How can AI Meaningfully Improve Diversity in Clinical Trials?

Arnaub K. Chatterjee
Senior Vice President - Medidata AI
Teaching Associate, Harvard Medical School

January 31, 2023

Medidata AI® Builds End-to-End Solutions to Advance Clinical Development

UNIQUE DATASETS

Clinical & Operational

~27K trials (7K ongoing),
~8M patients,
30K facilities, 107 countries

Real World

450M cases year
from 10,000 sites,
300+ RWD sources

PLATFORM & ANALYTICS

- Intelligent Trials
  - Study Feasibility
  - Performance Analytics
- Integrated Evidence
  - Synthetic Control Arm
  - Trial Design
  - Medidata Link
- Commercial Data Solutions
- Connected Patient

INSIGHTS & LABS

Industry Experts, Clinicians, Former Regulatory Officials, Data Scientists, Biostatisticians

Medidata AI Has the World's Largest Historical Clinical Trial Database, with 2 Distinct Flavors of Historic Clinical Trial Data

29K+ Studies

~8M Patients

Medidata Archive

Operational Data

Study, country and site level data with associated investigator names

35+ metrics spanning enrollment, quality, congestion and cycle times

True performance metrics based on actual events with real-time data capture

Example use cases:

- Analyze trial landscape in real-time
- Identify and rank countries and sites
- Forecast and re-forecast enrollment
- Generate ongoing, live study insights

Clinical Data

Study and patient-level data

100+ clinical fields, covering

- Primary & secondary outcomes
- Study & concomitant treatments
- Inclusion and exclusion criteria
- Adverse events

Rapid, precise meta-analysis at the subject level

Example use cases

- Synthetic Control Arm®
- Select patient populations
- Select comparators
- Validate endpoints
- Simulate outcomes using "virtual patients"

MEDIDATA

DASSAULT SYSTEMES

How can Medidata AI Address Diversity Objectives?

Leverage this rich and granular dataset to **help reduce disparity in terms of diversity** in clinical trials and **improve science and outcomes.**

Baseline and benchmark the diversity of trials in a specific indication

Identify sites that can accelerate a trial AND are more likely to enroll diverse patient populations

Site level patient demographics available to measure diversity

Age | **Sex** | **Race** | **Ethnicity**

Medidata/PPD Collaboration - TrueCast Diversity Data — Site Level

Knowing WHO and WHERE top recruiters of diverse patients are

Indication: Alzheimer's

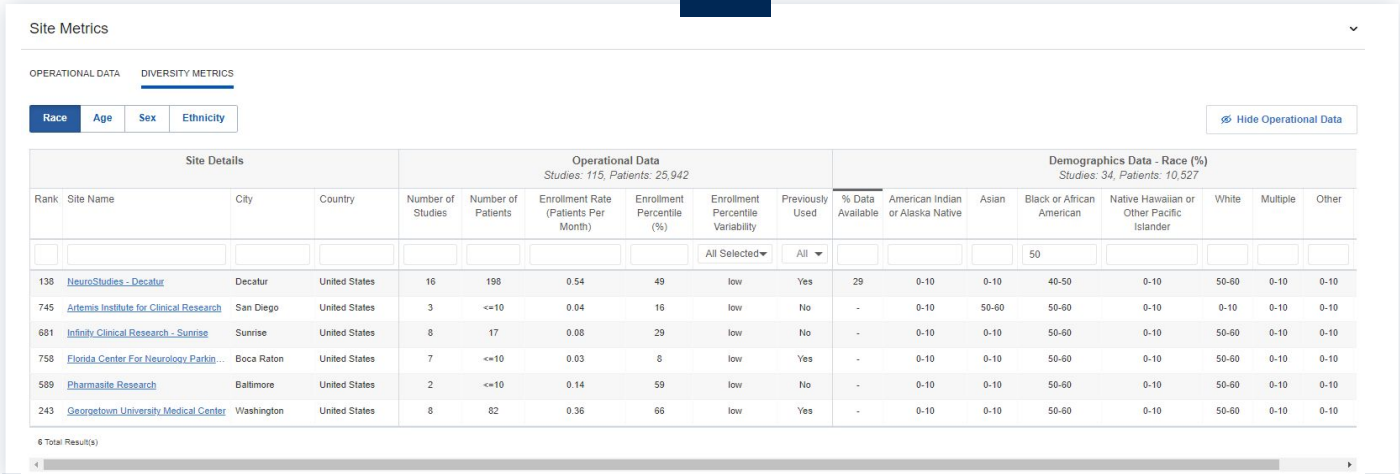
Diversity Metrics: Sex, Age, Race, Ethnicity

