Issues in Data Integration and Sharing for Novel Therapies: Lessons From an HCT Outcomes Registry
Mary M Horowitz, MD, MS
Important Issues to Address in Establishing a Mechanism for Data Integration and Sharing for Novel Therapies

- Ability to capture all patients of interest
- Ability to capture all data of interest - no matter where it is generated
- Ensuring data quality
- Maintaining long-term follow-up
- Ensuring confidentiality, security and regulatory compliance
- Making data rapidly available for multiple uses/users
Hematopoietic Stem Cell Transplantation (HCT): An Established Immune Therapy for Cancer – With a Tradition of Data Sharing

National: US, Japan, Germany, France, etc – 1980s-90s
International: Eurocord, Asian-Pacific
BMT Group; Eastern Mediterranean BMT Group – 1990s-2000s

- IBMTR Established
- EBMT Established
- NMDP Established
- NMDP & IBMTR Affiliated to form CIBMTR

Graph showing the increase in transplants from 1968 to 2013, with separate lines for autologous and allogeneic transplants.
CIBMTR 420,000 Cases Registered, up to ~10,000 variables per person (most with repeated observations, some extending over >30 years), >1000 publications
Characteristics of HCT Registry Data

• Population-based
  – Population defined by receipt of a specific type of therapy
  – No other eligibility criteria (i.e. do not need to be in a specific clinical trial or in a certain kind of center)
  – Requirement for consecutive reporting

• Longitudinal
  – No specific end date for follow-up
  – Depends on ability of treating physician/center to maintain contact

• Variable dataset
C.W. Bill Young Cell Transplantation Program*

US Department of Health and Human Services

Advisory Council on Blood Stem Cell Transplantation

HRSA/Division of Transplantation

National Cord Blood Inventory

Cord Blood Coordinating Center

Bone Marrow Coordinating Center

Stem Cell Therapeutic Outcomes Database

Office of Patient Advocacy/Single Point of Access

Components of the C. W. Bill Young Cell Transplantation Program

Individually contracted and accredited cord blood banks

Made reporting allogeneic transplant data mandatory in the US in 2005

Public Interface

Transplant centers, patients and families, referring physicians

* Created by the Stem Cell Therapeutic and Research Act of 2005 and the Stem Cell Therapeutic and Research Reauthorization Act of 2010

= HRSA Contract Organizations

= Other New Organizations or Relationships
Important Issues to Address in Establishing a Mechanism for Data Collection and Sharing for Novel Therapies

- Ability to capture all patients of interest
- Ability to capture all data of interest - no matter where it is generated
- Ensuring data quality
- Maintaining long-term follow-up
- Ensuring confidentiality, security and regulatory compliance
- Making data rapidly available for multiple users/uses
Important Issues to Address in Establishing a Mechanism for Data Collection and Sharing for Novel Therapies

• Ability to capture all patients of interest
• Ability to capture all data of interest - no matter where it is generated
• Ensuring data quality
• Maintaining long-term follow-up
• Ensuring confidentiality, security and regulatory compliance
• Making data rapidly available for multiple users/uses
Transplant Essential Data (TED) Form

Comprehensive Report Forms (CRFs)

Single CIBMTR Database (Research Database) with longitudinal follow-up

Non-US; US Auto
Voluntary

US Related*; Unrelated**
Mandatory in US (since 2007)

*Donor outcomes routinely collected
**Donor outcomes collected on subset
***Internationally harmonized dataset
Why Not Collect All Available Data on Everyone?

• **Because, in most centers, data submission is a manual process**
  – TED data are federally required
  – CRF data are not – we have to pay to defray the considerable cost of providing it
    • The number of cases for which we can get data is limited by our budget (NIH/NMDP)
• Transplant centers manually “integrate” data from multiple sources, including the EMR, graft processing and HLA labs, referring physicians, patients
  – Often doing this for multiple databases, including local research databases
AGNIS™
A Growable Network Information Service

• Initiated with an NIH grant in 2004
• Web-service based messaging system
  – AGNIS Service - Technical Description.  
    [Link](http://www.agnis.net/uploadedFiles/Documentation/docs/AGNISsService.pdf)
• Java-based server
• MySQL database repository
• Uses CDEs from caDSR/adds to caDSR
  – [Link](https://cabig.nci.nih.gov/concepts/caDSR/)
  – Curated thousands of variables in caDSR
AGNIS
Secure communication protocols/standard format

EBMT = European BMT Registry
APBMT = Asia Pacific BMT Registry
Renner R, et al. Integration of hematopoietic cell transplantation outcomes data: Data standards are not enough. Lecture Notes in Computer Science Vol. 9162 Springer Verlag, 2015. p. 139-146
 Challenges in Integration

Complex mappings
• Require in depth knowledge of clinical domain, business processes, multiple database system

Hematopoietic cell transplantation is rapidly changing
• Data standards must keep pace
• Revision of 8 AGNIS-supported forms in 2013 resulted in 801 questions added and 580 questions deleted

