



CMSS WEBINAR SERIES:
Registry Science and Research

CMSS Presents:

***Challenges and Opportunities for
Clinical Registries Collecting Health
Information from Patients***

December 13, 2022

ASCO[®]

AMERICAN SOCIETY OF CLINICAL ONCOLOGY
KNOWLEDGE CONQUERS CANCER

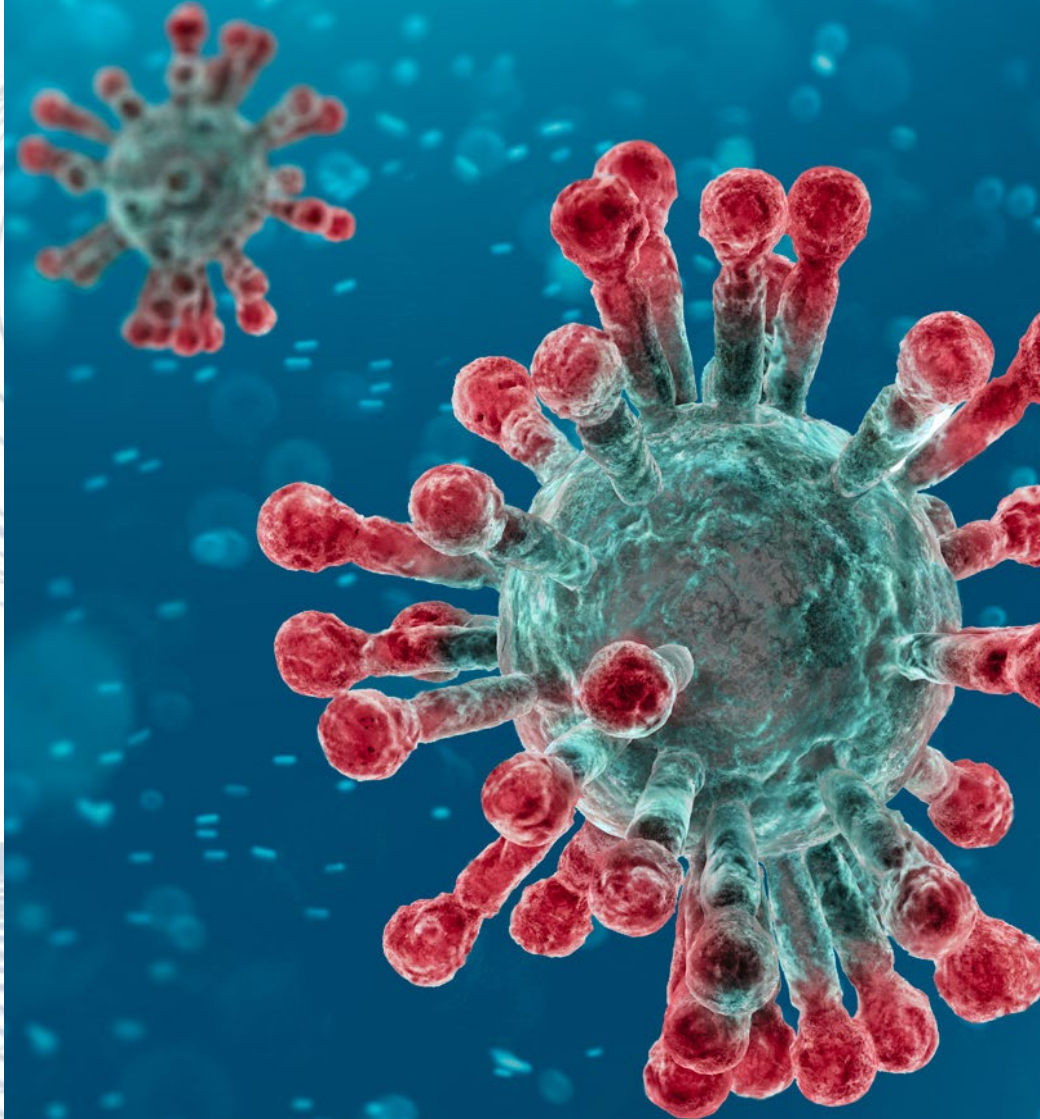
ASCO's COVID-19 Registry

*Liz Garrett-Mayer, PhD
Vice President, Center for Research and Analytics
American Society of Clinical Oncology (ASCO)*

**Challenges and Opportunities for Clinical
Registries Collecting Health Information from
Patients**

December 13, 2022

© 2020 American Society of Clinical Oncology (ASCO). All Rights Reserved Worldwide.



