Investigating Demographic and Socioeconomic Representation in Stanford's NIH RECOVER Cohort

Dena G. Liu[1], Aditi Mohite[1], Donna Williams[1], Kelly Olszewski[1], Ting Chen[1] , Chris Jamero[1], Zenita Leang[1], Mary Rithu Varkey[1], Maya Caple[1], Khaliun Gombojav[1], Ruth Talamoa[1], Minjoung Go[1,2], Andre Kumar[1,2]

[1]Department of Medicine, Stanford Center for Clinical Research, [2]Department of Medicine, Division of Hospital Medicine
Stanford University, Palo Alto, CA, USA

Stanford MEDICINE | Center for Clinical Research

BACKGROUND

COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), became a global pandemic in 2020.

- In 2024, **Long COVID** was defined as a chronic condition occurring at least three months post-infection, presenting as a continuous, relapsing, or progressive disease state affecting multiple organ systems.

Health Equity Challenges:

- Low-income populations and communities of color face higher risks of negative health outcomes from COVID-19 and Long COVID. Historically, their underrepresentation in research has limited universal applicability of treatments.
- In California:
  - 26% of households earned < $50,000 annually (2022).
  - Adults below the federal poverty level are twice as likely to experience Long COVID symptoms compared to adults with income levels 300% and above the FPL.

The NIH RECOVER (Researching COVID to Enhance Recovery) Initiative:

- A nationwide, multi-site observational study investigating post-acute sequelae of SARS-CoV-2 infection (PASC).
- Aims to support the development of diagnostic and therapeutic interventions for a diverse U.S. study population.
- Participants receive advanced diagnostic tests at no cost.
- Stanford, one of 86 RECOVER sites, contributes 8% of the national adult enrollment (1,022 of 12,561 participants).

Research Aims:

- Hypothesis: The demographic distributions of Stanford RECOVER participants differ significantly from state-level COVID-19 data.
- Examine income levels and insurance status to evaluate socioeconomic diversity in the Stanford RECOVER cohort.

METHODS

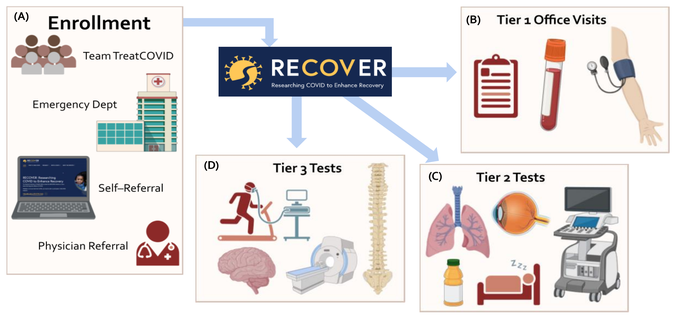

Figure 1. Overview of Participant Enrollment and Involvement in the NIH RECOVER Initiative.
(A)Participants were enrolled through the TreatCOVID team, physician referrals, the emergency department, and self-referrals until September 2023.
(B)Initial evaluations are conducted.
(C)Additional diagnostic tests, such as ECGs, chest CTs, neuropathy assessments, sleep studies, and vision tests may be triggered.
(D)Higher burden diagnostics, including lumbar punctures, skin biopsies, exercise testing, brain MRI, and neurocognitive testing may also be triggered.

Data Sources:

- **RECOVER Study Data:** Participant demographics and socioeconomic data were extracted from REDCap.
- **State-Level Data:** COVID-19 demographics and diagnostic testing costs were obtained from the California Department of Public Health and the California Health Care Access and Information site.

Analysis:

- Descriptive and chi-square analyses summarized demographics and compared cohort distributions to state-level data.
- Analyses were conducted using R and Excel.

