

Exploring Data Ark Data Commons: A Focus on Accessing MarketScan Dataset

Yiyuan Liu, PhD
The Minerva HPC Team

October 4, 2024



Icahn
School of
Medicine at
**Mount
Sinai**

Outline

- ▶ Data Ark Introduction
 - Mission
 - Available Datasets
 - Data Access
 - Data Onboarding Procedures and Policy
- ▶ Accessing MarketScan Dataset
 - Data Information
 - Data Access Workflow

Introduction to Data Ark

Data Ark Data Commons Increase the Power, Pace and Relevance of Our Science

Challenges



- ▶ Exhaustive searches for relevant datasets
- ▶ Repeated downloads of the same files across groups
- ▶ Difficulty in understanding opaque data structures

How Data Ark Helps



- ▶ Storage space for frequent-use research datasets
- ▶ A team managing the resource, simplifying access, training and user support

Data Ark website: <https://labs.ica hn.mssm.edu/minervalab/resources/data-ark/>

Data Ark Offers Mount Sinai Researchers Readily Available Datasets

- There are 19 datasets hosted under Data Ark currently

Access within 24 hours after DUA signed

Public Data Sets

- 1,000 Genomes Project
- BLAST
- gnomAD
- eQTLGen
- Genebass
- GTEx
- GWAS Summary Stats
- LDSCORE
- Reference Genome
- The Cancer Genome Atlas (TCGA)
- UKBB-LD

Mount Sinai Generated Data

- CBIPM-BioMe Data
- Living Brain Project
- Mount Sinai COVID-19 Biobank
- MSDW COVID-19 EHR Data Set
- MSDW OMOP EHR Data
- STOP COVID NYC Cohort

Restricted access

Public Data Sets

- UK Biobank Genotype (Offboarding in Planning)

User Group-acquired Data Sets

- MarketScan® (Annual Data Access License Expires on Nov 4, 2024)

Continued Access to MarketScan Data Contingent on Data Access License Renewal

How to Rapidly Access Public and Mount Sinai-Generated Datasets Through Data Ark

- ▶ Visit the dataset webpage on Data Ark
- ▶ Access instruction is provided in the 'Access' section
- ▶ User completion of dataset-specific [DUA](#) (data use agreement)
- ▶ Access is granted within 24 hours

How to Access the Restricted Dataset (MarketScan) Through Data Ark

► For MarketScan data

- Providing training certificates
- Sign the DUA
- PI signs the DUA
- License cost applicable after initial 90-day free-of-charge access
- Contact Dr. Parul Agarwal or Dr. Inga Peter (the current and incoming custodian for the MarketScan data and the MarketScan user group) for the cost of data access after the initial 90 days

Access

The following documents are required for each member of the research team:

- Mount Sinai HIPAA training certificate. Certificates of completion can be found at [Mount Sinai PEAK](#)
- CITI Basic Course and Refresher (if due) training certificates. Information for required certificates is [here](#)
- [Terms of Use for MarketScan®](#)

To use these data, you must read, agree and sign the [Data Use Agreement](#) (you must be logged in with **your Minerva ID and password** through the Mount Sinai campus network or secure remote VPN). If you don't have a Minerva ID, please open a ticket with us on MarketScan data access at hpchelp@hpc.mssm.edu

Data Onboarding Process

User-requested datasets must follow an approval process.

1. Data onboarding requested via the REDCap form (https://redcap.link/data_intake) that asks for the storage needed and prospective users.
2. Data Ark team verifies that the prospective users will use the data set.
3. If the dataset is < 1 TB, the Data Ark team will approve and start the onboarding process (webpage, copying data, verifying consent, build data usage agreement, notify users, etc.)
4. Dataset > 1 TB requires the Advisory Board for approval.

Eligibility for Cost-waived for Data Ark Hosting and Data Retention Policy

- The eligibility for cost-waived hosting on Data Ark is based on the number of user groups calibrated to the data size.

Data Size (in Terabytes)	# Of user groups/dataset	Cost waived/year
1 or less	≥ 2	\$100
3	≥ 3	\$300
10	≥ 10	\$1,000 (\$500 for 6 months)
20	≥ 20	\$2,000 (\$1,000 for 6 months)
30	≥ 20	\$3,000 (\$1,500 for 6 months)
100	≥ 20	\$10,000 (\$5,000 for 6 months)

- Data with annual low usage will be archived and offboarded.

Questions and Support: Contact Data Ark

All service requests must come through the ticket system:

hpchelp@hpc.mssm.edu

MarketScan Data Access

MarketScan Proprietary Data

- ▶ Longitudinal, retrospective, de-identified medical and prescription drug claims information of millions of Americans across states insured by private health plans, Medicare and Medicaid plans.
- ▶ For analysis of episodes of illness, health risk factors, treatment patterns, costs, and outcomes.
- ▶ MarketScan data served by Data Ark is owned by Merative.
- ▶ By early November each year, Merative, the data owner, settles into an annual use agreement with the school user group to license the MarketScan data use.
- ▶ MarketScan data hosted by Data Ark spans the years 2013 to 2021.

Data Elements in MarketScan

- ▶ MarketScan data contains information for millions of enrollees annually on

Dataset	Key Data
Enrollment	<ul style="list-style-type: none">• Demographics (age, gender, geographic region)• Plan type (HMO, PPO, etc.)• Enrollment start and end dates
Medical Claims	<ul style="list-style-type: none">• Inpatient admissions• Outpatient visits• Diagnoses (ICD codes)• Procedures (CPT/HCPCS codes)• Dates and place of service• Payments and charges
Outpatient Prescription Drug Claims	<ul style="list-style-type: none">• NDC drug codes• Dispense date• Days supply• Payments and charges
Provider	<ul style="list-style-type: none">• Type of provider (physician, facility, etc.)• Specialty for professionals• Geographic region

Documentation and Training Materials Available for MarketScan Data

- ▶ Documentation on MarketScan data and access is available on the Data Ark webpage [MarketScan Data](#) and [MarketScan User Handbook](#).
- ▶ Additional meta-data including data dictionary provided by Merative, the data owner, is available and **restricted to existing MarketScan users** through a directory [meta-data on One Drive](#).
- ▶ Slide deck and video recording for the MarketScan data training provided by Merative on January 31, 2024 is available and also **restricted to existing MarketScan users** through a separate directory [training materials on One Drive](#).

Data Ark and MarketScan User Group Jointly Ensure Our MarketScan Services Meet Users' Needs

What Data Ark does



- ▶ **Access permission granting according to the agreement with the user group,**
- ▶ **Technical support,**
 - a. SAS license renewal;
 - b. Computational resource and infrastructure for users' MarketScan data analyses.

What the user group does



- ▶ **License cost and terms negotiation with the data provider and among members;**
- ▶ **Setting out Data Use Agreement;**
- ▶ **User access authorization;**
- ▶ **Limited capacity for consultation on data analysis and SAS programming;**
- ▶ **Leader Dr. Parul Agarwal (Nov 2023- Nov 2024).**
- ▶ **Leader Dr. Inga Peter (Nov 2024 and onwards).**

Users' responsibilities include learning SAS programming and understanding MarketScan data elements.

License Cost Imposed by the Data Owner for Accessing the MarketScan Data

- ▶ For internally funded studies, users' access to the MarketScan data incurs **no cost for the first 90 days from the date of access granted**.
- ▶ **Thereafter, there is an associated cost, determined by the MarketScan user group.**
- ▶ For **externally funded studies**, the additional fee to use the school-acquired data sets in non-commercial projects will be **\$30,000 per study**.
- ▶ The additional fee to use the data in a **commercially funded study** will be **\$60,000 per study**.
- ▶ License cost-related inquiries shall be directed to the MarketScan user group leaders, **Dr. Parul Agarwal (Nov 2023- Nov 2024) and Dr. Inga Peter (Nov 2024 and onwards)**.