ASCO Registry Overview

Goal:

- Help the cancer community learn more about the patterns of symptoms and severity of COVID-19 among patients in cancer treatment, as well as how COVID-19 is impacting the delivery of cancer care and patient outcomes

Objectives:

- Analyze distribution of symptoms and severity of COVID-19 among people with cancer
- Examine impact of COVID-19 on cancer treatment and outcomes
- Document adaptations of cancer care to the pandemic

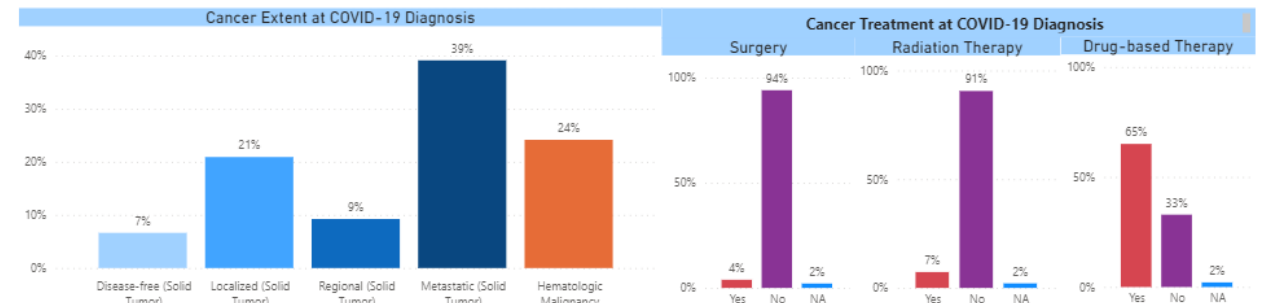
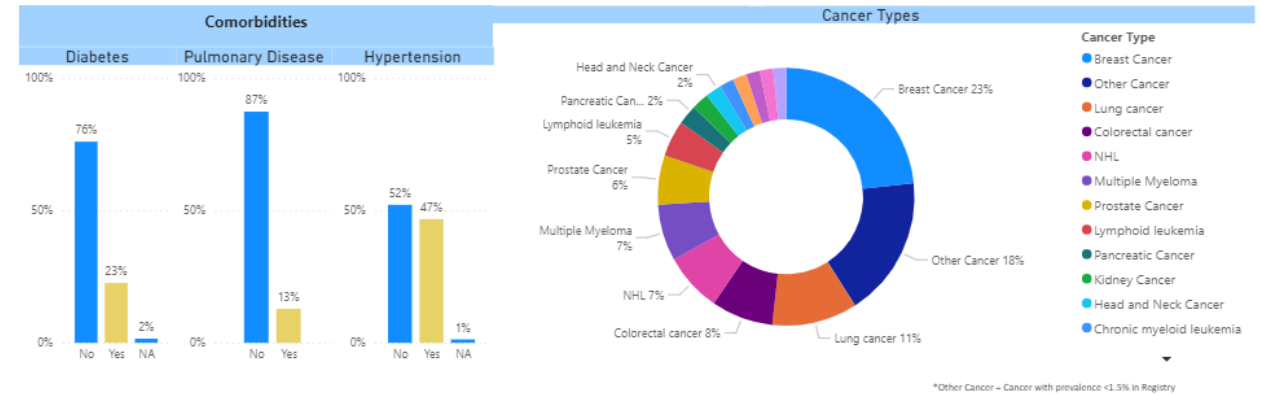
Participation:

- Launched April 2020
- As of December 2022, >6000 patient cases included in registry from 64 practices
- Up to 2 years of follow-up from first SARS-CoV-2 positive test

Target Patient Population

- Positive SARS-CoV-2 test and
- One of the following:
 1. Active cancer at the time of SARS-CoV-2 positive test
 2. Cancer-free for less than 1 year but on adjuvant treatment

ASCO COVID-19 Registry Data Dashboard (snapshot)



The ASCO Registry is supported by Conquer Cancer's [COVID Impacts Cancer Fund](#)