RESULTS

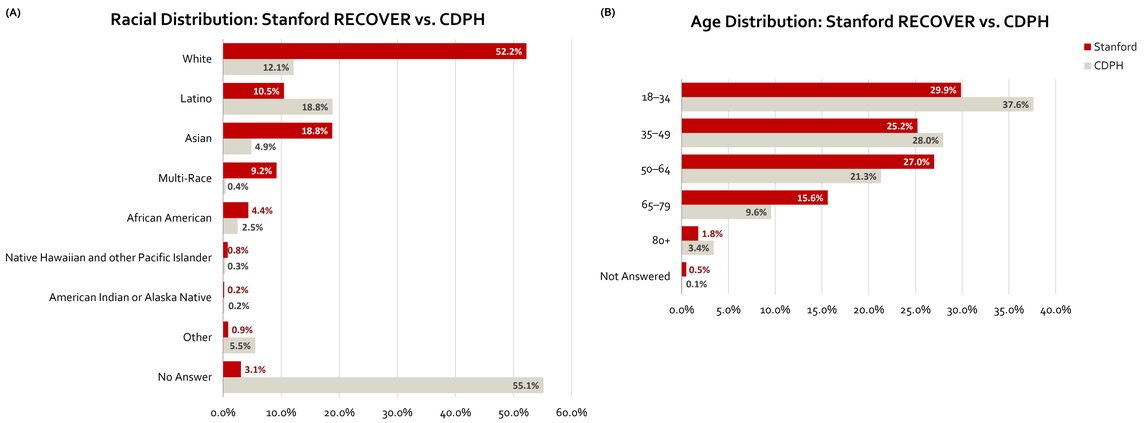

Figure 2. Demographic Distribution of Stanford RECOVER Participants (*N* = 1011) Compared to California State Data (CDPH) (*N* =16,825,002).
(A)Racial distribution: Stanford has a higher proportion of White and Asian groups as compared to state data (52.2% vs. 12.1% , 18.8% vs. 4.9% respectively), while Latino participants are underrepresented (10.5% vs. 18.8%). 'Other' includes smaller racial groups and unclassified responses.
(B)Age distribution: Older participants (50+ years) represent 44.4% of Stanford's cohort whereas younger participants (18–34) only account for 29.9%, inverse from state data.

Figure 3: Race and Age Group Representation: Stanford RECOVER Cohort vs. California State Data. Chi-square tests compared race and age group distributions. Smaller race groups ('American Indian or Alaska Native' and 'Native Hawaiian and Other Pacific Islander') were combined into 'Other,' and responses of 'Prefer not to answer' or missing data were excluded to meet chi-square test assumptions. Significant differences were found for **race** (χ^2 (5, N = 970) = 1360.8, p < 0.001) and **age** (χ^2 (3, N = 970) = 47.4, p < 0.001).

The heatmap shows **standardized residuals**, highlighting **overrepresentation** (red) or **underrepresentation** (blue) by race and age group. Darker shades indicate larger deviations from expected values, while lighter shades suggest closer alignment.

Table 1. Test Counts and Average Out-of-Pocket Costs for Tier 2/3 Tests

Test Type	Tests Performed	Average Out-of-Pocket Cost ($)
Tier 2		
Hepatitis B & C Test	329	$150
Vision Test	331	$150
Oral Glucose Tolerance Test	34	$200
UPSIT Smell Test	244	$200
6 Minute Walk Test	332	$250
MINI Neuropsychiatric Interview	140	$250
Chest CT	159	$275
ACTH & Cortisol Test	296	$350
Renal Ultrasound	60	$400
Pulmonary Function Test	173	$450
Fibroscan	105	$600
NIH Toolbox	36	$600
Home Sleep Study	95	$700
Neuropathy	159	$750
Echocardiogram	142	$800
Tier 3		
Comprehensive Audiometry	95	$250
Gastric Emptying Study*	3	$1,000
Cardiopulmonary Exercise Test	17	$1,200
Brain MRI	63	$1,500
Lumbar Puncture	2	$1,500
Skin Biopsy*	0	$2,000
Cardiac MRI*	1	$3,000
Complete Neurocognitive Test	12	$3,500

*Note: Average out-of-pocket costs were obtained from the California Department of Health Care Access and Information available Health Fare Payments Data (HPD) . *Indicates tests recently operationalized.*