Continued Access to the MarketScan Data Contingent on Successful License Renewal

- ▶ The current annual license expires on Nov 4, 2024.
- ▶ **Raw and user-generated data will be deleted from the MarketScan servers at Mount Sinai if the license is not renewed.**
- ▶ Users may keep only **processed data** on a local computer.

MarketScan Data Access via Minerva HPC and Windows-based Server

- ▶ MarketScan data is now accessible through Linux-based Minerva HPC (high-performance computing) and the legacy Windows-based server.
- ▶ **Advantages provided by Minerva HPC:**
 - Guaranteed computational resources (CPU capacity, memory and disk storage for user processed data) dedicated to each user job;
 - Concurrent sessions from multiple users without compromising on the computational performance;
 - Potentially, computational resources expandable based on users' needs;
 - User data import from or export to project directories, user account-specific directories.
- ▶ Data Ark/HPC team provides SAS program IDE (Integrated Development Environment) on [OnDemand of Minerva HPC](#) , and will announce additional programming (for example, R) environment deployed on [OnDemand of Minerva HPC](#).
- ▶ **Windows-based server is intended to be retired by Nov 2024.**

MarketScan Data Access through Minerva Prerequisite

School Affiliation Required Due to the Licensing Agreement

1. Obtain an active account of the school (Icahn School of Medicine at Mount Sinai)
 - a. Register an Icahn School of Medicine account through [Sailpoint](#), if not already.
2. School VPN tunnel setup when remotely and VIP two-factor authentication
 - a. Setup of VPN tunnel and VIP token is through [ASCIT](#), if not already.

Access to the following webpages requires campus network or school/hospital VPN tunnel enabled when remote.

1. Minerva account granted upon request and account activation
 - a. [Request a Minerva account](#), if not already.
 - b. Annual NIH and HIPAA forms and OnDemand profile completion through [forms](#) (authentication using either school or hospital ID and associated password **without** the VIP token).
2. Login to Minerva HPC once in order for the home directory for a new user to be created
 - a. Documentation on [logging in to Minerva](#) (authentication using only school ID and associated password **with** the VIP token).

MarketScan Data Access

5. Access MarketScan data

- a. Complete MarketScan DUA (data use agreement) through the [Data Ark DUA website](#) (authentication using only school ID and associated password **without** the VIP token).
- b. Submit training certificates required in the DUA to hpchelp@hpc.mssm.edu.
- c. PI completes a separate MarketScan DUA requested by the Data Ark/HPC team.

Please proceed to the next step only after receiving confirmation through the HPC ticketing system if you are a new user or access Minerva for the first time.

5. MarketScan data and SAS IDE is accessible via

- a. The legacy Windows-based server through [Storefront](#) (local installation of Citrix Workspace app required) scheduled for retirement before Nov 2024(or maybe sooner).
- b. [OnDemand of Minerva HPC](#).

Demo: Access Market Data via OnDemand (ondemand.hpc.mssm.edu)

Contact for Help

- ▶ **Computational infrastructure-related queries**
 - HPC ticket system hpchelp@hpc.mssm.edu.
- ▶ **Limited consultation on data analysis and SAS programming**
 - MarketScan user group leader
 - Dr. Parul Agarwal (**Nov 2023 - Nov 2024**)
 - Associate Professor, Department of Population Health Science and Policy
 - parul.agarwal@mountsinai.org
 - Dr. Inga Peter (**Nov 2024 and onwards**)
 - Professor, Department of Genetics and Genomic Sciences
 - inga.peter@mssm.edu

Please Acknowledge CTSA in Your Publications

- ▶ Please acknowledge the support from Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai by including the following acknowledgement in a publication of any material, whether copyrighted or not, based on or developed with Minerva HPC resources:

“This work was supported in part through the computational resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences.”





Thank You!

MarketScan Proposal

Icahn School of Medicine at Mount Sinai

Jake Luce
Account Manager



Contract Updates

Data Provisioning on SaaS program (Workspace) – No longer delivering through sFTP

Grant Funded Work - \$25,000 per project (*previously \$30,000*)

- If the award is less than \$250,000, the incremental fee to use MarketScan is *at least* \$10,000 per study or 10% of the total award per study, whichever is greater.
- If the award exceeds \$250,000, the incremental fee is fixed at \$25,000 per study.
- Client shall provide supporting documentation for the total amount awarded for each Study Rider request that is below the standard fee of \$25,000.

Classroom Dataset - \$10,000 per year

- Previously not an approved for MarketScan Data
- Unique dataset for use in classroom and teaching setting

Treatment Pathways Tool - \$40,000 per year

- Previously not licensed to the academic market
- Unique tool to create and analyze cohorts without programming experience
 - If interested, request a demo of the tool.

Rolling Data Model (*Required for Packages Only*)

- New data will be delivered upon availability (usually December or January)
- Upon Client access to the newest Data year, Client's access to the oldest Data year will be terminated.
- Any Data stored in accordance with the Storage section below must be deleted upon expiration.

Current Contract

Expires November 4, 2024

Current License	
<u>Dataset</u>	<u>Data Period</u>
Commercial – Set A	9 (2013-2021)
Medicare – Set A	9 (2013-2021)

12-month fee = \$74,248

Renew Legacy Scope

12-month Fee: \$82,500

12-Month License – Existing Scope	
<u>Dataset</u>	<u>Data Period</u>
Commercial – Set A	10 (2013-2022)
Medicare – Set A	10 (2013-2022)

Apollo Package

12-month Fee: \$80,500

12-Month License – Apollo Package	
Database	Data Period
Commercial & Medicare – Set A	10 years (2014-2023)
Medicaid	10 years (2014-2023)
National Weights File	10 years (2014-2023)
Standard Quarterly Updates (Labs, Commercial & Medicare)	Variable
Lab Results	10 years (2014-2023)
Dental	10 years (2014-2023)

36-month Fees:

Year 1	\$80,500
Year 2	\$84,525
Year 3	\$88,751
Total 36-month fee	\$253,776
Average per year	\$84,592

*This non-binding estimate for Merative contains preliminary planning estimates that can be used for project budgeting purposes. The estimates are based on preliminary pre-sales information provided by you. Merative reserves the right to change this non-binding estimate at its sole discretion. All final pricing information is subject to the successful negotiation of a binding, formal contract that is negotiated and executed by both parties. The Merative solution will not be made available to you prior to the successful negotiation and execution of a binding, formal contract by both parties.

Rolling Data Model *(Required for Packages)*

Atlas Package

12-month Fee: \$99,000

12-Month License – Atlas Package	
Database	Data Period
Commercial & Medicare – Set A	15 years (2009-2023)
Multi-State Medicaid	15 years (2009-2023)
Mortality Data	10 years (2014-2023)
Lab Results	15 years (2009-2023)
Dental	15 years (2009-2023)
Health and Productivity Management	10 years (2014-2023)
Health Risk Assessment	10 years (2014-2023)
Benefit Plan Design	8 years (2016-2023)
National Weights File	15 years (2009-2023)
Quarterly Updates (Labs, Commercial & Medicare)	Variable
Medicaid Updates	Semi-Annual

** Secures access to historically licensed data*

36-month Fees:

Year 1	\$99,000
Year 2	\$103,950
Year 3	\$109,148
Total 36-month fee	\$312,098
Average per year	\$104,032.50

*This non-binding estimate for Merative contains preliminary planning estimates that can be used for project budgeting purposes. The estimates are based on preliminary pre-sales information provided by you. Merative reserves the right to change this non-binding estimate at its sole discretion. All final pricing information is subject to the successful negotiation of a binding, formal contract that is negotiated and executed by both parties. The Merative solution will not be made available to you prior to the successful negotiation and execution of a binding, formal contract by both parties.