# of Industry Studies: 34

# of Sites: 500+

# of Patients: 10,500+

# of Countries: 40+



Site Metrics

OPERATIONAL DATA | DIVERSITY METRICS

Race | Age | Sex | Ethnicity

Hide Operational Data

Site Details				Operational Data (Studies: 115, Patients: 25,942)						Demographics Data - Race (%) (Studies: 34, Patients: 10,527)							
Rank	Site Name	City	Country	Number of Studies	Number of Patients	Enrollment Rate (Patients Per Month)	Enrollment Percentile (%)	Enrollment Percentile Variability	Previously Used	% Data Available	American Indian or Alaska Native	Asian	Black or African American	Native Hawaiian or Other Pacific Islander	White	Multiple	Other
								All Selected	All				50				
138	NeuroStudies - Decatur	Decatur	United States	16	198	0.54	49	low	Yes	29	0-10	0-10	40-50	0-10	50-60	0-10	0-10
745	Artemis Institute for Clinical Research	San Diego	United States	3	<=10	0.04	16	low	No	-	0-10	50-60	50-60	0-10	0-10	0-10	0-10
681	Infinity Clinical Research - Sunrise	Sunrise	United States	8	17	0.08	29	low	No	-	0-10	0-10	50-60	0-10	50-60	0-10	0-10
758	Florida Center For Neurology Parkin...	Boca Raton	United States	7	<=10	0.03	8	low	Yes	-	0-10	0-10	50-60	0-10	50-60	0-10	0-10
589	Pharmasite Research	Baltimore	United States	2	<=10	0.14	59	low	No	-	0-10	0-10	50-60	0-10	50-60	0-10	0-10
243	Georgetown University Medical Center	Washington	United States	8	82	0.36	66	low	Yes	-	0-10	0-10	50-60	0-10	50-60	0-10	0-10