Limited Dataset Challenges

- Established as “no consent” registry
- “Limited” dataset (i.e., no direct identifiers)
- Challenges:
 - Longitudinal tracking of patient information
 - Extracting data from Electronic Health Records
 - Missingness
 - Inability to connect to other sources



Limited Data Set

Tendencies for Bias

Missing cases → lack of representativeness

- Asymptomatic and mild cases less likely to be included

Missing data among cases

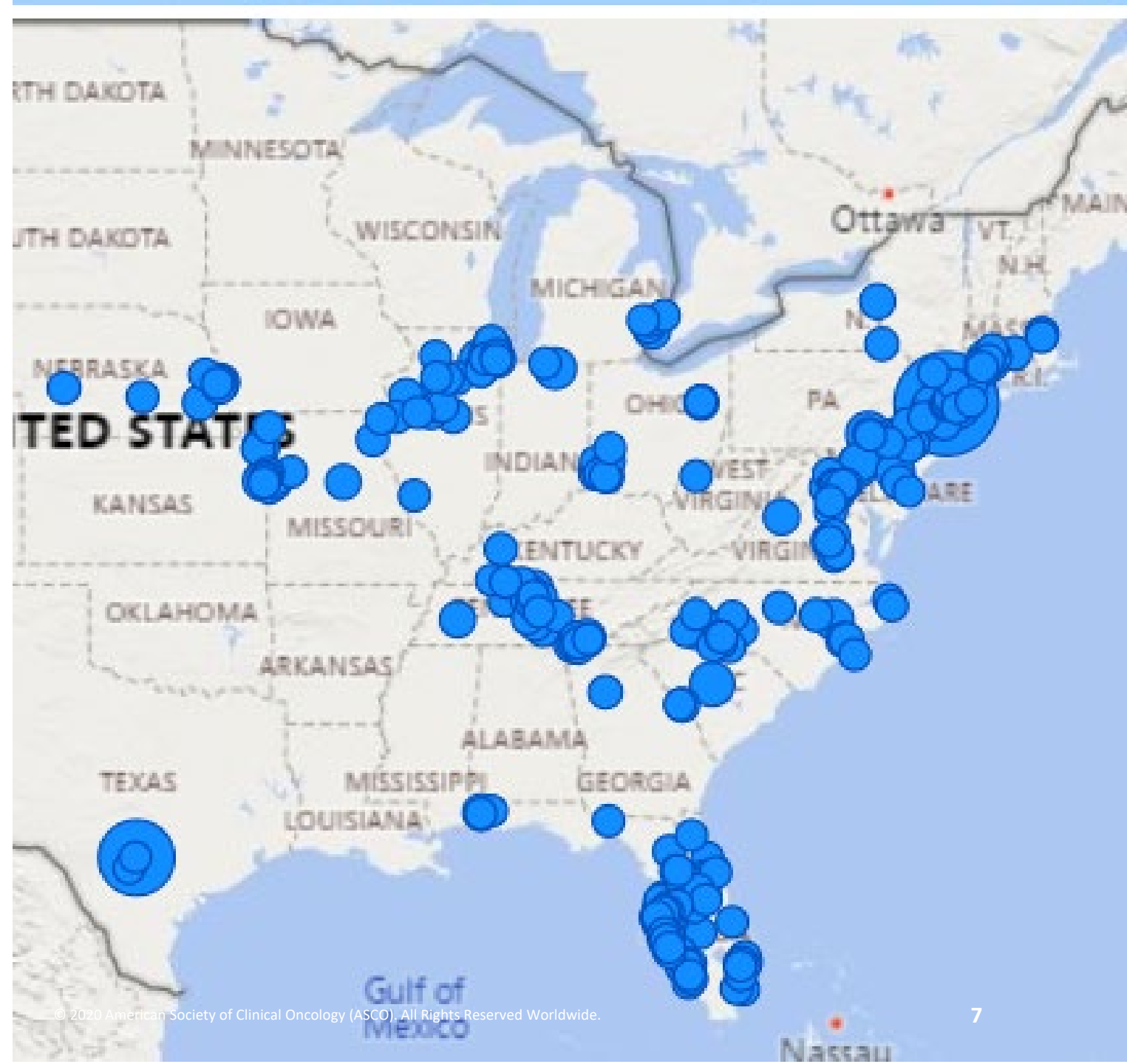
- Reliant on oncology practice EHRs for data on COVID-19
- More severe cases → more data

Changes in missingness over time

- Changes in COVID-19 severity and concern with changes in variants
- Changes in patient population
- Waning interest

De-identification

- Safe Harbor? Nope.
- Expert Deidentification
 - Date shifting
 - Geographic information
 - Social determinants of health considerations



Benefits of direct-to-patient approach for ASCO COVID-19 Registry?