Orion Package

12-month Fee: \$149,000

Database	Data Period
Commercial and Medicare Standard Edition Databases	25 most recent years
Multi-State Medicaid Database	23 most recent years
Lab Database	18 most recent years
Commercial and Medicare Mortality Edition Databases	10 most recent years
Health and Productivity Management Database	17 most recent years
Benefit Plan Design Database	8 most recent years
Health Risk Assessment Database	17 most recent years
Dental Database	Up to 18 most recent years
Interim updates of the Multi-State Medicaid Database	2 updates, starting with the current one at the time of provisioning
Standard Quarterly Updates and quarterly Early View releases of the Commercial and Medicare Databases	4 updates, starting with the current one at the time of provisioning
National Weights	15 most recent years

36-month Fees:

Year 1	\$149,000
Year 2	\$156,450
Year 3	\$164,273
Total 36-month fee	\$469,723
Average per year	\$156,574

*This non-binding estimate for Merative contains preliminary planning estimates that can be used for project budgeting purposes. The estimates are based on preliminary pre-sales information provided by you. Merative reserves the right to change this non-binding estimate at its sole discretion. All final pricing information is subject to the successful negotiation of a binding, formal contract that is negotiated and executed by both parties. The Merative solution will not be made available to you prior to the successful negotiation and execution of a binding, formal contract by both parties.



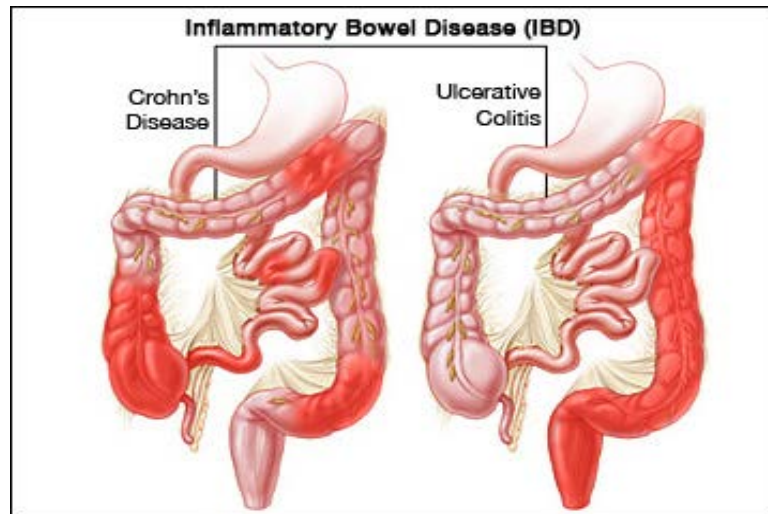
Merative MarketScan[®] database

Inga Peter

With the help of Eunyoung Yang

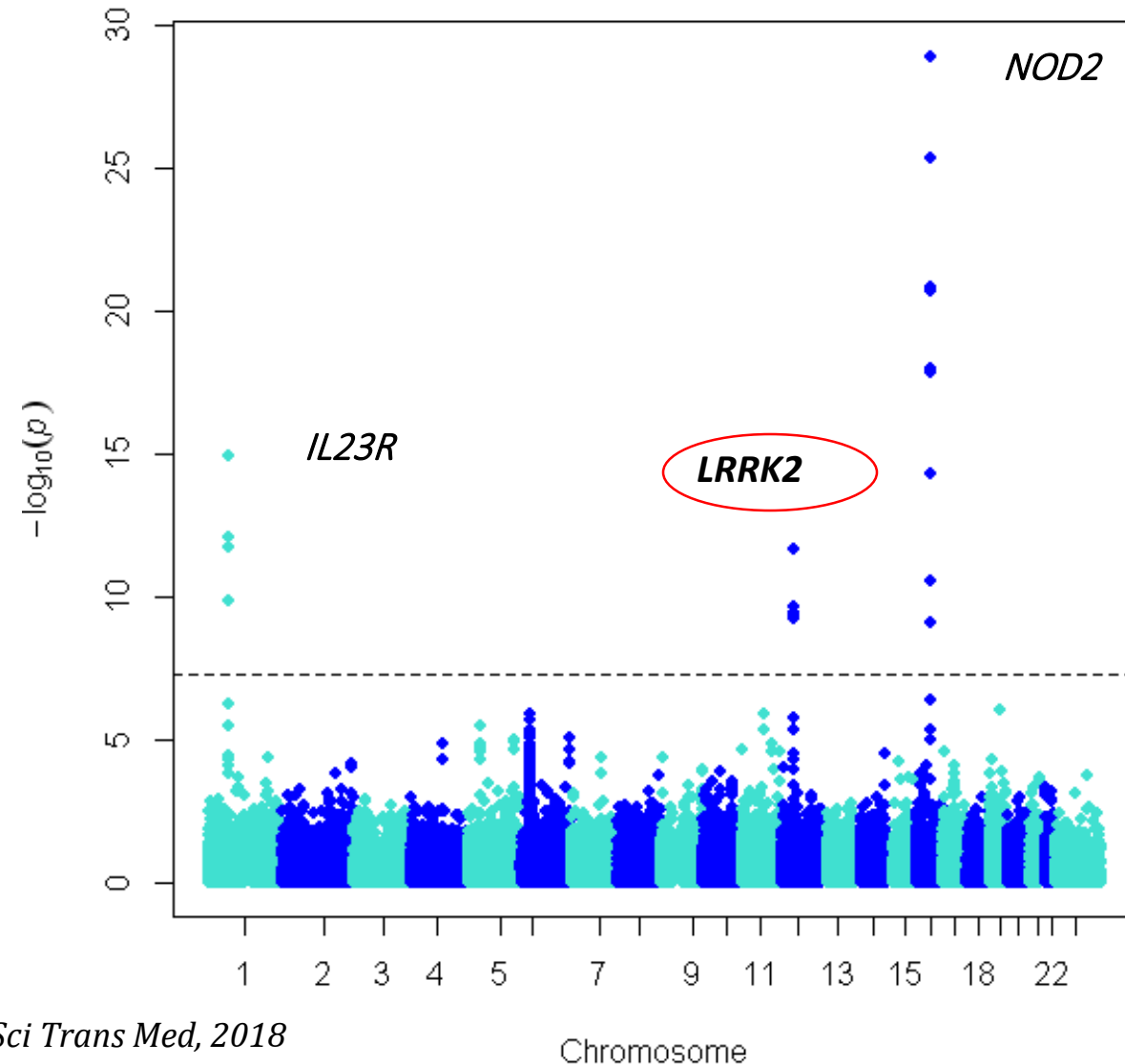
Inflammatory Bowel Disease (IBD)

- **IBD** - a chronic relapsing and remitting immune-mediated disease of the GI tract resulting from excessive response of the immune system to the commensal bacteria in genetically susceptible individuals



