



National  
COVID  
Cohort  
Collaborative



JOHNS HOPKINS  
M E D I C I N E

# ***Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and De-identified Clinical Dataset***

Stephanie Hong, BS, FAMIA  
Sr Healthcare Informatics Analyst

N3C  
Data Ingestion and Harmonization Workstream

Johns Hopkins School of Medicine  
Section of Biomedical Informatics and Data Science (BIDS)  
November 14, 2023



# National COVID Cohort Collaborative (N3C) and CMS (Medicare and Medicaid)

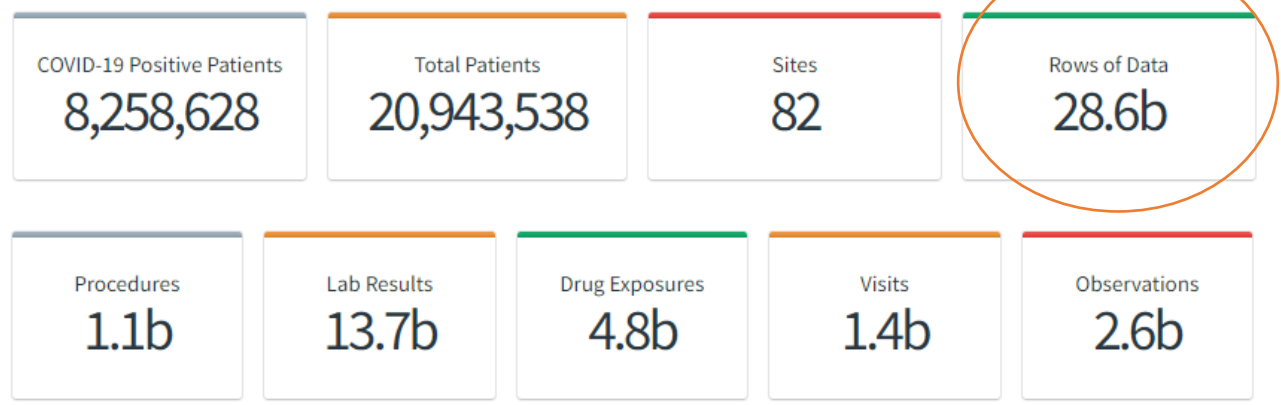
- The N3C Data Enclave is the largest centralized repository of COVID-related patient Electronic Health Record (EHR) data in the United States, available to researchers in the OMOP CDM format.  
*It is the largest OMOP instance to our knowledge.*
- CMS claims files are also transformed into OMOP CDM format.
- Using Privacy Preserving Record Linkage – **N3C data and CMS data are linked by patient id.**
- As a result, N3C EHR datasets in OMOP CDM format are **enriched** with additional CMS claims data.