- Case report forms vs. reliance on EHRs.
- Tracking and follow-up in control of registry
- Bias mitigation?
 - Missing data per case: yes (but not perfect)
 - Missing cases: maybe
 - Would require consent
 - Engagement could be a challenge
 - Likely still change in case-mix over time





Registry Science: Non-technical Challenges to Direct-to-Patient Registries

Steven Labkoff, MD, FACP, FACMI, FAMIA
Global Head, Clinical & Healthcare
Informatics
Registry Science Practice Leader

Disclosures



Financial Disclosures

- Pension with Pfizer
- Shareholder in Quantori
- Venture Partner at Boston Millennia Partners

Affiliations

- Collaborating Scientist, Division of Clinical Informatics, Beth Israel Deaconess Medical Center
- Finance Committee, American Medical Informatics Association

Case Study



- A Rare Cancer Patient Support and Research Foundation gets a large donation to help find a cure for the cancer
- The foundation endeavors to enroll the largest **direct-to-patient** registry ever conceived of (for this cancer) - over 5000 patients
- After nearly 3 years building out the platform the registry opens and in year 1, the platform garners well over 1000 patients
- However...
 - As the program rolls on, registrations start to slow
 - Follow up becomes problematic (for those that remain alive)
 - After year 1, the enrolled cohort is too homogeneous, and not reflective of the true prevalence of the disease in the real world
 - Enthusiasm begins to wane, and there are 9 years left on the study

What is a Direct-to-Patient Registry?

A direct-to-patient registry design is one in which recruitment and some or all related communication and data collection is conducted directly with the patients, without guidance from a medical care provider trained in registry procedures.

21st Century Patient Registries: Registries for Evaluating Patient Outcomes: A User's Guide: 3rd Edition, Addendum [Internet]. Dreyer N, Franklin P, Haynes K, Mehta P and Ritchey MB.

While there are many similarities, the differences create new challenges for the study team

Registry Differences

	Traditional Scientific Registry	Direct-to-Patient Registry
Protocol	Required	Required
IRB Review	Required	Required
Recruitment	Managed by Trial Coordinator	Can be all comers. Communications and interfaces must be patient-friendly and easily digested
Enrollment	Managed by Trial Coordinator; May be focused on a specific institution	Patient self-managed; Can target populations according to wider geography - all comers
Consent	Required, managed by Coordinator	Required, managed by Patient
Recruitment Strategy	Clinical Research Managers Process	Patient-facing interfaces Advertise broadly and screen via patient tools
Cohort Diversity	Challenge: Protocol Defined	Challenge: Protocol Defined
Data back to Patients?	Not usually	Generally YES
Loss to Followup	Frequently an issue	DTP can minimize loss to follow-up with outreach

While there are many similarities, the differences create bona fide challenges for the study team

Kinds of Data in a DTP Registry

Data Types	Comments
Patient Medical Records	With appropriate consent and abstraction. Patients will need to provide a written (certified e-signature) authorization
Genomic/proteomics/ epigenomic	Arrangements may be needed for tissue/specimen collection. As this is not being run via a CRO, Life Sci Org, or university - this can get complicated
Patient Reported Outcomes	Can be done easily with permission from authors. Patients need to be computer savvy
Outreach Notes	With appropriate consent, navigator notes can be used as another data type
Imaging	Special data storage arrangements may be needed
Other	Other specific data types may be possible depending on the protocol

Data types can be just as varied as in traditional registry programs. Obtaining and processing the data due to the DTP nature of the study may present different challenges