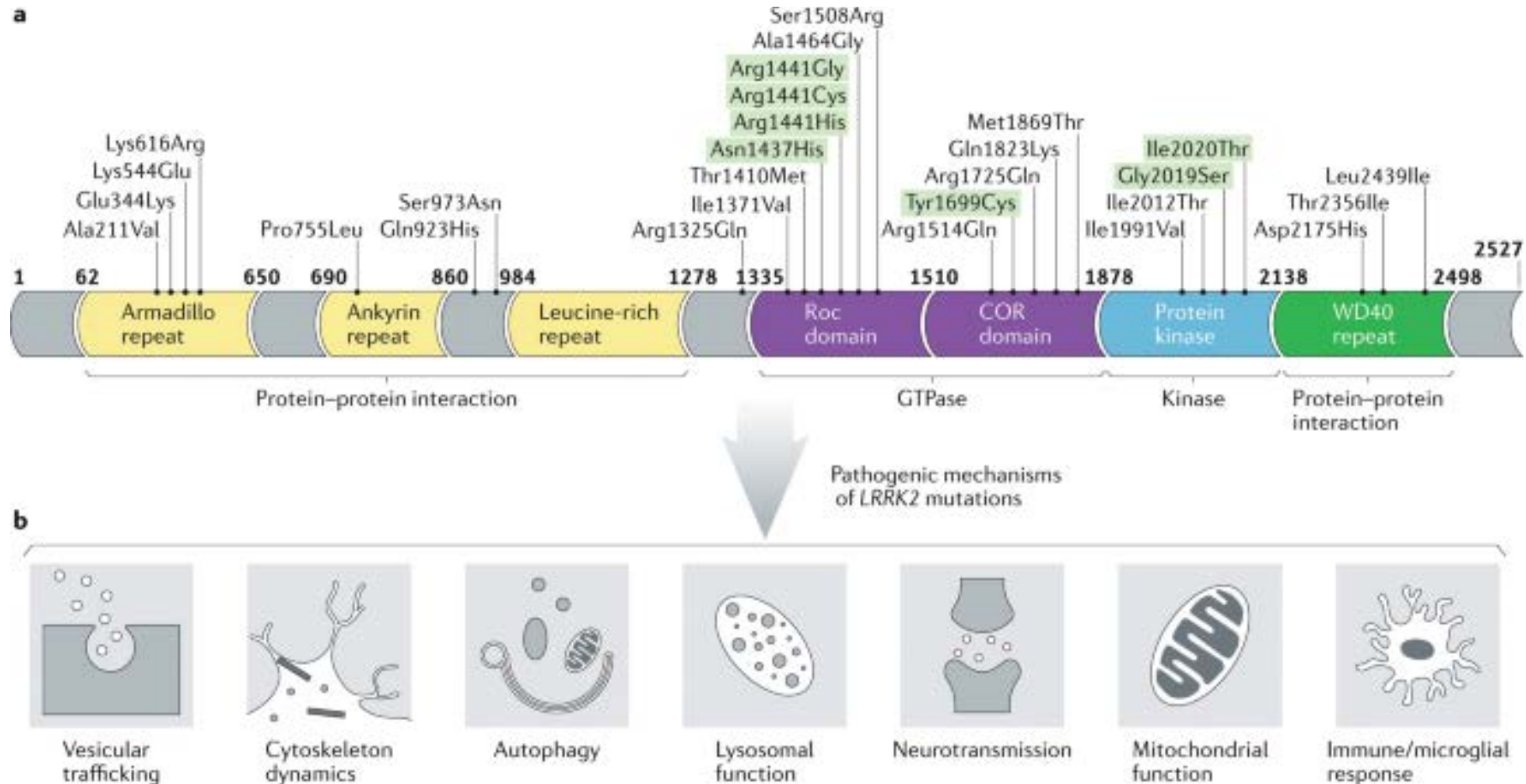
- Characterized by chronic intestinal inflammation that results in the following clinical symptoms:
 - Diarrhea
 - Bleeding
 - Abdominal pain
 - Fever
 - Weight loss
 - Joint pain
- Up to 40% of patients will eventually have a surgery
- Currently, no cure or prevention

Association Analysis of Coding Variants in the Ashkenazi Jewish Population: 4099 CD cases and controls



LRRK2 Leucine-rich repeat serine/threonine-protein kinase 2

Major risk for Parkinson's disease – a progressive movement disorder



Incidence of PD in IBD patients and matched cohorts

Truven administrative claim database, N>170 million person-lives

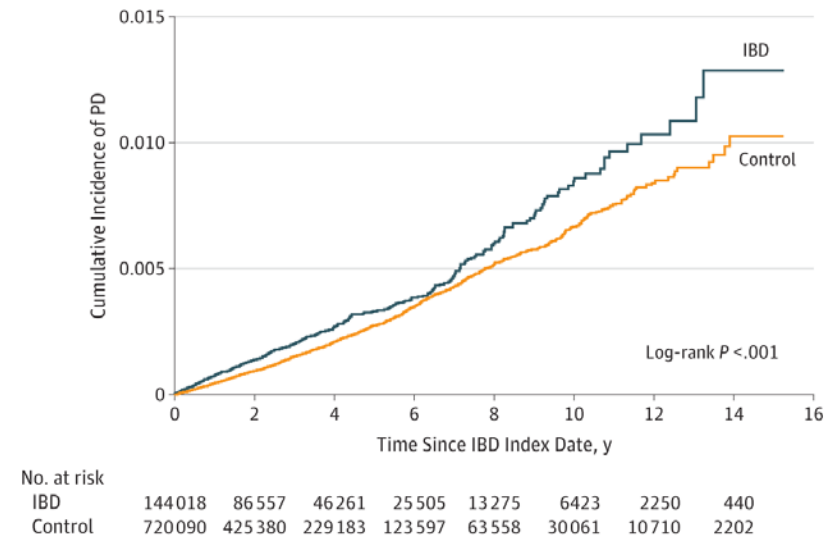


Table 2. Incidence of PD in Persons With IBD and in Matched Persons Without IBD

				Univariate Poisson Model ^c		Multivariate Poisson Model ^d	
Variable	PD Event	Person-years ^a	Rate ^b	Crude IR (95% CI)	P Value	Adjusted IR (95% CI)	P Value
All IBD							
Yes	371	508 033	0.73	1.28 (1.14-1.44)	<.001	1.28 (1.14-1.44)	<.001
Controls	1425	2 500 792	0.57	1 [Reference]		1 [Reference]	
Crohn disease							
Yes	122	194 865	0.63	1.26 (1.03-1.54)	.02	1.26 (1.03-1.53)	.02
Controls	480	966 379	0.50	1 [Reference]		1 [Reference]	
Ulcerative colitis							
Yes	243	302 512	0.80	1.30 (1.13-1.5)	<.001	1.31 (1.14-1.51)	<.001
Controls	913	1 480 956	0.62	1 [Reference]		1 [Reference]	

Abbreviations: IBD, inflammatory bowel disease; IR, incidence rate; PD, Parkinson disease.

^a Crohn disease and ulcerative colitis stratified person-years do not add up to all IBD person-years because undetermined IBD cases are not presented.

^b Incidence rate per 1000 person-years.

^c Analysis of patients with IBD vs matched controls, offset by time.

^d Analysis adjusted for time-varying age group and sex, and offset by time.

Incidence of PD among IBD Patients by Anti-TNF Exposure

Virtual repurposing prevention trial

Table 3. Incidence Analysis of PD Among Patients With IBD by Anti-TNF Exposure

Anti-TNF Exposure ^a	PD Event	Person-years	Rate ^b	Univariate Poisson Model ^c		Multivariate Poisson Model ^d	
				Crude IRR (95% CI)	P Value	Adjusted IRR (95% CI)	P Value
Yes	2	23 610	0.08	0.11 (0.03-0.45)	.002	0.22 (0.05-0.88)	.03
No	369	484 423	0.76	1 [Reference]		1 [Reference]	

Abbreviations: anti-TNF, anti-tumor necrosis factor; IBD, inflammatory bowel disease; IRR, incidence rate ratio; PD, Parkinson disease.