# Putting the patient back together again

## N3C is enriched with CMS, Mortality, and Viral Variants datasets

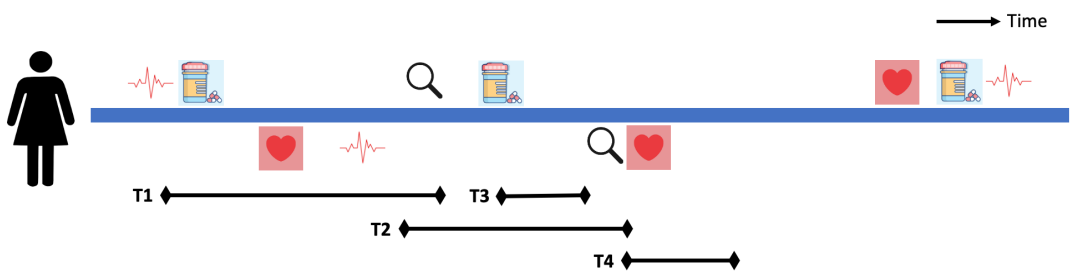
November 14, 2023

### N3C

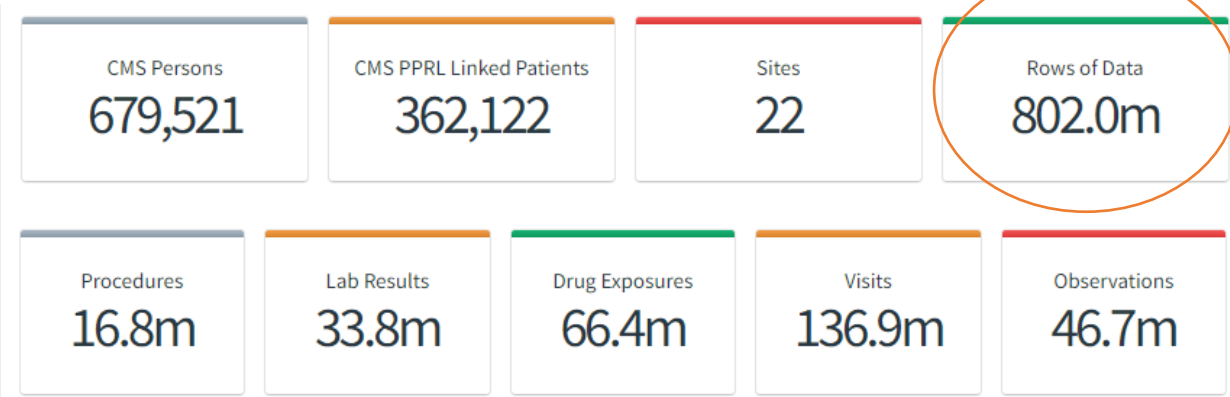


How much data is associated with Privacy Preserving Record Linkage(PPRL)?

Mortality (747,001)  
Viral Variants ( 51,855)



### CMS (Medicare + Medicaid) – CMS Opted-in only



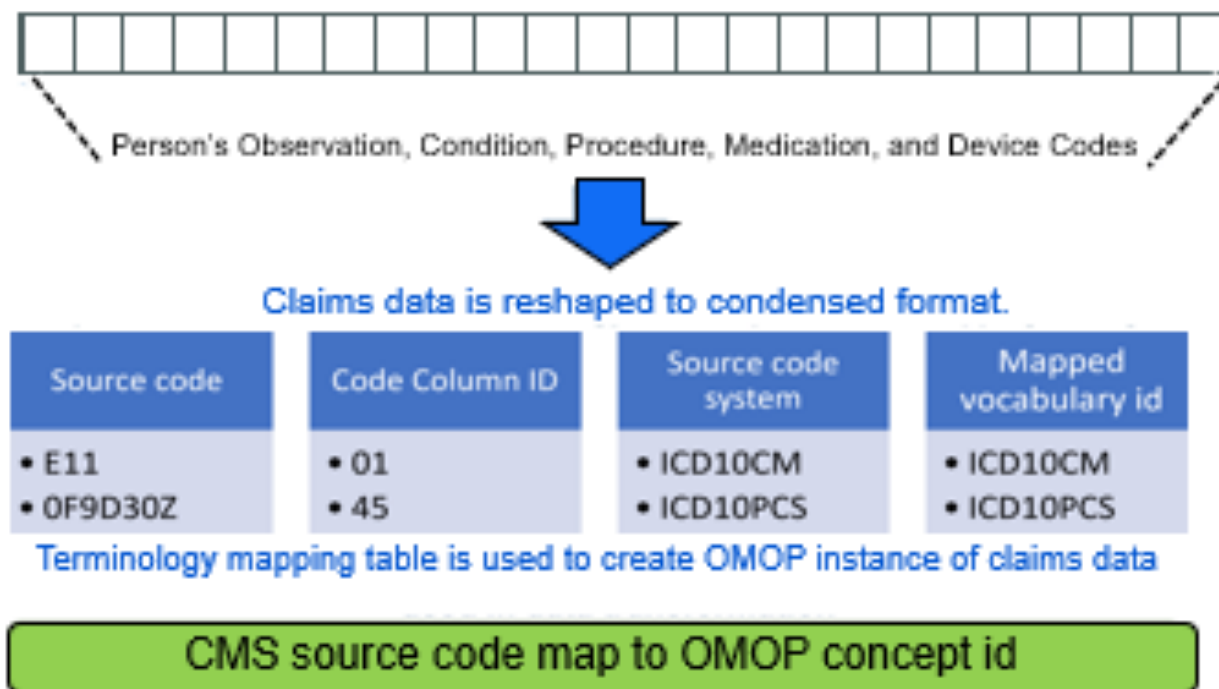
# CMS Claims files were in proprietary format