Some of the Challenges in DTP Long-term, Longitudinal Registries

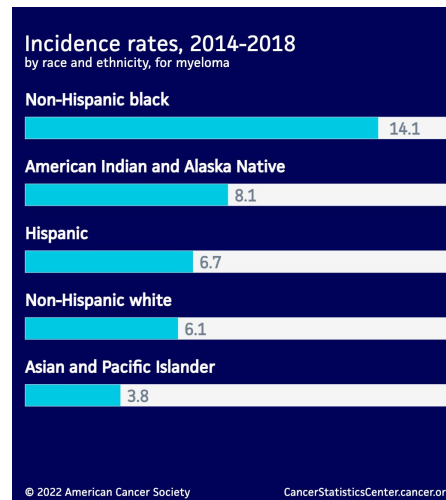
Representativeness	Care must be taken to ensure that you get a representative population; Most DTP Registries take “all comers”
Recontacting Patients	Despite consents that may allow for recontacting, when you do recontact patients, you need something to tell them or talk to them about that’s helpful
Gathering Tissue or other specimens for inclusion	Because a DTP registry is not managed out of a research institution, if tissue is needed, new workflows must be developed to ensure tissue can be obtained and studied outside routine procedures
Patient Literacy	Because DTP registries rely upon patients for nearly all data gathering, all interfaces, communications, and data visualizations must be simplified to meet patient’s literacy levels
Long-term Follow up and the WIIFM	Engagement will depend on ensuring you “give something back”

These are NOT technology issues/challenges, though technology can help

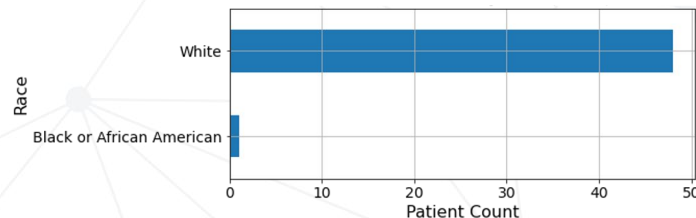
Representativeness

Getting the population right

- The ACS cites: Multiple Myeloma Incidence, African Americans have an incidence of 14%, other studies indicate closer to 18%
- Though the CoMMpass Study was able to recruit appropriate numbers of African Americans, the direct-to-patient study, CureCloud is having challenges
- Early in year 1 of recruiting, the CureCloud only was able to get < 1% of its cohort were African-American
- Possible Reasons:
 - Trust in clinical research
 - Access to technology
 - Marketing efforts focused on the wrong populations
- New strategies needed during recruitment



<https://www.cancer.org/cancer/multiple-myeloma/about/key-statistics.html>



Getting the representation of all relevant groups for a DTP registry can be challenging

Representativeness: When is “Enough” enough? (Why does it have to be SO BIG?)

		Target #	1000		
		% ND	% R1	% R2	% R3+
		15.0%	35.0%	25.0%	25.0%
CoMMpass Variants	% variants from CoMMpass	Newly Dx	Relapse 1	Relapse 2	Relapse 3+
1q Amp	15%	23	53	38	38
Del(13)	22%	33	77	55	55
Del(17p)	5%	8	18	13	13
Del(1p)	10%	15	35	25	25
Hyperdiploid	24%	36	84	60	60
Myc	6%	9	21	15	15
t(11:14)	8%	12	28	20	20
t(14:16)	5%	8	18	13	13
t(14:20)	2%	3	7	5	5
t(4:14)	14%	21	49	35	35
Total of mutations		150	350	250	250

		Target #	2500		
		% ND	% R1	% R2	% R3+
		15.0%	35.0%	25.0%	25.0%
CoMMpass Variants	% variants from CoMMpass	Newly Dx	Relapse 1	Relapse 2	Relapse 3+
1q Amp	15%	56	131	94	94
Del(13)	22%	83	193	138	138
Del(17p)	5%	19	44	31	31
Del(1p)	10%	38	88	63	63
Hyperdiploid	24%	90	210	150	150
Myc	6%	23	53	38	38
t(11:14)	8%	30	70	50	50
t(14:16)	5%	19	44	31	31
t(14:20)	2%	8	18	13	13
t(4:14)	14%	53	123	88	88
Total of mutations		375	875	625	625

<= 20 cases

20-40 cases

>40 cases

Representativeness: Why So Big?