^a Anti-TNF exposure status was defined as yes for all days from anti-TNF exposure index date to last date of anti-TNF exposure coverage or end of enrollment or PD index date, whichever was earliest; otherwise the anti-TNF

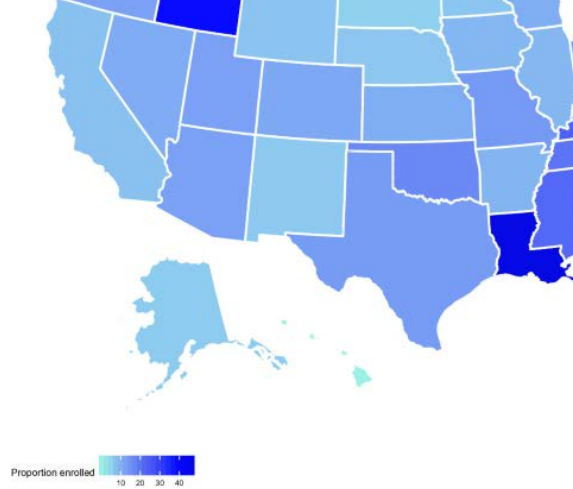
exposure status was defined as no.

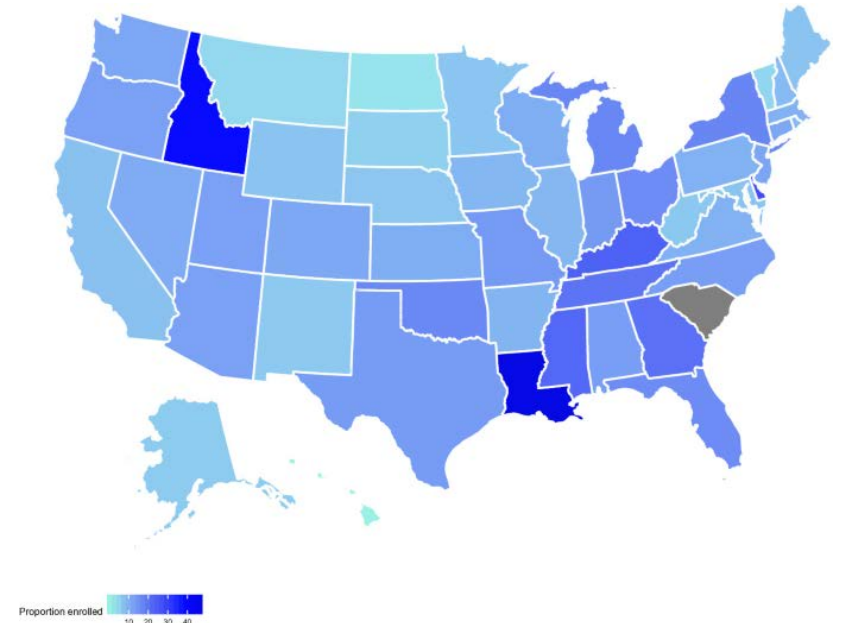
^b Incidence rate per 1000 person-years.

^c Unadjusted incidence ratio, offset by time.

^d Model adjusted for time-varying age group and sex, and offset by time.

IBM MarketScan database

- Including
 - De-identified patient-level health data
 - Medical, drug and dental
 - Productivity
 - Laboratory result
 - Health risk assessments (HRAs)
 - Hospital discharges
 - Electronic medical records (EMRs)
 - One of the largest collections of de-identified US patient data
 - Sample from multiple sources
 - Employers, managed care organizations, hospitals, EMR providers, Medicare and Medicaid
 - More than 37 billion service records
 - More than 273 million covered individuals
 - More than 120 contributing employers and 40 contributing health plans
 - Representation from more than 350 unique carriers
- 





Data from
273M+ patients



120+ Employers



9-12 Medicaid states



40+ Health plans



500+ Hospitals



Electronic Medical
Records

Data included

Eligibility
Claims
Encounters
Rx claims
Benefits plan info
Discharges
Clinical
Productivity
Mortality data*
HRAs

Data preparation

Standardization
Linking
Privacy protection
Enhancements
QA/Improvement
Customization

3 Core Claims Databases

Commercial

- Under-65 working population + dependents
- Claims linked to drug, enrollment, and benefit plan design

Medicare

- Retirees covered by previous employers
- Claims linked to drug, enrollment, and benefit plan design

Multi-state Medicaid

- Medicaid enrollees
- Inpatient services & Rx claims, enrollment, LT care, etc.

Specialty Databases

- Categories include health and productivity, health risk assessment, lab results, linked claims-EMR, hospital drugs, dental, and weather

What data is available

- IBM MarketScan has 3 core databases and 3 specialty databases
 - Three core databases
 - Commercial Claims and Encounters Database (CCAE)
 - Medicare Supplemental and Coordination of Benefits Database(MDCR)
 - Medicaid
 - Three specialty databases
 - Benefit Plan Design (BPD)
 - Health and Productivity Management (HPM)
 - Labs
- The databases that MS has licensed
 - Commercial Claims and Encounters Database (CCAE)
 - Medicare Supplemental and Coordination of Benefits Database (MDCR)

What data is available

- Commercial Claims and Encounters Database (CCAE)
 - Contains
 - Health care coverage eligibility and service use of individuals in plans or product lines with fee-for-service plans and fully capitated or partially capitated plans
 - Represents
 - The medical experience of insured employees and their dependents for active employees, early retirees, Consolidated Omnibus Budget Reconciliation Act (COBRA) continuees, and Medicare-eligible retirees with employer-provided Medicare Supplemental plans.
 - Tables
 - Medical/Surgical
 - Inpatient Admissions (I)
 - Facility Header (F)
 - Inpatient Services (S)
 - Outpatient Services(O)
 - Prescription Drug (D)
 - Enrollment (A,T)
- Medicare Supplemental and Coordination of Benefits Database (MDCR)
 - Represents
 - Medicare-eligible active and retired employees and their Medicare-eligible dependents from employer-sponsored supplemental plans

How to access the dataset

- Certification
 - Mount Sinai HIPPA training
 - CITI Basic Course and Refresher training
- Access
 - You must read and agree and sign the [Data use agreement](#) (you must be logged in with your Minerva ID and password through the Mount Sinai campus network or secure remote VPN)
 - If you don't have a Minerva ID, open a ticket – hpchelp@hpc.mssm.edu
- MarketScan data
 - You can access, analyze and use data only on the server.
 - Data is accessible using SAS application launched from Citrix Workspace.

MarketScan inpatient file

SAS

File Edit View Tools Data Solutions Window Help

Log - (Untitled)