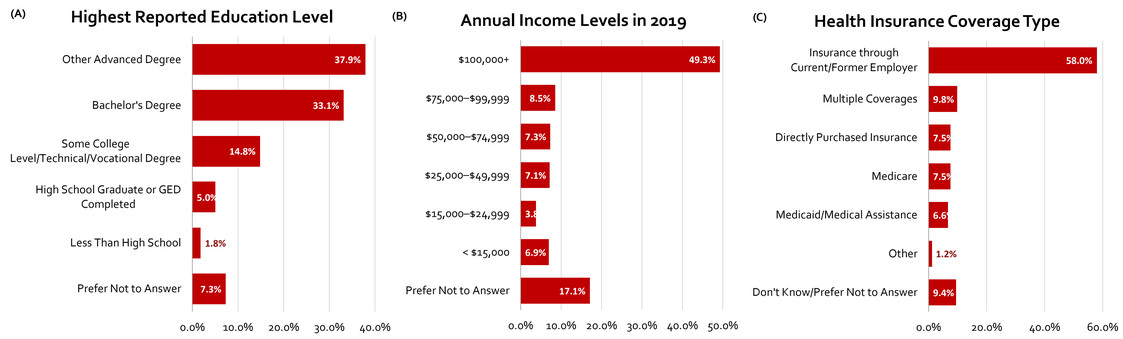

Figure 4. Socioeconomic and Educational Factors in RECOVER Participants at Stanford (*N = 1011*).
(A)Education: 71.0% of participants hold a bachelor's degree or higher, while less than 2% reported not completing high school, including one participant with no formal schooling. Advanced or professional degrees were the most reported (37.9%).
(B) Income (2019): Nearly half (49.3%) earned $100,000 or more, while 17.8% earned less than $50,000.
(C)Insurance: Most participants (58.0%) were insured through an employer. Public insurance (Medicare/Medicaid) covered 14.1%. 'Other' included coverage through Veteran Affairs (n = 2), TRICARE (n = 3), or 'Self-pay' (n = 7), accounting for a total of 1.2%.

CONCLUSIONS

Demographic and Socioeconomic Representation

- The cohort is notably overrepresented by White participants across all age groups and Asian participants aged <50 years.
- Latino participants are notably underrepresented across all age groups.
- The majority of participants have private insurance and higher income levels.
- Educational attainment is nearly twice as high as the state average for adults aged ≥25 years (71% vs. 35.9%).

Health Equity Implications

- Demographic disparities highlight the challenges of achieving equitable representation in studies and emphasize the need for targeted outreach to underrepresented groups.
- RECOVER not only reveals key insights into Long COVID but also uncovers critical incidental findings—helping participants discover and address previously unknown health conditions with their providers.
- By covering laboratory and advanced testing costs, the study reduces financial barriers, particularly for high-cost assessments.

Limitations

- Missing racial data (55.14%) in the CDPH dataset hinders direct comparisons, reflecting early pandemic data collection challenges.
- CDPH data from May 2023 may not fully reflect current demographic trends.
- Combining smaller racial categories to meet statistical requirements limits specific insights.

Future Directions

- Focus on including underrepresented populations.
- Reduce logistical and financial barriers, such as transportation and work schedules, by offering flexible clinic hours for low-wage workers.
- Analyze demographic shifts over time to evaluate changes in engagement and retention.

Acknowledgments: This research was supported by the National Institutes of Health (NIH) under the RESEARCHING COVID TO ENHANCE RECOVERY (RECOVER) Initiative, Prime Award No. OT2HL161847-01, Subaward No. ADU-15-21. We are grateful to the participants and study sites involved in the NIH RECOVER study, as well as our colleagues and collaborators for their invaluable contributions. Special thanks to Divya Pathak and Sumana Shashidhar for their guidance and support in this project, and to Ma Fides Mabunay for their operational support. We also acknowledge Dr. Upinder Singh for her foundational leadership as the former PI and the sub/co-investigators of the RECOVER study, including Dr. Andra Blomkalns, Dr. Catherine Blish, Dr. Hector Bonilla, Dr. Yvonne (Bonnie) Maldonado, Dr. Linda Geng, Dr. Jake Scott, Dr. Karen Jacobsen, Dr. Xiaolin (Kathleen) Jia, Dr. Prasanna Jagannathan, Dr. Neera Ahuja, Dr. Orlando Quintero, Dr. PJ Utz, Dr. Alfredo Urdaneta, and Dr. Hannah Valantine, for their significant contributions to the study.