## We examined each claim files for standard terminologies and coding systems

Medicare Claims	Claims Source File	Description	Terminology
	IP	Inpatient	ICD10CM, ICD10PCS
	OPL	Outpatient	ICD10CM
	PDE	Part D	NDC
	HH	Home Health	ICD10CM, HCPCS
	HS	Hospice	ICD10CM, HCPCS
	DM	Durable Medical Equipment	ICD10CM, HCPCS
	PB	Part B Carrier file	ICD10CM, HCPCS
	SN	Skill Nursing	ICD10CM, HCPCS
Medicaid Claims	Claims Source File	Description	Terminology
	IP	Inpatient	ICD10CM, ICD10PCS
	LT	Long term care	ICD10CM
	OT	Other Services	ICD10CM, CPT4, HCPCS
	RX	Prescriptions	NDC

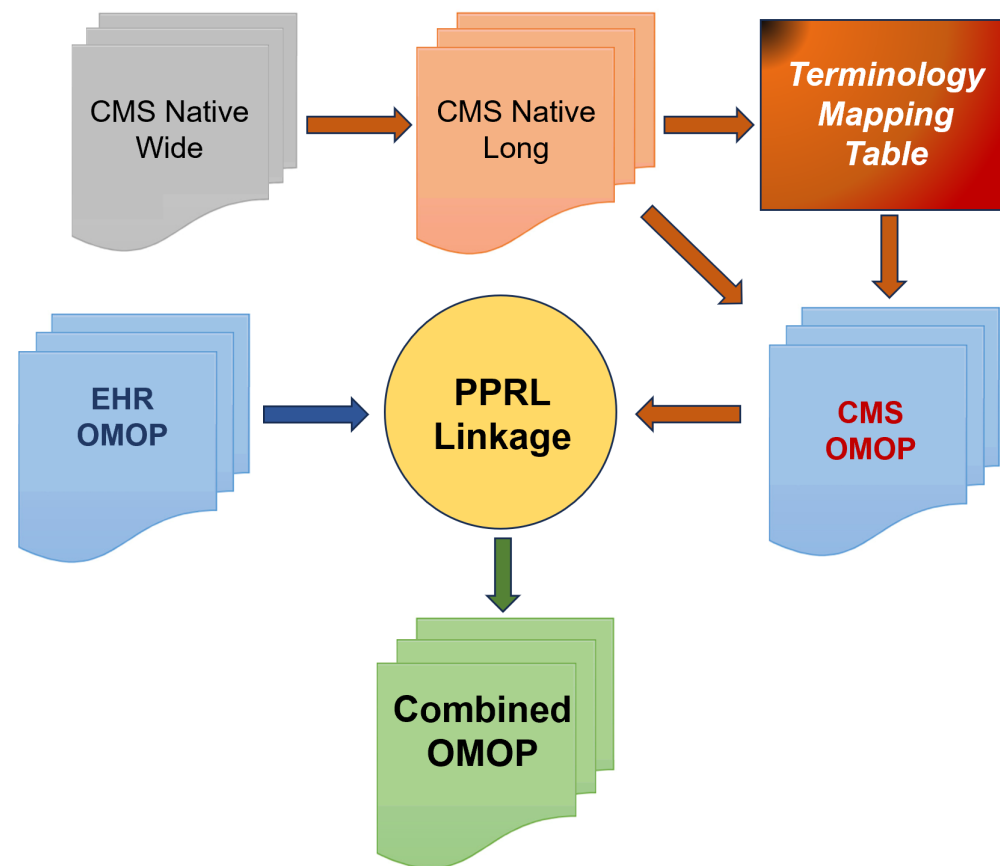
Some claims were 4000+ columns wide  
 with codes spread across from 1 to 45 columns  
 And sometimes one claim can spread across multiple rows  
 with Sequence number

- Dataset was pivoted to contain just the code and the code system
- the provenance of data is preserved
- the terminology mapping table is generated and used in ETL ( Extract-Transform and Load )