VIEWTABLE: Sasuser.Ccae162

	Sequence Number	Version	Enrolled Family ID	Enrollee ID	Patient Birth Year	Date Year Incurred	Date of Admission	Age of Patient	Case and Services Link	Length of Stay	Date of Discharge	Diagnosis Related Group	Diagnosis Code ICD Version Indicator	Net Payments: Hospital	Payments: Hospital	Coverage Indicator MHSA	Diagnosis Principal	Physician ID	Net Payments: Physician	Payments: Physician	PI
1	441425	20	8421308	.	2016	2016	01/01/2016	0	441425	2	01/03/2016	795 0		1920.44	1920.44		Z3801		0	0	6
2	705139	20	14941066	.	2016	2016	01/01/2016	0	705139	3	01/04/2016	794 0		4203.75	4428.75	1	Z3800	974007491	235.16	235.16	6
3	705143	20	13829204	.	2016	2016	01/01/2016	0	705143	1	01/02/2016	795 0		2526	2526	1	Z3800	114235348	152.95	152.95	6
4	739078	20	34418372	.	2016	2016	01/01/2016	0	739078	774	01/02/2016	774 0		4615.2	6269	1	O411230	942463449	2491.5	2831.25	6
5	739113	20	34418372	.	2016	2016	01/01/2016	0	739113	1	01/02/2016	794 0		1583.14	2228	1	Z3800	847739376	941.5999999	1177	6
6	739123	20	30753179	.	2016	2016	01/01/2016	0	739123	1	01/02/2016	871 0		4216.03	8270.04	0	A419		0	0	9
7	1058813	20	6422742	.	2016	2016	01/01/2016	0	1058813	2	01/03/2016	795 0		1407.4	1760.13	1	Z3800		0	0	8
8	1107932	20	3743647	.	2015	2016	01/01/2016	0	1107932	33	02/03/2016	56 0		210368.73	210368.73		G911		17578.75	17578.75	8
9	1165586	20	31275409	.	2016	2016	01/01/2016	0	1165586	1	01/02/2016	794 0		1683.11	2203.86	1	Z3801	463249211	18.42	123.02	6
10	1168694	20	31949157	.	2016	2016	01/01/2016	0	1168694	3	01/04/2016	792 0		12923.81	14359.79		Z3801		3587.94	3986.6	6
11	1177924	20	31641215	.	2016	2016	01/01/2016	0	1177924	2	01/03/2016	795 0		5086.83	5086.83		Z3800		0	0	6
12	1177932	20	31449569	.	2016	2016	01/01/2016	0	1177932	2	01/03/2016	795 0		3951.07	3951.07		Z3800		176	176	6
13	1327854	20	35978202	.	2016	2016	01/01/2016	0	1327854	1	01/02/2016	795 0		0	414		Z3800		0	0	9
14	441417	20	32183366	.	2016	2016	01/02/2016	0	441417	2	01/04/2016	793 0		2400	3000		Z3801		66.329999999	66.329999999	5
15	739114	20	34458536	.	2016	2016	01/02/2016	0	739114	1	01/03/2016	793 0		0	0	1	Z3800	723485403	0	175.65	6
16	933322	20	40121076	.	2016	2016	01/02/2016	0	933322	68	03/10/2016	856 0		298004.74	298004.74	1	T814XXA		2347.38	2934.12	6
17	1177935	20	31448486	.	2016	2016	01/02/2016	0	1177935	2	01/04/2016	794 0		11139.31	11139.31		Z3800		359.89	359.89	6
18	705130	20	28128018	.	2015	2016	01/03/2016	0	705130	3	01/06/2016	203 0		23670	23920	1	J218	911508495	652	652	4
19	705152	20	13285417	.	2016	2016	01/03/2016	0	705152	2	01/05/2016	793 0		7375.05	7375.05	1	Z3801	170661225	146.75	146.75	4
20	739107	20	33912914	.	2016	2016	01/03/2016	0	739107	2	01/05/2016	791 0		6442.33	8052.85	1	Z3800	523559655	0	296.24	9
21	739244	20	36041746	.	2016	2016	01/03/2016	0	739244	2	01/05/2016	795 0		594.86	363.46	1	Z3800		0	0	6
22	1177940	20	35509169	.	2016	2016	01/03/2016	0	1177940	4	01/07/2016	795 0		7198.57	7198.57		Z3800		0	0	6
23	1177943	20	31839096	.	2016	2016	01/03/2016	0	1177943	3	01/06/2016	795 0		2530.26	2530.26		Z3801		0	0	6
24	1177955	20	31745875	.	2016	2016	01/03/2016	0	1177955	2	01/05/2016	795 0		2842.98	2842.98		Z3800		0	0	6
25	170604	20	6070900	.	1952	2016	01/04/2016	64	170604	2	01/06/2016	470 0		24419.18	26817.47	1	T8484XA		3286.63	3459.62	8
26	528174	20	20658171	.	2016	2016	01/04/2016	0	528174	4	01/08/2016	794 0		3116	3116		Z3801	970076706	256.26	256.26	4
27	698132	20	36636417	.	1997	2016	01/04/2016	18	698132	3	01/07/2016	886 0		3328.88	3328.88		F913		1005.11	1005.11	5
28	705140	20	13381292	.	2016	2016	01/04/2016	0	705140	1	01/05/2016	795 0		8308.66	8308.66	1	P599	369564702	526.04	526.04	6
29	739002	20	35016295	.	2016	2016	01/04/2016	0	739002	1	01/05/2016	795 0		1186.3	1482.87	1	Z3800	109653042	64.5	80.619999999	9
30	739103	20	21629480	.	2016	2016	01/04/2016	0	739103	1	01/05/2016	795 0		560.4	700.5	1	Z3800		84.39	298.77	8
31	739104	20	19237336	.	2016	2016	01/04/2016	0	739104	3	01/07/2016	793 0		2256	2820	1	Z3801		0	233.17	8
32	1058821	20	28701977	.	2016	2016	01/04/2016	0	1058821	3	01/07/2016	794 0		3951.21	4944.1199999	1	Z3801		0	0	8
33	1177933	20	31465308	.	2016	2016	01/04/2016	0	1177933	2	01/06/2016	795 0		4799.92	4799.92		Z3800		242.97	242.97	6
34	1177954	20	31740648	.	2016	2016	01/04/2016	0	1177954	3	01/07/2016	793 0		14520.31	14520.31		Z3801		72.26	245	6
35	1178110	20	34094453	.	2016	2016	01/04/2016	0	1178110	2	01/06/2016	794 0		2797.76	2797.76		Z3800		0	0	6
36	670338	20	29052545	.	2016	2016	01/05/2016	0	670338	2	01/07/2016	795 0		1969.59	2188.43		Z3800		45.3	221.2	2
37	698134	20	33391900	.	2016	2016	01/05/2016	0	698134	2	01/07/2016	794 0		3988.91	3988.91		Z3800		244.03	244.03	5
38	705141	20	16119988	.	2016	2016	01/05/2016	0	705141	5	01/10/2016	794 0		2333.59	2892.85	1	P9683	722627155	282.8	314.21	6
39	705142	20	13412786	.	2016	2016	01/05/2016	0	705142	2	01/07/2016	795 0		3038.43	3376	1	Z3800	391958727	281.13	468.55	8
40	738924	20	40026537	.	2016	2016	01/05/2016	0	738924	7	01/12/2016	208 0		0	0	1	T797XXA	268401157	0	0	5
41	739051	20	30046823	.	2016	2016	01/05/2016	0	739051	4	01/09/2016	392 0		14483.32	16394.9	1	K589	831471728	428.4	428.4	6
42	739145	20	34882135	.	2016	2016	01/05/2016	0	739145	2	01/07/2016	795 0		1087.9	1359.88	1	Z3800		92.409999999	115.51	8
43	739170	20	36039954	.	2016	2016	01/05/2016	0	739170	2	01/07/2016	795 0		1225.42	1872.34	1	Z3800		214.4	268	6
44	887626	20	15881788	.	1991	2016	01/05/2016	25	887626	3	01/08/2016	885 0		5121.21	5886.26	1	F29		0	0	6
45	887630	20	18785021	.	1998	2016	01/05/2016	17	887630	7	01/12/2016	331 0		52981.7	52981.7	1	K50018		2984.78	2984.78	4
46	1042209	20	33555637	.	2016	2016	01/05/2016	0	1042209	1	01/06/2016	644 0		-0.01	0.01	1	C73		0	0	4
47	1042234	20	25677934	.	2016	2016	01/05/2016	0	1042234	4	01/09/2016	603 0		6400.54	6400.54	1	L03114		-220.42	-220.42	4

Output - (Untitled) Log - (Untitled) Editor - Untitled1 VIEWTABLE: Sasuser.C...

