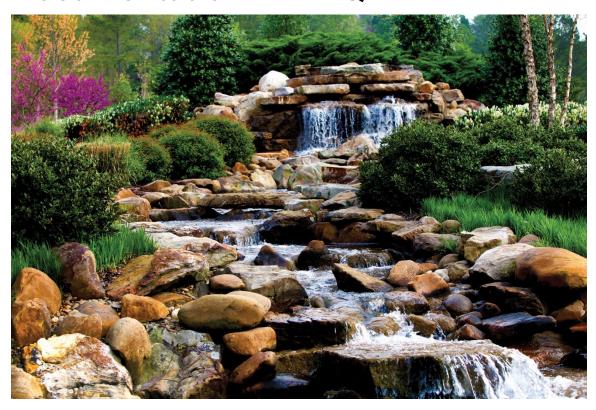# VA EDH Data Curation Documentation FY22-Q4

Hilda B. Klasky
Kevin Sparks
Jeremy Logan
Alec Hamaker
Matthew Whitehead
Heidi Hanson
Rochelle Watson
Anuj Kapadia

September 2022

OAK RIDGE
National Laboratory

Computational Sciences & Engineering Division

VA EDH DATA CURATION DOCUMENTATION FY22-Q4

Hilda B. Klasky
Kevin Sparks
Jeremy Logan
Alec Hamaker
Matthew Whitehead
Heidi Hanson
Rochelle Watson
Anuj Kapadia

September 2022

CONTENTS

1. INTRODUCTION

The health and well-being of the Nation's men and women who have served in uniform is the highest priority for the U.S. Department of Veterans Affairs (VA). VA is committed to providing timely access to high-quality, recovery-oriented, evidence-based mental health care that anticipates and responds to Veterans' needs and supports the reintegration of returning Service members into their communities. VA is working to eliminate suicide among all Veterans by developing and implementing innovative suicide prevention approaches and resources.

Health outcomes, such as suicide are typically modeled as a function of genetics and environment, where environment refers to factors beyond medical, e.g., air quality, access to transportation and food, homelessness status, etc. Mental health outcomes for each individual are considered to be associated with multiple stressors that fall under a variety of categories socioeconomic, economic, physical environment. Understanding the relationships between these stressors, covariates, and health outcomes, requires curated, standardized data that can be input into the VA's Recovery Engagement and Coordination for Health - Veterans Enhanced Treatment (REACH VET) or other health outcomes model.

Environmental Determinants of Health (EDH) as defined by the World Health Organization (WHO) is clean air, stable climate, adequate water, sanitation and hygiene, safe use of chemicals, protection from radiation, healthy and safe workplaces, sound agricultural practices, health-supportive cities and built environments, and a preserved nature are all prerequisites for good health.

1.1 BACKGROUND

With funding from the VA Office of Mental Health and Suicide Prevention (OMHSP), the EDH project has developed novel datasets associated with select health outcomes, a methodology for converting spatiotemporal data from one spatial reference (such as a 1km grid) to another (such as US Census Tracts), and health outcomes modeling capabilities. The datasets are an advancement to the Agency for Healthcare Research and Quality (AHRQ) Social Determinants of Health (SDoH) covariates as key gaps are addressed, a finer spatial resolution (Census Tract), and environmental covariates are included.

The process of curating and standardizing these datasets is non-trivial, as they are often measured at different spatial and temporal resolutions and have different spatial and temporal granularities. For example, the US Census data products typically use census blocks, block groups, or counties, whereas air pollutants from the EPA and weather data are available on 1km grids, and some economic data may be available only at a zip code level. In this context, standardized refers to the datasets all being at the same scale of spatial extent (e.g., US Census Tract and/or County), and curated refers both to a process that is repeatable, has data provenance, and which uses appropriate methodologies for converting covariates.

The data contained in the EDH dataset are drawn from multiple sources, and variables may have differing degrees of availability, patterns of missing data, and methodological considerations across sources, geographies, and years.