		Target #	5000		
		% ND	% R1	% R2	% R3+
		15.0%	35.0%	25.0%	25.0%
CoMMpass Variants	% variants from CoMMpass	Newly Dx	Relapse 1	Relapse 2	Relapse 3+
1q Amp	15%	113	263	188	188
Del(13)	22%	165	385	275	275
Del(17p)	5%	38	88	63	63
Del(1p)	10%	75	175	125	125
Hyperdiploid	24%	180	420	300	300
Myc	6%	45	105	75	75
t(11:14)	8%	60	140	100	100
t(14:16)	5%	38	88	63	63
t(14:20)	2%	15	35	25	25
t(4:14)	14%	105	245	175	175
Total of mutations		750	1,750	1,250	1,250

<= 20 cases

20-40 cases

>40 cases

Possible Approaches to Attain Better Representativeness

- Start with a true marketing plan/strategy
 - Consider using a consumer marketing agency
 - Go wider than just social media (can introduce bias due to population harmony)
 - Go where the patients are
 - Eg: Work with the religious community
 - Gain notable speaker/sponsors/supporters from the community in question
- Be in constant review mode looking at your study's demographics as you are accruing



Recontacting Patients

- Requires specialized staff (patient navigators or nurses)
- Personalized contact creates trust (long-term engagement)
 - Can discuss “problems” or “topics” to help allay anxiety - and drive engagement
 - A two-way street: Provide new information when possible. Collect new information from subjects
- Generates new data sources
 - Can improve data quality by checking first-hand on donated data
 - Can flag issues early



The more meaningful the contact, the better long-term relationship and engagement

Complexity of Obtaining Tissue or Other Biospecimens

- DTP registries usually do not require patients go to a clinical research center/hospital. Therefore if tissue or biospecimens are needed, there are new challenges
- Novel workflows may need to be built
- Contracts with phlebotomy organizations or laboratory specimen collection organizations may be needed
- Work flows to management tissue blocks could be required
- All of the above must keep the complexity to the patient minimized



Image Reference:
<https://uknow.uky.edu/research/centers-and-institutes/center-clinical-and-translational-science/new-biospecimen-bank>

Data from tissue is generally a critical component of a registry. Ensuring tissue collection can be cleanly and easily managed requires new thinking and workflows

Patient Literacy

- Standard registries are managed by scientifically trained staff - clinical managers, nurses, clinicians, etc.
- DTP Registries rely on patients to provide all information to the registry
- Patient literacy needs to be taken into account in order to ensure that patients understand all aspects of the program - from consent to the return of data
- Keep information as simple as possible
- All patient communications needs to be reviewed/vetted for literacy at no more than an 8th grade reading level

Health Literacy Communication Practices

USE PLAIN LANGUAGE

Plain language is communication all patients can understand the first time they hear it. Avoid technical jargon!

DISTILL INFORMATION

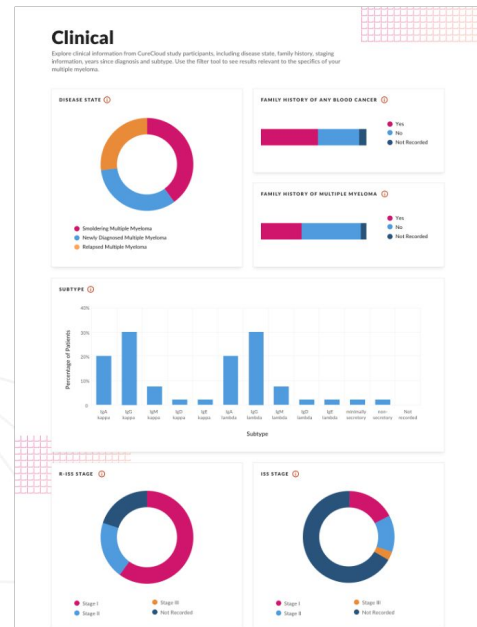
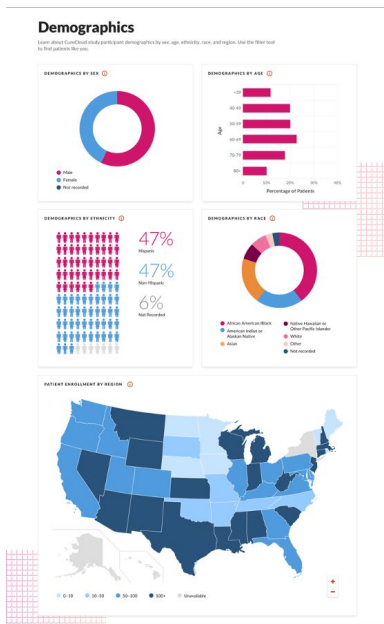
Emphasize the top 1-3 things your patient should know about his/her condition and its management.