Cost
• ~$100,000 to map to AGNIS
• Pales in comparison to amount being spent on EHRs at same centers
• Still unattainable – low priority on the IT list
• No “meaningful use” credit
AGNIS – Some Successes

• EBMT-CIBMTR connectivity
  – TEDs submitted for 7,500 patients from >50 EBMT centers
• Other current AGNIS users
  – 5 centers make direct submissions/1 doing direct retrieval/1 both
  – 4 vendors supporting >40 centers active with production clients
• Additional users developing AGNIS submission or retrieval:
  – 3 centers and 6 vendors currently active; others in discussion or planning stages
BRIDG Model
Biomedical Research Integrated Domain Group
Domain Analysis Model

HCT Domain

Physical Model

Integration Engine

FormsNet Website
FormsNet Database

Research Database

AGNIS

NCI
National Cancer Institute

CDISC
Clinical Data Interchange Standards Consortium

FDA
Food & Drug Administration

HL7
Health Level Seven

Have interoperability outside of CIBMTR
• Can be extended to include other BRIDG classes for other needs
• Easier mapping to FDA, NCI, clinicaltrials.gov
Transplant Centers Are Not the Only Sources of Data

- HLA Typing laboratories
- Other genetic testing laboratories – commercial and research
- Cord Blood Banks
- Donor and collection centers
- NMDP captures these in a variety of ways for a subset of transplants and transfers to Data Warehouse eliminating need for manual entry by transplant center
Integrated Immunobiology Database

- Data from NMDP HLA contract labs, cord blood banks
- High-res HLA, KIR typing on stored donor-recipient samples
  - >17,500 pairs with HLA data; >10,000 pairs with KIR data
- Full HLA gene sequencing of donor/recipient pairs
  - 90 pairs completed; >300 in process
- Working committee/collaborative study testing
  - Somatic point mutation sequencing in MDS patients (N=1523)
  - T-cell epitope repertoire sequencing in HLA mismatched HCT (N=98)
  - qPCR-based telomere length assessment in HCT donors (N=2810)
  - GWAS of unrelated HCT recipients with acute leukemia (N=734)
  - Sequencing of HLA region genomic variation in HLA matched and mismatched unrelated donor HCT (N=3233)
Important Issues to Address in Establishing a Mechanism for Data Collection and Sharing for Novel Therapies

- Ability to capture all patients of interest
- Ability to capture all data of interest—no matter where it is generated
- **Ensuring data quality**
- Maintaining long-term follow-up
- Ensuring confidentiality, security and regulatory compliance
- Making data rapidly available for multiple users/uses
Data Quality

- High quality data = useful data
- Complete, timely data
  - Continuous process improvement program
  - Requirement is that 90% of forms due are submitted
  - Assessed on a trimester

- Accurate data
  - On-site audit program
  - Four year cycle
  - Plans to adopt as audit program for FACT
Deficiencies in EHR Data

- Incompleteness 24-86% - if it is not there, it can’t be captured, not even by Watson
  - e.g. someone has to note the % skin involvement for cutaneous acute GVHD, the performance score at onset of chronic GVHD, the co-morbidity index before start of conditioning
  - Not all missingness is random
- CPOE errors 51-91%
- Inaccuracies, errors 4-5%
- Inconsistencies variable

CIBMTR’s EMR User Group Working to Standardize HCT-related Data Collection at Point of Care

• Goal: facilitate ease of data collection, consistent with national data standards, and submission for use of research

• Working with Epic on tools that will use CIBMTR-defined data standards in caDSR and BRIDG
  – Tools will be part of the foundation system for Epic
  – Standard database fields in the EMR data warehouse, facilitating ease of submission to CIBMTR and others
CIBMTR’s EMR User Group Working to Standardize BMT-related Data Collection at Point of Care

• aGVHD doc flow sheet was released in EPIC’s Foundation system 2014
• BMT Smartform, essential descriptive data about the BMT, including BMT date and graft type, released in 1/2016; data elements tied to BRIDG CDEs
• Standard data elements defined on forms/worksheets through the EMR user group have been shared with OTTR/Cerner for their development purposes to benefit the HCT community
Quality Check - Have you Ever Looked at Your Own Problem List?

- Or tried to adjudicate an adverse event based on the inpatient notes? Beware the dreaded “CUT and PASTE”
  - Need to include documentation as part of education of ALL medical providers

- With better ways of sharing data electronically, perhaps all of those well-trained data professionals currently spending time in front of data entry screens can focus on making sure that providers are getting the right data entered at the right times
What About Billing/Claims Data?