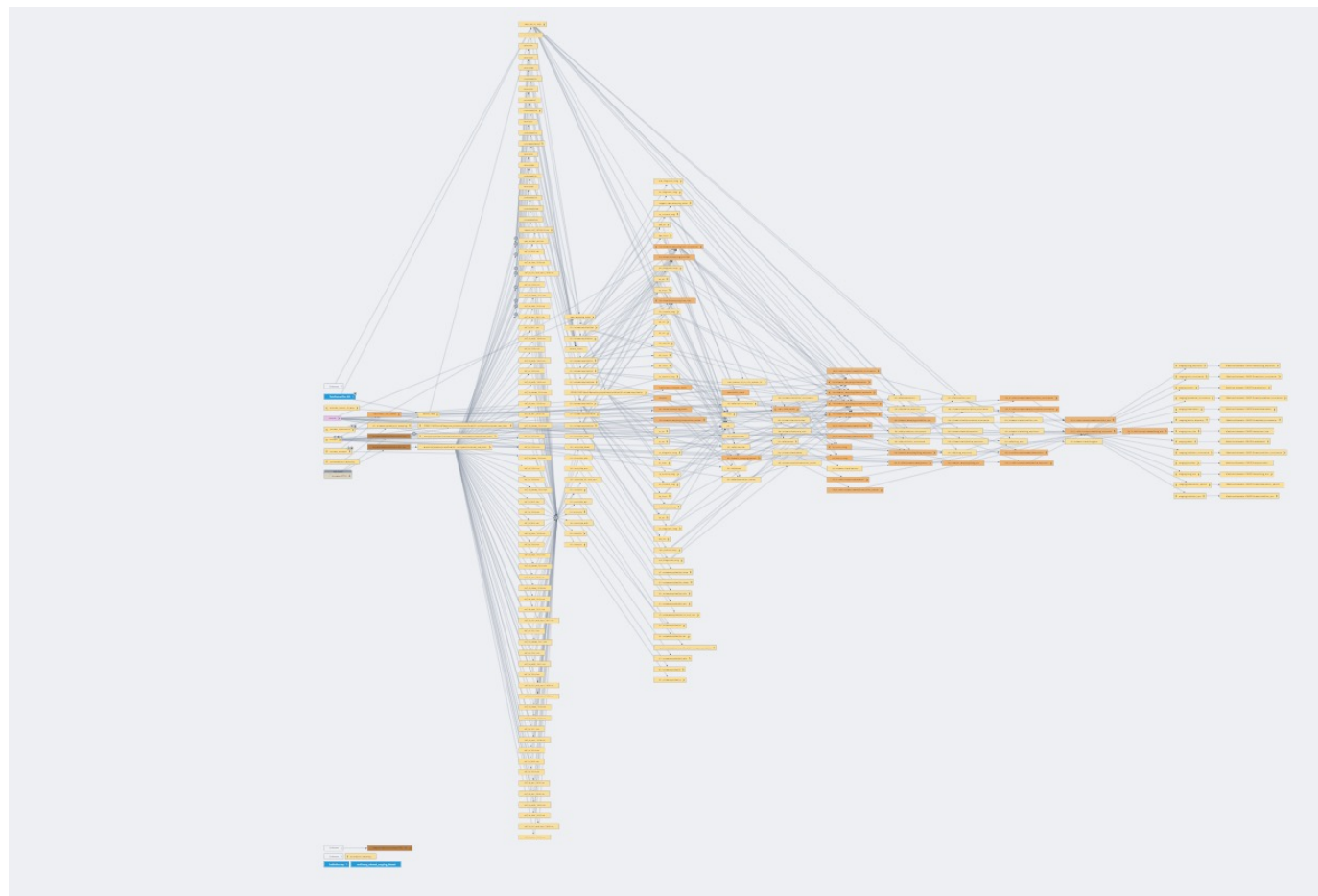
# Pipeline

- Visits - Claim dates, i.e. admit and discharge dates, are used to create the visit
- Visit types - Revenue center code are reference to determine the visit types in addition to the claim type
- N3C EHR OMOP CDM and CMS OMOP CDM are linked by PPRL patient IDs.
- De-duplication - In cases where N3C person\_id is duplicated, a Global ID is provided for each. (It is possible to have data of the same person be represented or submitted to N3C by two different data partner.)