<https://www.emra.org/emresident/article/health-literacy/>

Patient literacy can help drive engagement. By ensuring patient understand the program the more likely they are going to be to stay with it

Long-term Engagement: The WIIFM: What's in it for ME!

- Provides value back to patients
- Give them something they ONLY get from the organization/study
- Disseminate up-to-date information
- Use outreach tools such as newsletters
- Share the data: Democratize the data
- Provide seminars/discussions about the registry's intermediate findings
- If there is news coming out, ensure it's broadly communicated
- ENSURE the Reading Level is no higher than 8th Grade
- DO NOT SIT ON FINDINGS!



WIIFM: Democratize Data/Findings

Patients will engage as long as there is a tangible “something” there to help them in their journey

Long-term Engagement

- Keeping patients in the study for the duration of the program is critical to success
- Marketing, Marketing, Marketing
 - Data Visualizations
 - Webinars
 - New Findings
 - Recontacting for updates
- Updated WIIFMs
- Constant cadence of information - such as publications
- Let the patients know when new findings are made as a result of their efforts



Image Reference:
<https://www.globalbankingandfinance.com/how-to-start-a-digital-marketing-agency/>

Building Direct-to-Patient Registries for Medical Specialty Societies

Opportunities, Challenges, Solutions

2022-12-13

Leon Rozenblit, JD, PhD
Head, Registry Practice Center of Excellence

DTP Registries for MSSs: Opportunities, Challenges, Solutions

Context: Broad research use vs QI only; Unit of analysis is the patient vs the provider; “Patient centricity” for MSSs vs PAGs

Opportunities

- **New and diverse** data streams
- **PM-EMR** Ingestion
- Wearables and At-Home **Sensors**
- Build **trust** and engagement

Challenges

- **Data Integration**
- **Identified** data
- **Human** subjects
- **Consent** management
- Representativeness, **sampling bias**, D&I
- **Very** long-term follow up

Solutions

- **Active engagement** & return of value
- **Data rights** brokerage
- Centralized research data **governance**
- Systematic research data **management** on a robust infrastructure



DATAVANT

Connecting the world's health
data to improve patient
outcomes.

CMSS Webinar: **Challenges and Opportunities for Clinical Registries Collecting Health Information from Patients**

Samantha Robicheau

Solutions Lead, Registries & Clinical
Research Networks
December 13, 2022



New Solutions for Privacy-Preserving Record Linkage

Opportunities for
Registries



“Each person in the world creates a Book of Life. This Book starts with birth and ends with death. Record linkage is the name of the process of assembling the pages of this Book into a volume.”

– HALBERT L. DUNN, M.D., 1946
National Office of Vital Statistics

But **health data**
is **fragmented**
across
thousands of
organizations



There are pockets of connected patient data, but the industry lacks a standard

Rare disease patients see on average 7 doctors over 7 years to get a diagnosis



About 35% of patients switch doctors every 2 years

The average person sees 18 doctors in their lifetime

What could we do if data was linked?

COMPLEX DISEASE

EHR



-Omics

Multi-disciplinary diseases impact multiple organ systems and treatment approaches often take a multi-modal approach which can and should be analyzed together.

*Drive deeper understanding of complex diseases and leading to **advances in scientific discovery** by making the most of research investments.*

SPECIALTY FOCUSED

Registry



Claims

Rare disease registries focus on treatment by a particular specialist, but outcomes can be impacted by other comorbidities that can be identified through claims.

*Insight into the **global burden of disease** can inform treatment recommendations and potential areas for future research.*

DIRECT TO PATIENT

PROs



EHR

Collecting patient experience with rare diseases from patients impacted linked to their clinician-reported treatments will better capture the natural history of disease.

*Build out a longitudinal view of the **natural progression of rare disease** that accurately ties treatment patterns to patient outcomes.*

WHOLE PERSON CARE

Registry



SDOH

Combining clinical quality data from a registry with social-determinants of health allows you to understand the social and behavioral impact on patient health, informing whole person care.