2. DOCUMENTATION OVERVIEW

This data source documentation report provides researchers with information on the structure and contents of the datasets, as well as descriptions of the data sources utilized to populate the data files. This document covers EDH's FY22-Q4 dataset curation documentation. The datasets included follow:

- • National Mental Health Services Survey

- National Survey on Drug Use and Health

- Occupational Employment and Wage Statistics

2.1 RECOMMENDED CITATION FOR FY22-Q4 DATASET PACKAGE

Klasky, H.B., Sparks, K., Logan, J., Hamaker, A., Whitehead, M., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY22-Q4, ORNL/SPR-2022/2587, PUB ID 183700. 2022.

2.2 PREVIOUS DOCUMENT RELEASES

Since its beginning, the following EDH project releases (datasets and data curation documentation) have been made available:

For the EDH Data Curation Documentation delivered on FY22-Q3, which contains the following datasets:

- Medicare Part D Opioid Prescribing Rates

- High Intensity Drug Trafficking Areas (HIDTA)

- Small-Area Estimates of Housing Characteristics

- Internet Access Services

- Facebook Social Connectedness Index

- Veteran Population for the Civilian Population

refer to: Klasky, H.B., Sparks, K., Logan, J., Tuccillo, J., Whitehead, M., Hamaker, A., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY22-Q3, ORNL/SPR-2022/2487 - Pub ID 178645. 2022.

For the EDH Data Curation Documentation delivered on FY22-Q2, which contains the following datasets:

- Eviction Rates (by county)

- Income Inequality (American Community Survey Income Inequality Measures Based on Income to Poverty Ratio by Census Block Group)

- Individual-Oriented Social Vulnerability Index (IOSVI), Census Block Groups, and

- National Instant Criminal Background Check System (NICS), Lethal Means Access.

refer to: Christian, J.B., Klasky, H.B., Sparks, K., Peluso, A., Tuccillo, J., Rastogi, D., Branstetter, M., Whitehead, M., Hamaker, A., and Watson, R., VA EDH Data Curation Documentation - FY22-Q2, Rev. 2, ORNL/SPR-2022/2391 - Pub ID 174092. 2022.

For the EDH Data Curation Documentation delivered on FY22-Q1, which contains the following datasets:

- Social Capital Index Dataset (2019 - updated)

- Social Vulnerability Index Dataset (2018)

- Block Group Area Deprivation Index Dataset for Washington, DC (2019)

- Low Food Access Area Dataset for Washington, DC (2017)

refer to: Christian, J.B., Klasky, H.B., Sparks, K., Peluso, A., Tuccillo, J., Devineni, P., and Watson, R. VA EDH Data Curation Documentation - FY22-Q1, Rev. 2, ORNL/SPR-2022/2316- Pub ID 172755. 2022.

Consult the following source for the FY21 dataset curation documentation: Christian, J.B., Branstetter, M, Klasky, H.B., Tuccillo, J., Sparks, K., Rastogi, D., Watson, R., Yoon, H-J., Kim, Y., VA EDH Data Curation Documentation - FY 2021, Rev. 2, ORNL/SPR-2021/2366 - Pub ID 170648. 2021.

3. CONTENTS AND STRUCTURE

The data in the EDH datasets were gathered from a variety of publicly available data sources. In the FY22-Q4 release the data sources included follow:

- National Mental Health Services Survey (2018).

- National Survey on Drug Use and Health (2018 and 2019).

- Occupational Employment and Wage Statistics (2021).

3.1 DATASET CURATION DOCUMENTATION STANDARD FORMAT

Each data source description follows a standard format with the next fields:

- Sponsor (name of the organization that provided the raw data, e.g., Health Resources and Services Administration [HRSA] for the Area Health Resources Files [AHRF])

- Description (brief, general description of the data)

- Inclusion in the EDH datasets

    – Lists the social or environmental determinants of health domains to which the data source has contributed variables

    – Includes additional information about the data source relevant to the EDH dataset

- Resources (links to original data source documentation, data download sites, and other relevant information)

- Update frequency: how often is each dataset going to be updated.