# Generated terminology mapping table is used to transform the CMS claims data to OMOP CDM

- Data pipeline with over **250** transformation steps
- **158** data health checks
- **82 DTAs in N3C** |
- **30 DTAs in N3C CMS PPRL linkage**
- **22 sites CMS opted-in**
- N3C total patients : **20,868,921**
- N3C PPRL-linked CMS patients : **359,096**
- Total rows of data in N3C: **28.3billion**
- Total rows of data in CMS: **653,366,927**
- N3C dataset enriched by CMS





## **RESULTS**

*Among the PPRL-linked patients, in average, additional concepts are available from CMS:*

<b>Claims</b>	<b>Domain</b>	<b>PPRL-Linked patient/ domain</b>	<b>Average # of additional concepts added per patient</b>
<b>Medicare</b>	Condition	71%	78
<b>Medicare</b>	Procedure	60%	6.33
<b>Medicare</b>	Drug_exposure	75%	21.83
<b>Medicare</b>	Measurement	60%	16.48
<b>Medicare</b>	Observation	66.9%	8.6
<b>Medicare</b>	Device	47.6%	6.8
<b>Medicaid</b>	Condition	20.8%	33.9
<b>Medicaid</b>	Procedure	20.2%	23
<b>Medicaid</b>	Drug_exposure	21.9%	20.9
<b>Medicaid</b>	Measurement	17.8%	17.44
<b>Medicaid</b>	Observation	18.3%	6.68
<b>Medicaid</b>	Device	13.9%	6.3



# OHDSI Global Symposium 2023 Showcase-4

- **Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and De-identified Clinical Dataset**

<https://www.ohdsi.org/2023showcase-4>

Thank you!

- N3C mentors
- OHDSI mentors
- JHU mentors

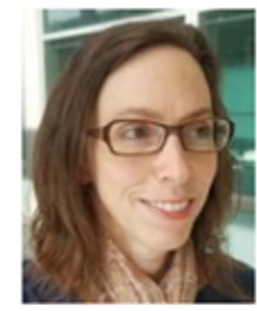


**Clair Blacketer, MPH**  
Director at Janssen Inc.



**Emily Pfaff, PhD, MS**  
Assistant Professor of Medicine

**AREAS OF INTEREST**  
Clinical Informatics



**Melissa Haendel, PhD**  
Professor, Biomedical Informatics  
**Chief Research Informatics Officer**



**Kenneth R. Gersing M.D.**  
Director of Informatics  
Division of Clinical Innovation  
Office of the Director



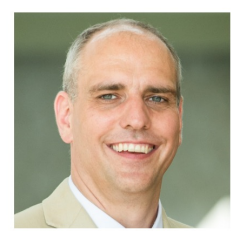
**Christopher G Chute, MD DrPH**  
Bloomberg Distinguished Professor in Health Informatics  
Professor Medicine, Public Health, Nursing  
Chief Research Information Officer, Johns Hopkins Medicine  
Deputy Director, Institute for Clinical and Translational Research  
Head, Section of Biomedical Informatics and Data Science



**Kristin Kostka**  
Director of the OHDSI Center at the Roux Institute



**ANDREW WILLIAMS, PHD**  
Informatics Co-Leader; Faculty; Assistant Professor  
Tufts CTSI, ICRHPS; Tufts University School of Medicine



**Christian Reich**  
Executive Director, OHDSI Center at the Roux Institute



**Harold P. Lehmann, M.D., Ph.D.**  
Professor of Medicine  
**EXPERTISE**  
General Pediatrics  
**RESEARCH INTERESTS**  
Biostatistics; Evidence-based medicine; Decision



**Paul Nagy, Ph.D.**  
Paul Gregory Nagy, Ph.D.  
Deputy Director, Johns Hopkins Medicine Technology Innovation Center  
Associate Professor of Radiology and Radiological Science



National  
COVID  
Cohort  
Collaborative

# Thank you

## shong59@jh.edu



Stephanie Hong, FAMIA<sup>1</sup>, Thomas Richards, MS<sup>6</sup>, Benjamin Amor PhD<sup>2</sup>, Tim Schwab<sup>2</sup>, Philip Sparks<sup>2</sup>, Maya Choudhury<sup>2</sup>, Saad Ljazouli, MS<sup>2</sup>, Peter Leese, MSPH<sup>4</sup> Amin Manna, MEng<sup>2</sup>, Christophe Roeder, MS<sup>3</sup>, Tanner Zhang, MD, MS<sup>1</sup>, Lisa Eskenazi<sup>1</sup>, Bryan Laraway, MS<sup>3</sup>, Nirvisha Garara, MS, Rasi Talluri, James Cavallon<sup>1</sup>, Eric Kim<sup>1</sup>, Shijia Zhang, MS<sup>1</sup>, Emir Amaro Syailendra, MD<sup>1</sup>, Shawn O'Neil, PhD<sup>3</sup>, Davera Gabriel, RN FHL7, FAMIA<sup>1</sup> Sigfried Gold, MS<sup>1</sup>, Tricia Francis, MS<sup>1</sup>, Andrew Girvin, PhD<sup>2</sup>, Emily Pfaff, PhD, MS<sup>4</sup>, Anita Walden, MS<sup>3</sup>, Harold Lehmann, MD, PhD<sup>1</sup>, Melissa Haendel, PhD<sup>3</sup>, Ken Gersing MD<sup>5</sup>, Christopher G Chute, MD DrPH<sup>1</sup>, on behalf of the N3C Consortium



