitis
- Epididymitis
- Testicular Pain
- Wound Dehiscence
- Bladder Injury
- Bowel Injury
- Hematemia
- Chronic Incisional Pain
- Incisional Hernia
- Reoperation for Incisional Hernia
- Pulmonary Embolism
- Nephrolithiasis
- Proteinuria
- Hypertension
- Diabetes Mellitus
- Microalbuminuria
- Myocardial Infarction
- Dysrhythmia
- Kidney Transplant Waiting List
- Sudden Death
- Chronic/Maintenance Dialysis
- Cardiac Arrest
- Stroke
- Kidney Transplant

B) Proportion of living donors
- Condition: Surgical, Non-Surgical

Chen et al., JAMA Network Open (in Press)
Prototype of running clinical trials in an untrustworthy environment using blockchain

Daniel R. Wong, Sanchita Bhattacharya & Atul J. Butte

UCSF
Take Home Messages

• Holistic approach to analyzing clinical trials data

• Open-access clinical trials are a valuable resource to evaluate new hypotheses, gain novel insights, and inform a better trial design

• Minimum information requirements will be the data format standard for interoperability.

• 10Kimmunomes- reference dataset generated by integrating individual level data from publicly available immunology studies

Embrace open-access clinical trials!
Acknowledgements

Atul Butte
Mark Davis
Garry Nolan

UCSF
Zicheng Hu
Kelly Zalocusky
Jieming Chen
Daniel Wong

Northrop Grumman Health Solutions
Elizabeth Thomson
Patrick Dunn
Henry Schaefer
John Campbell
Morgan Crafts

Data Providers
Mazen Nassrallah
Ravi Shankar

Funding Support
National Institute of Allergy and Infectious Diseases (NIAID)
National Institutes of Health (NIH)
Health and Human Services (HHS)
Contract #: HHSN316201200036W

@ImmPortDB
Thanks!

Sanchita.Bhattacharya@ucsf.edu

@sanchitab