- Variable definitions and specifications (in tabular format)

    – Variable name

- Variable label

- Source table, if multiple data tables were available from the original data source (optional)

- Numerator (for derived variables, optional)

- Denominator (for derived variables) or original variable (when renamed for the EDH dataset, optional)

- Variable availability across years (in tabular format)

  - Variable name

  - Data year availability (e.g. 2009 to 2018)

3.2 DATASET CONVENTIONS

Regarding datasets' versioning, Microsoft SQL Server database system is used to supply the datasets. A table is used to hold each dataset. The following schema names for the quarterly releases are (or will be) included in the database: OMHSP_FY22Q1, OMHSP_FY22Q2, OMHSP_FY22Q3, OMHSP_FY22Q4, OMHSP_FY23Q1,…, and so on. These will aid in distinguishing between releases when we deliver the same data set from one release to the next.

Variables in the EDH dataset were created from these several data sources in one of two ways:

1. Drawn directly from the original data source. When the data was available from the data source as needed, we renamed the original variables for clarity and consistency across years, and to fit the naming conventions of the SDOH beta data files.

2. Derived using data from the original data source. For some data sources, it was necessary to calculate percentages or rates for inclusion in the beta data files. The numerators and denominators for the variables and their sources are shown following each data source description.

The following conventions were followed in constructing the EDH datasets to provide researchers with a consistent and easy-to-use resource:

- Variable assignment to annual datasets. Variables appear in the annual datasets that correspond with (1) the single year represented by the original data source (e.g., Nursing Home Compare data for facilities in 2016 appears in the 2016 county dataset), or (2) the last year in a period represented by the data (e.g., American Community Survey data aggregated over 2012 to 2016 is in the 2016 dataset).

- Variable availability. The availability of each variable changes across data years. Following each data source description in this report is a table showing the availability of each variable in the annual datasets. When a variable is not available, we indicate it with NA (not available) or simply '-'.

- Variable naming. Except for the geographic ID variables, all variable names begin with a data source acronym followed by an underscore and a descriptive title.

- • Missing values. The datasets use a blank to denote a missing value, almost exclusively. The one exception is the provider ratio variables from the County Health Rankings (CHR) data, which have negative values for counties where the number of providers is zero. This is described further in the description of the CHR data.

Detailed information about each data source is included in the following sections of this report.

4. NATIONAL MENTAL HEALTH SERVICES SURVEY

4.1 SPONSOR

Substance Abuse & Mental Health Data Archive

4.2 DESCRIPTION

The National Mental Health Services Study (N-MHSS) is a yearly survey that gathers information related to the features and services of all recognized mental health treatment institutions in all 50 states, the District of Columbia, and all U.S. territories and authorities. The survey also gathers information related to the population serviced in these facilities as of a certain survey reference date every alternate year (as of 2014).

The N-MHSS provides numerous specific variables at the facility level (for example, "Facility type [Psychiatric hospital, Outpatient mental health facility, etc.]", "Does facility offer mental health intake", and "Does facility offer substance abuse treatment"). This dataset includes specific features that were agreed upon during our conversations with our VHA's sponsors for the state-level survey in 2018. Look at the URL links in the Resources section below for more information about the columns in this table.

4.3 INCLUSION

Year: 2018

Geographical unit: state level

Data: This data set was compiled using facility-level National Mental Health Services Survey (N-MHSS) 2018 data, which was then aggregated to state-level statistics. The survey contains various question categories, which are represented here as A category questions and B category questions, and may be found at the original survey link in the Resources section below. In addition, we are sharing the data survey's un-edited csv file in the table 'national_mental_health_services_survey'.

*Regarding the B-Questions-related datasets, it should be noted that the following B-questions in the survey lacked data for several of the category numeric ranges. The following variables are not displayed in the respective data table in those cases. The variables missing can be thought of as zero data.

- [OMHSP_FY22Q4].[b8_est_mental_health_admin_vet] is missing columns: [101_250], [251_500], [501_1000], [1001_1500], and [more_1500].