National  
COVID  
Cohort  
Collaborative



Melissa A. Haendel,<sup>1,4,7,8,10,13,14,52,78,101</sup> Christopher G. Chute,<sup>1,4,8,10,13,14,52,78,100,101</sup> Tellen D. Bennett,<sup>9,10,13,14,52,100,101</sup> David A. Eichmann,<sup>4,9,10,13,78,101</sup> Justin  
Guinney,<sup>4,9,10,14,78,101</sup> Warren A. Kibbe,<sup>9,10,52,78,101</sup> Philip R.O. Payne,<sup>4,9,10,78,101</sup> Emily R. Pfaff,<sup>9,10,13,15,52,78</sup> Peter N. Robinson,<sup>4,9,10,15,52,78,100</sup> Joel H. Saltz,<sup>10,13,14,15,52,78,101</sup>  
Heidi Spratt,<sup>9,10,100</sup> Christine Suver,<sup>10,78,101</sup> John Wilbanks,<sup>10,78,101</sup> Adam B. Wilcox,<sup>10,78,101</sup> Andrew E. Williams,<sup>10,13,78</sup> Chunlei Wu,<sup>9,13,14,78</sup> Clair Blacketer,<sup>15,52</sup> Robert L.  
Bradford,<sup>9,52</sup> James J. Cimino,<sup>10,14,101</sup> Marshall Clark,<sup>9,15,52</sup> Evan W. Colmenares,<sup>9,15,52</sup> Patricia A. Francis,<sup>78</sup> Davera Gabriel,<sup>9,10,13,14,15,52</sup> Alexis Graves,<sup>7,9,78</sup> Raju  
Hemadri,<sup>9,15,52</sup> Stephanie S. Hong,<sup>9,15,52</sup> George Hripscak,<sup>10,52</sup> Dazhi Jiao,<sup>9,15,52</sup> Jeffrey G. Klann,<sup>14,52,101</sup> Kristin Kostka,<sup>9,15,52</sup> Adam M. Lee,<sup>9,15,52</sup> Harold P.  
Lehmann,<sup>9,15,52</sup> Lora Lingrey,<sup>9,15,52</sup> Robert T. Miller,<sup>9,15,52</sup> Michele Morris,<sup>9,15,52</sup> Shawn N. Murphy,<sup>9,15,52</sup> Karthik Natarajan,<sup>9,15,52</sup> Matvey B. Palchuk,<sup>9,15,52</sup> Usman  
Sheikh,<sup>9,78</sup> Harold Solbrig,<sup>9,15,52</sup> Shyam Visweswaran,<sup>10,15,52,101</sup> Anita Walden,<sup>7,10,13,14,52,101</sup> Kellie M. Walters,<sup>10,14,101</sup> Griffin M. Weber,<sup>10,101</sup> Xiaohan Tanner  
Zhang,<sup>9,15,52</sup> Richard L. Zhu,<sup>9,15,52</sup> Benjamin Amor,<sup>78</sup> Andrew T. Girvin,<sup>15,78</sup> Amin Manna,<sup>78</sup> Nabeel Qureshi,<sup>15,78</sup> Michael G. Kurilla,<sup>10,78</sup> Sam G. Michael,<sup>10,78</sup> Lili M.  
Portilla,<sup>101</sup> Joni L. Rutter,<sup>1,101</sup> Christopher P. Austin,<sup>101</sup> Ken R. Gersing,<sup>78,101</sup>