Truven Health MarketScan Research Database: Claims from ~100 payers for >115 million patients
• 25% of the covered lives in the US in 2009
Goal: assess costs in year after AML diagnosis for patients age 50-64 receiving either HCT or chemotherapy
• Started with 29,915 patients with diagnosis of AML
• Only able to identify 985 (3.2%) where we were sufficiently confident of the diagnosis and treatment path to allow analysis.
Some Data Must Still Be Hunted Down by the Transplant Center

• Fragmented health care system mean that patients move from one provider/payer to another without assurance that data flow continues

• Example: Assessment of needle biopsy before definitive surgery for breast cancer (quality measure) using billing data
Of 85 patients designated as not having a needle biopsy prior to definitive surgery, ALL had had the procedure done at their referring hospital
Important Issues to Address in Establishing a Mechanism for Data Collection and Sharing for Novel Therapies

- Ability to capture all patients of interest
- Ability to capture all data of interest - no matter where it is generated
- Ensuring data quality
- **Maintaining long-term follow-up**
- Ensuring confidentiality, security and regulatory compliance
- Making data rapidly available for multiple users/uses
Some Data Must Still Get Be Hunted Down by the Transplant Center

- Fragmented health care system mean that patients move from one provider/payer to another without assurance that data flow continues
  - Deaths occurring in one system may not be reflected in another
- Follow-up, especially when necessary over long periods, requires manual processes
- Tracking by SSN would be great – but low patient/institution acceptance
Issues Pertinent to Cellular Therapy Follow-up

• Consider Jane Doe:
  – 12 year with early relapse of ALL
  – CAR-T-cell therapy #1 – no response
  – CAR-T-cell therapy #2 – different construct, different institution - response
  – BMT – third institution
    • *Viral-specific T-cell therapy posttransplant*

• Who reports what to where? How does the data *flow*?
CIBMTR Infrastructure for Collecting Cellular Therapy Data

Cellular Therapy

- Pre-CTED
- Post CTED

Hematopoietic Cell Transplantation

- Pre-TED
- Post-TED

Form 2804/2814

Unique ID Assignment

CRF

Comprehensive Data

Basic Level Of Data Collection

CIBMTR

CENTER FOR INTERNATIONAL BLOOD & MARROW TRANSPLANT RESEARCH
Important Issues to Address in Establishing a Mechanism for Data Collection and Sharing for Novel Therapies

• Ability to capture all patients of interest
• Ability to capture all data of interest- no matter where it is generated
• Ensuring data quality
• Maintaining long-term follow-up
• Ensuring confidentiality, security and regulatory compliance
• Making data rapidly available for multiple users/uses
Human Subjects Protection – HIPAA/Consent/Security

- CIBMTR consider Public Health Authority – no HIPAA authorization required
  - Strict “authority to operate” IT/security standards in compliance with US regulations – annual inspections
- Research protocol at NCT01166009
  - The transplant center is responsible for obtaining the necessary institutional review and approval for participation in the CIBMTR Research Database.
  - Since 2013, CIBMTR consent forms have requested permission for CIBMTR to contact patients directly
Important Issues to Address in Establishing a Mechanism for Data Collection and Sharing for Novel Therapies

• Ability to capture all patients of interest
• Ability to capture all data of interest - no matter where it is generated
• Ensuring data quality
• Maintaining long-term follow-up
• Ensuring confidentiality, security and regulatory compliance
• Making data rapidly available for multiple users/uses
Data Back to Centers (DBtC)

- Introduced in 2009 - provides validated data to centers from CIBMTR Database – without requiring AGNIS mapping
- Downloadable file paired with data dictionary
- 2015: 614 data downloads by 153 centers; 1,567 unique visits
- Data downloads, data dictionary, and code definition tables include all current versions of TED Forms
- Refreshed quarterly

Numbers of Centers Downloading Data by Year
**eDBtC (enhanced DBtC) to Launch in 2016**

- Dramatically improved user interface
- Visualization of center analytics and descriptive statistics
  - ~60 selectable data dimensions, TED and CRF variables
  - ~30 predefined filters
- Expanded available data to CRF data
- Ad Hoc analysis – explore data, including outcomes
- Data will be refreshed monthly
It Takes a Village………..

CIBMTR PUBLICATIONS

by Federal U24 Grant Periods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer-reviewed pubs</td>
<td>34</td>
<td>66</td>
<td>119</td>
<td>235</td>
<td>207</td>
</tr>
<tr>
<td># Different authors</td>
<td>177</td>
<td>295</td>
<td>357</td>
<td>775</td>
<td>1050</td>
</tr>
<tr>
<td># Different institutions</td>
<td>98</td>
<td>158</td>
<td>211</td>
<td>339</td>
<td>418</td>
</tr>
</tbody>
</table>

Why? We provide data - and statistical support and samples

* First 3 years of grant period 2013-2017
So – Where Do We Go

- Better informatics infrastructure – Flatiron, Watson, etc – and interoperability will be a huge help
  - Who will pay for it?
  - Will take time to penetrate the market – but we need to start collecting data now
  - Need to use and adapt existing infrastructures (like outcomes registries) where possible

- No single technology will address all of the issues
  - and some issues are not technological
  - But the issues are addressable
  - Requires collaboration among providers, payers, public and private research funders, patients, entrepreneurs
Thank You