- [OMHSP_FY22Q4].[b4c_num_beds_designated_mental_health_treat] is missing columns: [501_1000] and [more_1500].

- [OMHSP_FY22Q4].[b4a_num_clients_receiving_24hr_res_mental_health] is missing column: [1001_1500].

4.4 RESOURCES

To learn more about the N-MHSS data:

Link to landing page: https://www.datafiles.samhsa.gov/dataset/national-mental-health-services-survey-2018-n-mhss-2018-ds0001

Direct link to zip file: https://www.datafiles.samhsa.gov/sites/default/files/field-uploads-protected/studies/N-MHSS-2018/N-MHSS-2018-datasets/N-MHSS-2018-DS0001/N-MHSS-2018-DS0001-bundles-with-study-info/N-MHSS-2018-DS0001-bndl-data-tsv.zip

Link to questionnaire: https://www.datafiles.samhsa.gov/sites/default/files/field-uploads-protected/studies/N-MHSS-2018/N-MHSS-2018-datasets/N-MHSS-2018-DS0001/N-MHSS-2018-DS0001-info/N-MHSS-2018-DS0001-info-questionnaire-specs.pdf

Link to pdf codebook: https://www.datafiles.samhsa.gov/sites/default/files/field-uploads-protected/studies/N-MHSS-2018/N-MHSS-2018-datasets/N-MHSS-2018-DS0001/N-MHSS-2018-DS0001-info/N-MHSS-2018-DS0001-info-codebook.pdf

4.5   UPDATE FREQUENCY

Every fiscal year, or as requested by the sponsor, this dataset will be updated and distributed. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

Table 1. National Mental Health Services Survey (N-MHSS)

Variable name	Description
A-Questions	The A-Questions-related dataset provided in this release is as follows.
a_questions_state_crosstabs_final_fips	The dataset (table name) that includes the following columns
FIPS	Two-digit fips code for the state represented.
State	US state
hosp_inpatient_24hour	Survey question a3, mental health treatment service settings at this facility.
residential_24hour	Survey question a3, mental health treatment service settings at this facility.
partial_hosp_day_treatment	Survey question a3, mental health treatment service settings at this facility.
outpatient	Survey question a3, mental health treatment service settings at this facility.
mental_health	Survey question a7, What is the primary treatment focus of this facility, at this location?
mental_and_substance_abuse	Survey question a7, What is the primary treatment focus of this facility, at this location?
general_health	Survey question a7, What is the primary treatment focus of this facility, at this location?
public_own	Survey question a9, Is this facility operated by.
private_own	Survey question a9, Is this facility operated by.
individual_psychotherapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.

couple_family_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
group_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
cognitive_behavioral_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
dialectical_behavior_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
behavior_modification	Survey question a11, mental health treatment approaches offered at this facility, at this location.
integrated_dual_disorders_treatment	Survey question a11, mental health treatment approaches offered at this facility, at this location.
trauma_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
activity_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
electroconvulsive_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
telehealth_therapy	Survey question a11, mental health treatment approaches offered at this facility, at this location.
psychotropic_medication	Survey question a11, mental health treatment approaches offered at this facility, at this location.
housing_services	Survey question a12.11, services and practices offered regarding housing services.

supported_housing_programs	Survey question a12.12, services and practices offered regarding supported housing.
vocational_rehabilitation_services	Survey question a12.14, services and practices offered regarding vocational rehabilitation services.
supported_employment_services	Survey question a12.15, services and practices offered regarding supported employment.
psychiatric_emergency_walkin_services	Survey question a12.18, services and practices offered regarding psychiatric emergency walk-in services.
suicide_prevention_services	Survey question a12.19, services and practices offered regarding suicide prevention services.
dedicated_mental_health_treatment_for_trauma	Survey question a14.9, mental health treatment program or group exclusively for trauma patients.
dedicated_mental_health_treatment_for_veterans	Survey question a14.11, mental health treatment program or group exclusively for veterans.
crisis_intervention_team	Survey question a15, facility offer a crisis intervention team that handles acute mental health issues.
_	—
B-Questions	**The B-Questions-related datasets provided in this release are as follows.**
b3a_num_patients_24hour_inpatient_mental_health	Survey question b3a, dataset regarding patients receiving 24-hour hospital inpatient mental health treatment.
b3c_num_inpatient_beds_for_mental_health	Survey question b3c, dataset regarding beds at this facility designated for providing mental health treatment.
b4a_num_clients_receiving_24hr_res_mental_health_treat	Survey question b4a, dataset regarding clients receiving 24-hour residential mental health treatment.