C:\Program Files\SASHome\SASFoundation\9.4

Contents of 'Sasuser'

Ccae133 Ccae143 Ccae153 Ccae162 Ccae171

Ccae181 Ccae192_a Ccae201_a Ccae211_a Ccae133

Ccae143 Ccae153 Ccae162 Ccae171 Ccae181

Ccae192_a Ccae201_a Ccae211_a Ccae133 Ccae143

Ccae153 Ccae162 Ccae171 Ccae181 Ccae192_a

Ccae201_a Ccae211_a Ccae133 Ccae143 Ccae153

Ccae162 Ccae171 Ccae181 Ccae192_a Ccae201_a

Ccae211_a Ccae133 Ccae143 Ccae153 Ccae162

Ccae171 Ccae181 Ccae192_a Ccae201_a Ccae211_a

Ccae133 Ccae143 Ccae153 Ccae162 Ccae171

Ccae181 Ccae192_a Ccae201_a Ccae211_a Ccae133

Ccae143 Ccae153 Ccae162 Ccae171 Ccae181

Ccae192_a Ccae201_a Ccae211_a Ccae133 Ccae143

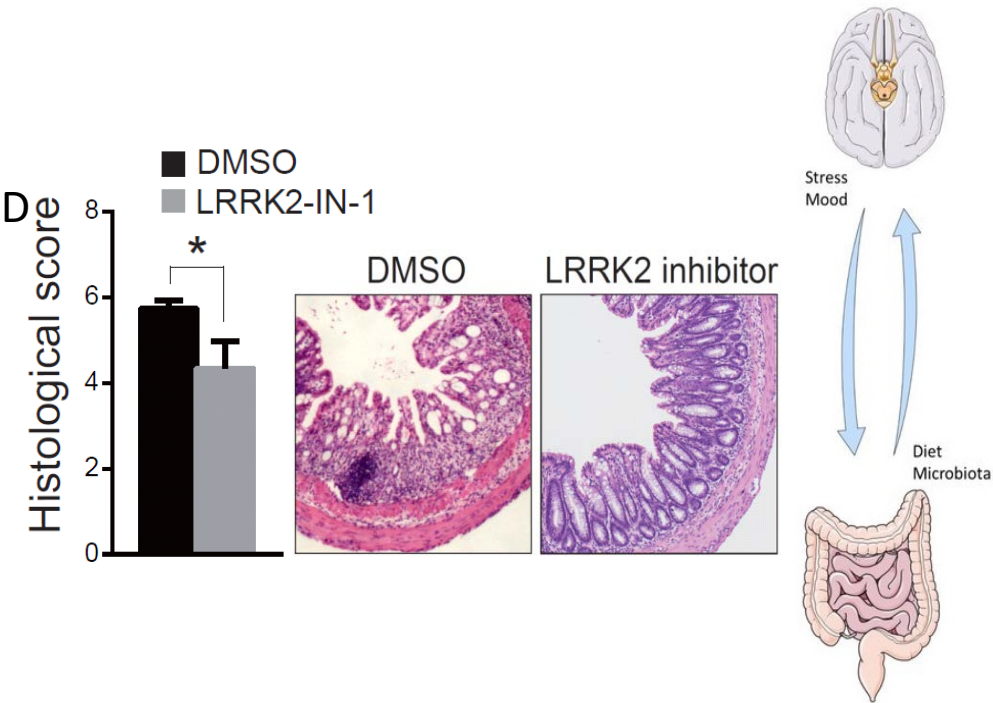
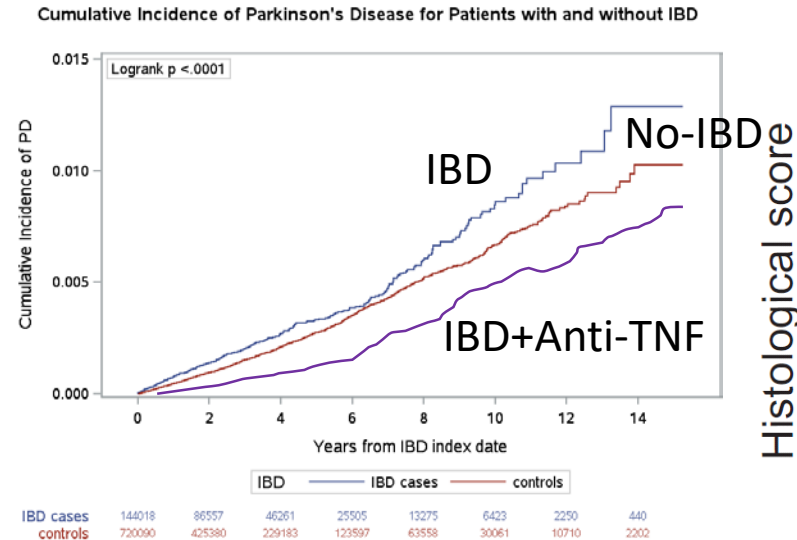
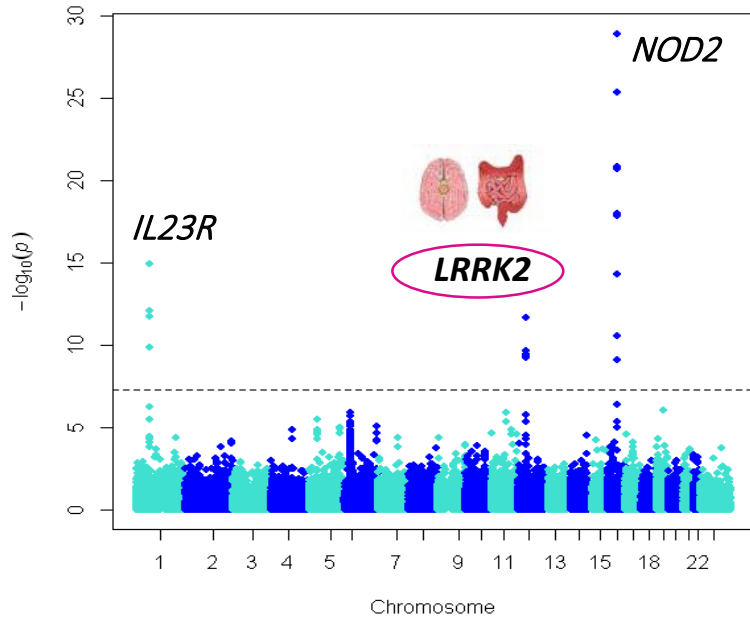
How to use it

- Access
 - MarketScan users are encouraged to use Minerva and save processed data in their project or home folders for further operations including downloading, which is indifferent to workflows adopted by other Minerva users.
 - The downloading restriction some users experienced was related to the Windows-based server, which is non-expandable in computational resources and is destined for decommissioning.
 - The Data Ark team increased the accessibility of MarketScan data by providing guaranteed computation resources on Minerva, which is unparalleled by the Windows-based server historically serving MarketScan data.

Skills needed to work with the dataset

- Proficiency in any major programming language (R, SAS)
 - Used to be only SAS programming
- Basic knowledge of medical codes
 - The International Classification of Diseases (ICD) code
 - Current Procedural Terminology (CPT) code
 - Diagnosis-related groups (DRGs)
 - The CMS Healthcare Common Procedural Coding System (HCPCS) procedure codes
 - Etc.
- Basic knowledge of format/elements of the databases
 - How variables are formed
 - How the files are connected

From Genetic Association to New Therapies for IBD and PD



Genetic link: Variation in the *LRRK2* gene, the strongest genetic risk of Parkinson's disease, is associated with risk for IBD

Clinical link: Patients with IBD are at higher risk of developing Parkinson's disease unless treated with anti-TNF

Pharmacological link: *LRRK2* inhibitors mitigate experimental colitis and reduce inflammatory biomarkers in human samples – *future potential treatment for IBD*



MERATIVE MARKETSCAN RESEARCH DATABASES

Data assets for government, non-profit, and academic research

Merative MarketScan Research Databases can
help researchers:

- Conduct comparative effectiveness research
- Gauge the effect of health policy
- Calculate the total burden of illness
- Improve healthcare treatment guidelines
- Inform business and public healthcare policy
- Benchmark best practices
- Articulate therapeutic value

Dynamic research quality

The Merative™ MarketScan® Research Databases contain de-identified records for more than 273 million patients since 1995, in order to provide a patient-centric perspective reflecting real-world treatment patterns and the cost of care. Data are characterized by strong longitudinal integrity, depth of clinical detail, and unique data elements that address laboratory test results, dental care, hospital drugs, clinical data, health and productivity data, self-reported health risks, and benefit plan design information..