Shaymaa Al-Shukri,<sup>4,15</sup> Adil Alaoui,<sup>101</sup> Ahmad Baghal,<sup>15</sup> Pamela D. Banning,<sup>15,100</sup> Edward M. Barbour,<sup>8,15</sup> Michael J. Becich,<sup>15,52,101</sup> Afshin Beheshti,<sup>14</sup> Gordon R. Bernard,<sup>8,15</sup> Sharmodeep Bhattacharyya,<sup>100</sup> Mark M.  
Bissell,<sup>9,15</sup> L. Ebony Boulware,<sup>14,100</sup> Samuel Bozzette,<sup>100,101</sup> Donald E. Brown,<sup>101</sup> John B. Buse,<sup>14</sup> Brian J. Bush,<sup>9,101</sup> Tiffany J. Callahan,<sup>14,52</sup> Thomas R. Campion,<sup>5,15</sup> Elena Casiraghi,<sup>9,15</sup> Ammar A. Chaudhry,<sup>13,14</sup>  
Guanhua Chen,<sup>9</sup> Anjun Chen,<sup>15</sup> Gari D. Clifford,<sup>8,15</sup> Megan P. Coffee,<sup>14,100</sup> Tom Conlin,<sup>14</sup> Connor Cook,<sup>7,78</sup> Keith A. Crandall,<sup>9,14,101</sup> Mariam Deacy,<sup>78</sup> Raouel R. Dietz,<sup>78</sup> Nicholas J. Dobbins,<sup>8,9</sup> Peter L. Elkin,<sup>15,52,100</sup>  
Peter J. Embi,<sup>52,101</sup> Julio C. Facelli,<sup>8,15</sup> Karamarie Fecho,<sup>13</sup> Xue Feng,<sup>9</sup> Randi E. Foraker,<sup>8,13,15</sup> Tamas S. Gal,<sup>8,15</sup> Lingqiang Ge,<sup>14</sup> George Golovko,<sup>15,101</sup> Ramkiran Gounpeddi,<sup>14,15</sup> Casey S. Greene,<sup>13,14</sup> Sangeeta  
Gupta,<sup>52,101</sup> Ashish Gupta,<sup>13,101</sup> Janos G. Hajagos,<sup>9,15</sup> David A. Hanauer,<sup>15,52</sup> Jeremy Richard Harper,<sup>9,14,52</sup> Nomi L. Harris,<sup>14</sup> Paul A. Harris,<sup>101</sup> Mehadi R. Hassan,<sup>9</sup> Yonggun He,<sup>15,52,100</sup> Elaine L. Hill,<sup>9,14</sup> Maureen E.  
Hoatlin,<sup>14</sup> Kristi L. Holmes,<sup>4,101</sup> LaRon Hughes,<sup>14</sup> Randeep S. Jawa,<sup>14</sup> Guoqian Jiang,<sup>14</sup> Xia Jing,<sup>7,14</sup> Marcin P. Joachimiak,<sup>8,15</sup> Steven G. Johnson,<sup>9,14,101</sup> Rishikesan Kamaleswaran,<sup>9,15,78</sup> Thomas George  
Kannampallil,<sup>15,101</sup> Andrew S. Kanter,<sup>15,52</sup> Ramakanth Kavuluru,<sup>9,13,14</sup> Kamil Khanipov,<sup>8,14</sup> Hadi Kharrazi,<sup>9,14</sup> Dongkyu Kim,<sup>15,52</sup> Boyd M. Knosp,<sup>8,15</sup> Arunkumar Krishnan,<sup>9</sup> Tahsin Kurc,<sup>9,15</sup> Albert M. Lai,<sup>101</sup> Christophe G.  
Lambert,<sup>52,101</sup> Michael L. Larsonov,<sup>14</sup> Stephen B. Lee,<sup>1,14</sup> Michael D. Lesh,<sup>9</sup> Olivier Lichtarge,<sup>14</sup> John Liu,<sup>9</sup> Sijia Liu,<sup>8,9,101</sup> Hongfang Liu,<sup>9,15</sup> Johanna J. Lomba,<sup>1,15,78,101</sup> Sandeep K. Mallipattu,<sup>9,14,15</sup> Chaitanya K.  
Mamillapalli,<sup>14</sup> Christopher E. Mason,<sup>15</sup> Jomol P. Mathew,<sup>8,15,52</sup> James C. McClay,<sup>101</sup> Julie A. McMurry,<sup>1,4,7,9,13,14,78</sup> Paras P. Mehta,<sup>14</sup> Ofer Mendelevitch,<sup>9</sup> Stephane Meystre,<sup>8,14,15</sup> Richard A. Moffitt,<sup>9,13,15</sup> Jason H.  