b4c_num_beds_designated_mental_health_treat	Survey question b4c, dataset regarding residential beds designated for providing mental health treatment.
b5a_num_clients_less24hr_mental_health_treat	Survey question b5a, dataset regarding clients receiving less-than-24-hour mental health treatment
b7_num_mental_health_treat_admin_12month	Survey question b7, dataset regarding treatment admissions in previous 12-month period.
b8_est_mental_health_admin_vet	Survey question b8, dataset regarding the veterans' estimated percent of admissions.
—	—
B-Questions datasets columns	**The following are the columns in each of the B-Questions datasets provided in this release*.**
fips	US Census Block Group Federal Information Processing Standards (FIPS) code.
state	US State
1_10	No. of cases with frequency between 1 to 10
11_20	No. of cases with frequency between 11 to 20
21_30	No. of cases with frequency between 21 to 30
31_40	No. of cases with frequency between 31 to 40
41_50	No. of cases with frequency between 41 to 50
51_75	No. of cases with frequency between 51 to 75
76_100	No. of cases with frequency between 76 to 100
101_250	No. of cases with frequency between 101 to 250

251_500	No. of cases with frequency between 251 to 500
501_1000	No. of cases with frequency between 501 to 1000
1001_1500	No. of cases with frequency between 1001 to 1500
more_1500	No. of cases with frequency greater than 1500

Table 2. Variable availability across years, (N-MHSS)

Variable name	2018
a_questions_state_crosstabs_final_fips dataset	X
b3a_num_patients_24hour_inpatient_mental_health dataset	X
b3c_num_inpatient_beds_for_mental_health dataset	X
b4a_num_clients_receiving_24hr_res_mental_health_treat dataset	X
b4c_num_beds_designated_mental_health_treat dataset	X
b5a_num_clients_less24hr_mental_health_treat dataset	X
b7_num_mental_health_treat_admin_12month dataset	X
b8_est_mental_health_admin_vet dataset	X

5. NATIONAL SURVEY ON DRUG USE AND HEALTH

5.1 SPONSOR

Substance Abuse and Mental Health Services Administration

5.2 DESCRIPTION

The National Survey on Drug Use and Health (NSDUH) quantifies a) misuse of prescription medications, alcohol, cigarettes, and illicit and prescription substances; b) substance abuse disorder and therapy for drug abuse major depressive episode and care for depression; c) extreme emotional anguish, mental sickness, and mental health services. The data provides estimates of drug use and mental illness at the state level.

The amount of drug abuse and mental illness across various subgroups is also identified by the NSDUH data, along with changes over time and the requirement for treatment services. This dataset includes a subset of data from the 2018-2019 state-level NSDUH that was agreed upon by our VHA sponsors.

5.3 INCLUSION

Year: annual averages based on 2018 and 2019 NSDUHs.

Geographical unit: state level.

Data:

This NSDUH dataset (see table 'national_survey_on_drug_use_and_health') includes 51 selected state-level variables from 2018-2019, as specified in the variable description table below. The figures are estimates (in thousands). In addition, for each variable, we provide lower and upper 95% confidence intervals, which are values that exist in the source.