*Create risk prediction models that can help inform care delivery and **highlight unaddressed health disparities**.*

Linkage is more important now than ever



The Past

10 years ago

Today

THE BASICS

EMERGING DIGITIZATION

COMPLETE DIGITIZATION OF PATIENT DATA

Pharmacy claims

Medical claims

EHR data

Lab data

Genomics

PROs

Disease registries

Decentralized trials

Wearable devices

Nutrition

Lifestyle

Purchasing

Media consumption

There is a real-world data ecosystem that has de-identified data ready to be linked

Top 10 claims data sources

Top 5 Labs + specialty and genetic labs

Top 5 EHR sources inpatient & outpatient

U.S. Mortality data

>20 specialty pharmacies

120 health plans

9 of Top 10 social determinants sources

>300 million
covered lives

>100 billion
patient records

There are two approaches to link patient data

Traditional Linkage v Privacy-Preserving Record Linkage (PPRL)

Traditional

- Links directly on personally identifiable information (PII) such as SSN or a combination of First Name, Last Name, and DOB matching to link the records

SSN

SSN

123-45-6789 = 123-45-6789

- **PII is exposed.**

PPRL

- Takes PII elements and creates an irreversible “hashed code” that can be used to uniquely represent the individual
- Also called “Tokenization”
- These irreversible hashed codes (tokens) are then encrypted and compared so that the resulting matches can be used to link data or records of an individual.
- **PII is never exposed to conduct entity resolution, nor to actually link together data sets**

Within PPRL there are many techniques dating back to the 1990s



Early PPRL methods

- Simple, Single Hash: National Institutes of Health (NIH) Global Unique Identifiers (GUID) approach
- Requires all of the following to create the token:
 - First Name
 - Middle Name
 - Last Name
 - Sex
 - Date of Birth
 - City/Municipality of Birth

Modern PPRL techniques

- Handle messy input data, misspellings
 - John v Jon
- Creates multiple tokens on different combinations of PII (because not all data sources have access to the same set of PII elements)

Note: Tokenization ≠ deidentification. Tokenization is a strategy to apply when de-identifying records, so that they remain linkable, despite redaction of PII.

Under the Hood: Token designs that yield high-precision matches

Multiple tokens can and should be created if possible to maximize likelihood of matching



Jane Smith
SSN: 123-456-999
DOB: 23rd March 1962

Designs must balance:


- Uniqueness
- Data entry patterns e.g. typos
- Availability across datasets

last name + 1st initial of first name + sex + DOB
last name (soundex) + first name (soundex) + sex + DOB
last name + first name + DOB + zip3
last name + first name + sex + DOB
SSN + sex + DOB
SSN + first name
first name + email
first name + cell phone
last name + 1st 3 characters of first name + sex + DOB + zip3
last name + 1st 3 characters of first name + sex + DOB
last name + 1st 3 characters of first name + sex + zip3
last name + 1st initial of first name + DOB + zip3
SSN + DOB
DOB + Email

Privacy-Preserving Record Linkage in a Nutshell


Identifiable demographic attributes from the data holder → “Tokens” are generated using privacy-preserving software → Authorized recipient links and matches tokens across multiple data sources

Registry A



Jane Smith
SSN: 123-456-999
DOB: 23rd March 1962

Registry B



Jane Smith
SSN: 123-456-999
DOB: 23rd March 1962

EuRZghHw8gY=

nWTsCSRJsI9=

Matched but without disclosing any PII



8y9oJbdg=

8y9oJbdg=

Lessons Learned & Opportunities

For optimal data linkages
for registries

1

Data elements: Collect robust PII in a safe and compliant manner in order to create tokens that lead to high-precision matches

2

Data ownership: Raw registry data stored on a data platform may no longer be owned by the individuals who are now interested in record linkage projects.

3

IRB protocol: Standard language includes sharing only de-identified data, but this could exclude the possibility for sharing & linking tokenized data between different registries or across different research projects.

Recommended additional reading

Sept 2022–**Privacy Preserving Record Linkage (PPRL) for Pediatric COVID-19 Studies**

*National Institute of Child Health and
Human Development (NICHD) Office
of Data Science and Sharing (ODSS)*

[LINK](#)

Nov 2022–**Assessing the impact of privacy-preserving record linkage on record overlap and patient demographic and clinical characteristics in PCORnet®, the National Patient-Centered Clinical Research Network**

*Journal of the American Medical
Informatics Association*

[LINK](#)

Thank you!