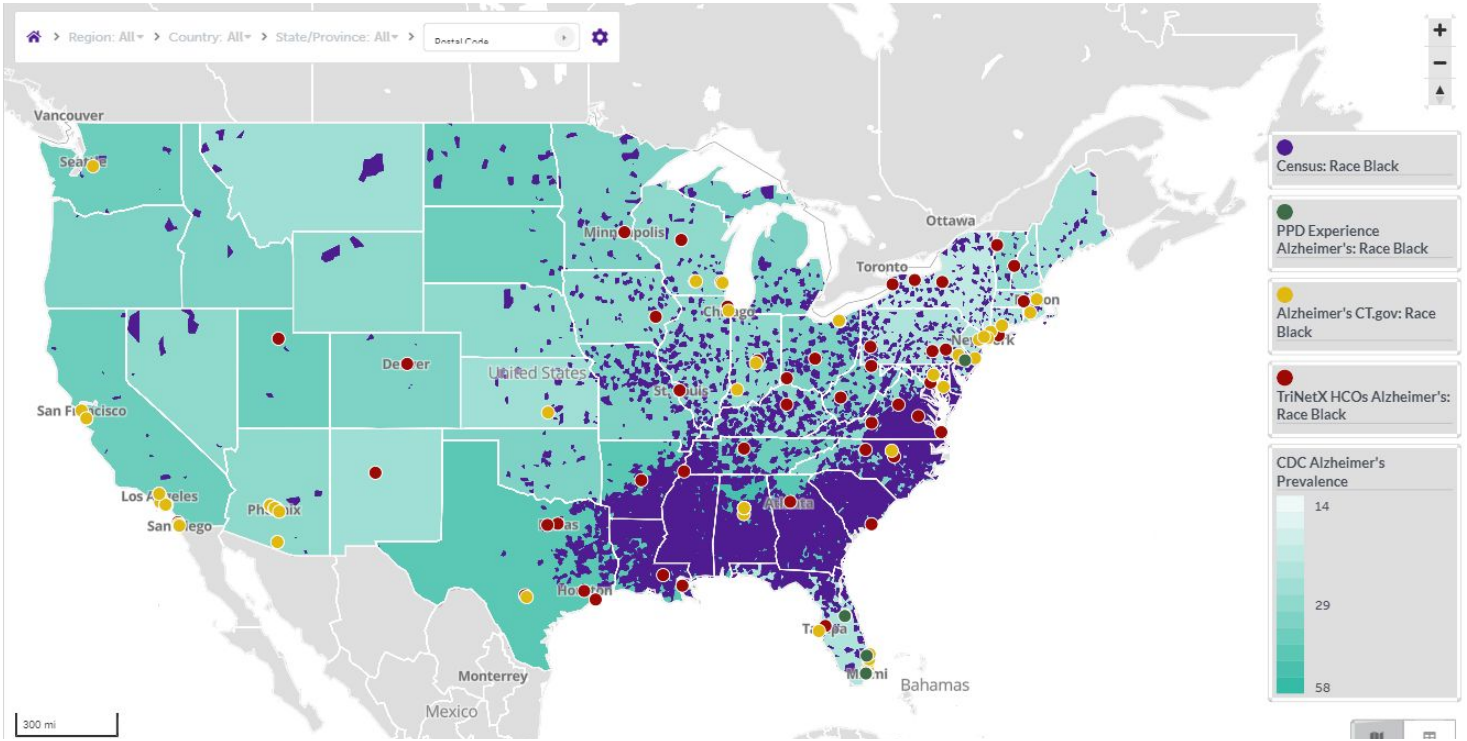
6 Total Result(s)

Power is in volume and granularity of site performance diversity data

MEDIDATA

DASSAULT SYSTEMES

Integrating Data to Drive More Representative Trials

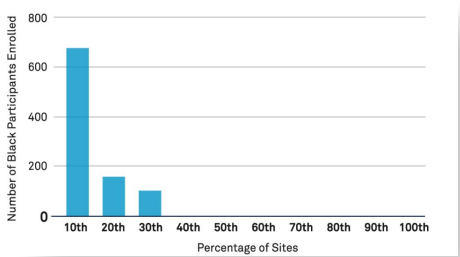


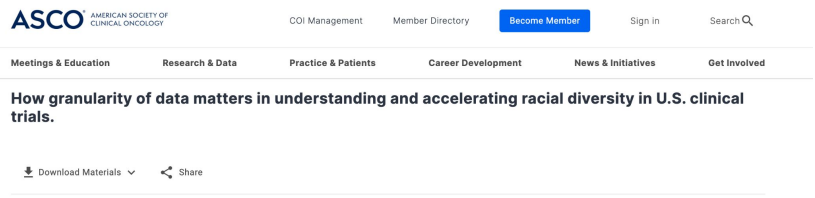
Example: Diverse patient recruitment is uneven across sites

Based on Medidata AI analysis across 4,000 trials and ~1M patients in last 10 years

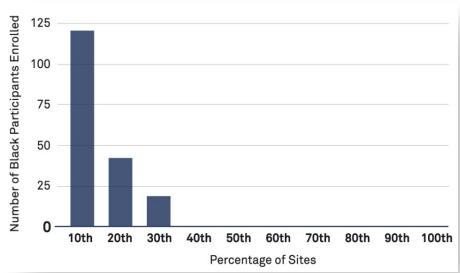
Black patients enrolled in only 30% of lung cancer and Alzheimer's sites in the US

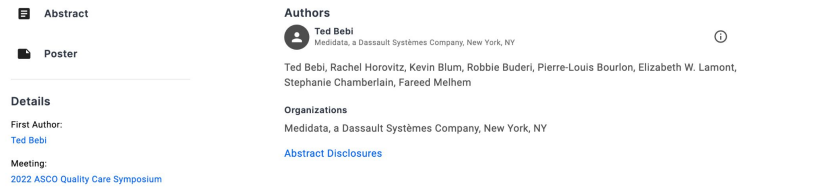
Lung Cancer





Alzheimer's Disease





MEDIDATA

DASSAULT SYSTEMES

Example: Association Between State Medicaid Policies and Accrual of Black or Hispanic Participants to Cancer Clinical Trials (in submission)