5.4 RESOURCES

To learn more about the NSDUH data:

Landing page link: https://www.samhsa.gov/data/report/2018-2019-nsduh-estimated-totals-state

Direct link to zip file:
https://www.samhsa.gov/data/sites/default/files/reports/rpt32879/NSDUHsaeTotal2019/2019NSDUHsaeCountTabsCSVs.zip

5.5 UPDATE FREQUENCY

Every fiscal year, or as requested by the sponsor, this dataset will be updated and distributed. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

Table 3. National Survey on Drug Use and Health (NSDUH)

Variable name	Description
FIPS	US Census Block Group Federal Information Processing Standards (FIPS) code.
marijuana_use	Table 3, marijuana use in the past month, by age group and state.
marijuana_use_95_ci_lower	Lower 95% confidence interval.
marijuana_use_95_ci_upper	Upper 95% confidence interval.
non_treatment_substance_use	Table 26, needing but not receiving treatment at a specialty facility for substance use in the past year.
non_treatment_substance_use_95_ci_lower	Lower 95% confidence interval.
non_treatment_substance_use_95_ci_upper	Upper 95% confidence interval.
drug_use	Table 6, illicit drug use other than marijuana in the past month, by age group and state.
drug_use_95_ci_lower	Lower 95% confidence interval.
drug_use_95_ci_upper	Upper 95% confidence interval.
any_mental_illness	Table 27, any mental illness in the past year, by age group and state.
any_mental_illness_95_ci_lower	Lower 95% confidence interval.
any_mental_illness_95_ci_upper	Upper 95% confidence interval.
alcohol_use	Table 13, alcohol use in the past month, by age group and state.
alcohol_use_95_ci_lower	Lower 95% confidence interval.
alcohol_use_95_ci_upper	Upper 95% confidence interval.
serious_mental_illness	Table 28, serious mental illness in the past year, by age group and state.
serious_mental_illness_95_ci_lower	Lower 95% confidence interval.
serious_mental_illness_95_ci_upper	Upper 95% confidence interval.

alcohol_binge	Table 14, binge alcohol use in the past month, by age group and state.
alcohol_binge_95_ci_lower	Lower 95% confidence interval.
alcohol_binge_95_ci_upper	Upper 95% confidence interval.
received_mental_health_services	Table 29, received mental health services in the past year, by age group and state.
received_mental_health_services_95_ci_lower	Lower 95% confidence interval.
received_mental_health_services_95_ci_upper	Upper 95% confidence interval.
drug_disorder	Table 20, illicit drug use disorder in the past year, by age group and state.
drug_disorder_95_ci_lower	Lower 95% confidence interval.
drug_disorder_95_ci_upper	Upper 95% confidence interval.
major_depressive_episode	Table 30, major depressive episode in the past year, by age group and state.
major_depressive_episode_95_ci_lower	Lower 95% confidence interval.
major_depressive_episode_95_ci_upper	Upper 95% confidence interval.
pain_reliever_disorder	Table 21, pain reliever use disorder in the past year, by age group and state.
pain_reliever_disorder_95_ci_lower	Lower 95% confidence interval.
pain_reliever_disorder_95_ci_upper	Upper 95% confidence interval.
suicidal_thoughts	Table 31, individuals that had serious thoughts of suicide in the past year, by age group and state.
suicidal_thoughts_95_ci_lower	Lower 95% confidence interval.
suicidal_thoughts_95_ci_upper	Upper 95% confidence interval.
alcohol_disorder	Table 22, alcohol use disorder in the past year, by age group and state.
alcohol_disorder_95_ci_lower	Lower 95% confidence interval.
alcohol_disorder_95_ci_upper	Upper 95% confidence interval.

suicide_planning	Table 32, individuals that made any suicide plans in the past year, by age group and state.
suicide_planning_95_ci_lower	Lower 95% confidence interval.
suicide_planning_95_ci_upper	Upper 95% confidence interval.
substance_use_disorder	Table 23, substance use disorder in the past year, by age group and state.
substance_use_disorder_95_ci_lower	Lower 95% confidence interval.
substance_use_disorder_95_ci_upper	Upper 95% confidence interval.
attempted_suicide	Table 33, attempted suicide in the past year, by age group and state.
attempted_suicide_95_ci_lower	Lower 95% confidence interval.
attempted_suicide_95_ci_upper	Upper 95% confidence interval.
population	This figure represents the state's population in 2019, who were 18 years of age or older.