Moore,<sup>8,9</sup> Hiroki Morizono,<sup>15,14,15,52</sup> Christopher J. Mungall,<sup>15,52</sup> Monica C. Munoz-Torres,<sup>7,10,78</sup> Andrew J. Neumann,<sup>78</sup> Xia Ning,<sup>14</sup> Jennifer E. Nyland,<sup>13,14</sup> Lisa O'Keefe,<sup>78</sup> Anna O'Malley,<sup>78</sup> Shawn T. O'Neil,<sup>78</sup> Jihad S.  
Obeid,<sup>10,14,15</sup> Elizabeth L. Ogburn,<sup>13</sup> Jimmy Phuong,<sup>9,15,52,100,101</sup> Jose D Posada,<sup>8,15</sup> Prateek Prasanna,<sup>14,52</sup> Fred Prior,<sup>9,14,15</sup> Justin Prosser,<sup>9,78</sup> Amanda Lienau Purnell,<sup>101</sup> Ali Rahnavard,<sup>8,52</sup> Harish Ramadas,<sup>9,52,78</sup> Justin  
T. Reese,<sup>9,10</sup> Jennifer L. Robinson,<sup>14,100</sup> Daniel L. Rubin,<sup>101</sup> Cody D. Rutherford,<sup>9,101</sup> Eugene M. Sadhu,<sup>8,15</sup> Amit Saha,<sup>9</sup> Mary Morrison Saltz,<sup>15,52,101</sup> Thomas Schaffter,<sup>78</sup> Titus KL Schleyer,<sup>14</sup> Soko Setoguchi,<sup>8,14,15</sup>  
Nigam H. Shah,<sup>8,14</sup> Noha Sharafeldin,<sup>14</sup> Evan Sholle,<sup>15,52</sup> Jonathan C. Silverstein,<sup>15,52,101</sup> Anthony Solomonides,<sup>101</sup> Julian Solway,<sup>14,101</sup> Jing Su,<sup>101</sup> Vignesh Subbian,<sup>9,52,101</sup> Hyo Jung Tak,<sup>15</sup> Bradley W. Taylor,<sup>9,14</sup> Anne  
E. Thessen,<sup>14,101</sup> Jason A. Thomas,<sup>15</sup> Umit Topaloglu,<sup>15,52</sup> Deepak R. Unni,<sup>8,9,15,52</sup> Joshua T. Vogelstein,<sup>14</sup> Andréa M. Volz,<sup>14</sup> David A. Williams,<sup>14,15</sup> Kelli M. Wilson,<sup>9,78</sup> Clark B. Xu,<sup>8,9,15</sup> Hua Xu,<sup>9,10,14</sup> Yao Yan,<sup>9,15,52</sup>  
Elizabeth Zak,<sup>8,15</sup> Lanjing Zhang,<sup>101</sup> Chengda Zhang,<sup>14</sup> Jingyi Zheng,<sup>14</sup>

<sup>1</sup>CREDIT\_00000001 (Conceptualization) <sup>4</sup>CREDIT\_00000004 (Funding acquisition) <sup>7</sup>CRO\_00000007 (Marketing and Communications) <sup>8</sup>CREDIT\_00000008 (Resources) <sup>9</sup>CREDIT\_00000009 (Software role) <sup>10</sup>CREDIT\_00000010 (Supervision role) <sup>13</sup>CREDIT\_00000013 (Original draft) <sup>14</sup>CREDIT\_00000014 (Review and editing) <sup>15</sup>CRO\_00000015 (Data role) <sup>52</sup>CRO\_00000052 (Standards role) <sup>78</sup>CRO\_00000078 (Infrastructure role) <sup>100</sup>Clinical Use Cases <sup>101</sup>Governance

# CMS PPRL Linked N3C Sites Dashboard

- <https://unite.nih.gov/workspace/slate/app/n3c-cms-pprl-linked-sites-dashboard>

\*Note, It is possible to have data of the same person be represented or submitted to N3C by two different data partner.

\*\* Same person can be found in both Medicare and Medicaid.