Among the most published databases in the US

Launched in 1989, the MarketScan Research Databases have served as the basis for more than 2,700 peer-reviewed articles since 1990, making a substantial contribution to the body of literature used to formulate policy decisions and improve healthcare for Americans. Researchers can utilize the MarketScan Research Databases to access information to help them understand disease progression, treatment patterns, health outcomes and the associated costs to patients, employers, and insurers.

- Nationally representative: age, gender, geography and type of coverage
- Linked to absenteeism and disability records, lab test results, dental claims and health risk assessments
- Continuous enrollment over multiple years with records since 1992
- Robust cohort sizes for detailed analyses
- Payment details, including copayments
- Demographic, clinical, utilization and financial profiles of patients
- Individual patients linked to their actual claims
- Data from providers of care: inpatient, outpatient, outpatient pharmacy, mail order and specialty pharmacy
- Detailed plan and eligibility descriptions
- Statistically de-identified to meet HIPAA privacy requirements
- Online access to time-saving information and tools

Contributing to more than 2,700 published studies over 30 years.

Databases

The privately insured population

The Merative™ MarketScan® Commercial Database provides researchers with in-depth, cross-sectional and longitudinal views of healthcare practices and costs for the American working population and their dependents. Fully adjudicated, patient-level claims are linked with other MarketScan Research Databases. Linking data at the patient level can help enhance the insights derived from claims-based research.

The Medicare population

For the fast-growing, over-65 age group, retrospective studies and forecasting analyses may be conducted using the Merative™ MarketScan® Medicare Database. The database includes the employer-paid portion and out-of-pocket patient expenses for both Advantage and Supplemental plans, as well as the Medicare-covered portion of payment (represented as Coordination of Benefits Amount or COB) for Supplemental plans. The inclusion of both Advantage and Supplemental data into one database provides a more representative, complete, and longitudinal view of the commercially-insured age 65+ U.S. population, and creates opportunities for unique analyses, such as comparing costs and utilization between Advantage and Supplemental enrollees.

The Medicaid population

The Merative™ MarketScan® Multi-State Medicaid Database pools claims from 47 million Medicaid enrollees from 8-12 geographically dispersed states. In addition to inpatient, outpatient, drug and enrollment files that are similar to those found in the other MarketScan Research Databases, the MarketScan Multi-State Medicaid Database includes a long-term care claims file that captures nursing home stays, home healthcare and long-term care services. In addition to standard demographic variables, such as patient age and gender, this database includes variables that may be of particular value to researchers investigating Medicaid populations, such as aid category (for example, blind or disabled, Medicare eligible) and race.

Health and productivity

Researchers have used data from the MarketScan Research Databases to quantify the total burden of illness (direct and indirect costs) and calculate the return on investment for insurer-and employer- based demand management initiatives. The Merative™ MarketScan® Health and Productivity Management Database contains data on workplace absence, short-term and long-term disability and workers' compensation dating back to 1997 and is directly linked to individual patients' de-identified medical and drug claims.

Health risks and behaviors

When you need to understand the relationship between patient behavior and health outcomes, health risk assessment (HRA) data can be helpful. HRA data provide self-reported information on clinical variables that may otherwise be unavailable. Like other

MarketScan Research Databases, the Merative™ MarketScan® Health Risk Assessment Database standardizes and links HRA data with the claims experience of patients – presenting an opportunity for innovative research. The MarketScan Health Risk Assessment Database has significant overlap with the MarketScan Health and Productivity Management Database and can be used to enhance health and productivity studies.

Lab test results

Merative™ MarketScan® Lab Database clinically enriches the medical and prescription drug data of millions of patients already in the MarketScan Research Databases by linking patients' claims data with lab test results obtained from large independent laboratories. These test results allow you to measure outcomes using clinical guidelines and markers, which can help you to better understand how well a drug is performing in a real-world clinical setting and how differences in treatment patterns can potentially affect clinical outcomes.

Dental care and medical conditions

By linking dental and medical/pharmacy claims, the Merative™ MarketScan® Dental Database allows researchers to understand the potential relationship between dental care and medical conditions. This is one of the only research databases of its kind that allows researchers to follow the entire continuum of healthcare, including services received in the dental office.

One of the largest
claims databases
available for
licensing, with more
than 273 million
unique patients since
1995.

Sample data elements

Demographic information

- Enrollee identifier
- Relation to primary insured employee
- Age of patient
- Gender of patient
- Industry group
- Employment classification
- Employment status
- Census region
- State
- Metropolitan statistical area

Insurance features

- Plan type
- Coverage limitations
- Maximum out-of-pocket

Financial information

- Net payments
- Coordination of benefits
- Capitation payments
- Total admission payments
- Payments to facilities accounts and providers
- Deductible and copayment amounts

Inpatient and outpatient medical information

- Admission date/service date
- Admission type/service type
- Diagnosis related group
- Discharge status
- Length of stay
- Major diagnostic category – Place of service
- Provider identification – Provider type
- Primary/secondary diagnosis and service codes

Drug information

- Average wholesale price – Days supplied
- Generic product ID
- National drug code
- Prescription drug payment – Refill number
- Therapeutic class
- Retail/mail order

Enrollment information

- Date of enrollment/ disenrollment
- Member days
- Payments to facilities accounts and providers
- Deductible and copayment amounts

Additional data elements for health and productivity

Absenteeism information

- Dates/hours of absence
- Absence Type (sick, leave, disability, recreational, etc.)

Short-term disability information

- Case days
- Disability type
- Total Payments

Workers' compensation information

- Body part injured
- Case days
- Cause of injury
- Indemnity payments – Medical payments

Data Licensing Options

- Complete years of data are available for purchase by university, non-profit organizations, and Federal, State, and local agencies
- Custom subsets are available, developed for specific grant - or internally funded studies to reflect specific populations or data years of interest
- Treatment Pathways, an online analysis tool, is available to support researchers and provides an interactive solution for hypothesis-driven investigations

Research Services Options

- Merative analysts, programmers, and subject matter experts are available to collaborate, design, and execute research studies



Learn more

Visit our [website](#) or [schedule](#) time with us to talk about how real-world data from MarketScan can help you generate valuable insights.

About Merative

Merative is a data, analytics and technology partner for the health industry, including providers, payers, life sciences companies and governments. With trusted technology and human expertise, Merative works with clients to drive real progress. Merative helps clients reassemble information and insights around the people they serve to improve healthcare delivery, decision making and performance. Merative, formerly IBM Watson Health, became a new standalone company as part of Francisco Partners in 2022. Learn more at merative.com.

© Copyright Merative 2022

Merative
100 Phoenix Drive
Ann Arbor, MI 48108

Produced in the United States of America, June 2022.

Merative, the Merative logo, and merative.com are trademarks of Merative, registered in many jurisdictions worldwide. Other product and service names might be trademarks of Merative or other companies. The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. Merative products are warranted according to the terms and conditions of the agreements under which they are provided.

MSN-3207081496 Rev 1.0