- The Medicaid program plays a key role in the financing of health care for low-income patients and may serve as a critical policy lever to improve enrollment of racial and ethnic minority patients in cancer clinical trials

- A team from Medidata, Cornell University and the University of Pennsylvania studied secular changes in enrollment of Black patients in Eastern Cooperative Oncology Group trials within 17 states before and after state-mandated clinical trial coverage over 2000-2019

- State-mandated Medicaid coverage of the routine costs of trial participation was associated with a short-term increase in the proportion of Black trial participants

Example: Regulatory Application of a Synthetic Control Arm

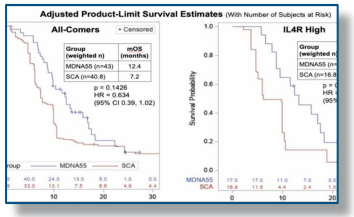
MEDICENNA™

Early in Development

Late Stage and Regulatory Use

Health Authority and Payer Discussions

- For the first time, the FDA approved the use of hybrid SCA in a Phase III study design in Recurrent Glioblastoma (rGBM)
- Medicenna used a Medidata SCA to demonstrate efficacy in a single-arm Phase II trial, with the goal of using the results to design a Phase 3 SCA-based registrational trial
- The SCA demonstrated large efficacy effect size in all-comers subjects, as well as in subjects expressing high levels of IL4R.



The FDA's acceptance of this unique design, will expedite completion of the Phase 3 trial in rGBM allowing earlier access of MDNA55 for a disease with poor prognosis and high unmet need," said **Dr. Fahar Merchant, President and CEO of Medicenna**.

(October 15, 2020) **The full press release can be viewed here.**

Example: Using AI and synthetic patient generation to explore historical trials and refine I/E criteria

SITUATION:
Tyrosine kinase inhibitors (TKIs) treat certain advanced malignancies (e.g., Non-small Cell Lung Cancer) with high efficacy when targeted driver mutations are present

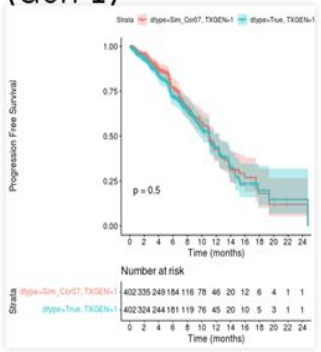
A leading biotech seeking to develop a new TKI sought to enable its data scientists to explore historical trial data to

- **Understand why some trials show higher effects** than others for the same drug(s)
- **To support phase 3 trial design decisions** (e.g., choice of geography, comparator, sub-population)

WHAT WE DID:
Built a Source clinical trial dataset from multiple studies, then created a suite of Simulant datasets, including

- **Very large datasets** to enable deep learning by the client's data scientists
- **Oversampled datasets** to amplify & study the interaction of race & geography
- **Multiple replicates** of all datasets for robust validation of all insights

Comparison with simulants Data Set: PFS (Gen 1)



	dtype=Sim_Cor07, TXGEN=1	dtype=True, TXGEN=1
0.5 years	0.77 (0.73 0.82)	0.72 (0.67 0.77)
1 years	0.43 (0.36 0.51)	0.44 (0.37 0.52)
1.5 years	0.18 (0.1 0.32)	0.2 (0.12 0.33)
2 years	0.12 (0.05 0.27)	0.15 (0.07 0.32)
Median survival time (days)	336 (313,377)	336 (296,379)

How can AI address diversity in enrollment and what can success look like?

- Create trust and authenticity in the data
- Create trust in the methodology - focus on generalizability
- New methods such as external control arms and synthetic patients can address patient burden, simulate outcomes, power enrollment and de-risk trials

MEDIDATA

DASSAULT SYSTEMES