Table 4. Variable availability across years, (NSDUH)

Variable name	2019
FIPS	X
marijuana_use	X
marijuana_use_95_ci_lower	X
marijuana_use_95_ci_upper	X
non_treatment_substance_use	X
non_treatment_substance_use_95_ci_lower	X
non_treatment_substance_use_95_ci_upper	X
drug_use	X
drug_use_95_ci_lower	X
drug_use_95_ci_upper	X
any_mental_illness	X
any_mental_illness_95_ci_lower	X
any_mental_illness_95_ci_upper	X
alcohol_use	X
alcohol_use_95_ci_lower	X
alcohol_use_95_ci_upper	X
serious_mental_illness	X
serious_mental_illness_95_ci_lower	X
serious_mental_illness_95_ci_upper	X
alcohol_binge	X
alcohol_binge_95_ci_lower	X
alcohol_binge_95_ci_upper	X
received_mental_health_services	X

received_mental_health_services_95_ci_lower	X
received_mental_health_services_95_ci_upper	X
drug_disorder	X
drug_disorder_95_ci_lower	X
drug_disorder_95_ci_upper	X
major_depressive_episode	X
major_depressive_episode_95_ci_lower	X
major_depressive_episode_95_ci_upper	X
pain_reliever_disorder	X
pain_reliever_disorder_95_ci_lower	X
pain_reliever_disorder_95_ci_upper	X
suicidal_thoughts	X
suicidal_thoughts_95_ci_lower	X
suicidal_thoughts_95_ci_upper	X
alcohol_disorder	X
alcohol_disorder_95_ci_lower	X
alcohol_disorder_95_ci_upper	X
suicide_planning	X
suicide_planning_95_ci_lower	X
suicide_planning_95_ci_upper	X
substance_use_disorder	X
substance_use_disorder_95_ci_lower	X
substance_use_disorder_95_ci_upper	X
attempted_suicide	X
attempted_suicide_95_ci_lower	X

attempted_suicide_95_ci_upper	X
population	X

6. OCCUPATIONAL EMPLOYMENT AND WAGE STATISTICS

6.1 SPONSOR

United States Census Bureau and US Bureau of Labor Statistics

6.2 DESCRIPTION

This dataset contains estimates from the Occupational Employment and Wage Statistics (OEWS) for each US state between April 1 and July 1, 2020. Each OEWS column represents the "Estimated total employment rounded to the nearest 10 (excludes self-employed)", as described in the original data file from the Bureau of Labor Statistics.

6.3 INCLUSION

Year: OEWS estimates for April 1, 2020 to July 1 2021 are included in this dataset.

Geographical unit: state level.

Data: With the VHA sponsor, we identified five main occupational groups:

- Life, physical and social science occupations, we only included clinical and counseling psychologist.

- Community and social service occupations, we have included marriage and family therapists, SUD, behavioral and mental counselors, and mental health social workers.

- Healthcare practitioners and technical occupations: we included psychiatrists (we did not include psych techs or aids, as they don't provide direct clinical care under a license)

- Healthcare support occupations.

- Other psychiatric support occupations and psychiatric aids.

6.4 RESOURCES

To learn more about the OEWS data:

Data on occupation obtained from: https://www.bls.gov/oes/special.requests/oesm21st.zip

Data on population obtained from:
https://data.census.gov/cedsci/table?g=0100000US%240400000&y=2021&tid=PEPPOP2021.NST_EST2021_POP

Working topics are available here: https://www.bls.gov/oes/current/oes_al.htm

6.5 UPDATE FREQUENCY

Every fiscal year, or as requested by the sponsor, this dataset will be updated and distributed. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

Table 5. Occupational Employment and Wage Statistics (OEWS)

Variable name	Description
FIPS	US Census Block Group Federal Information Processing Standards (FIPS) code. Two-digit FIPS code for the state represented.
19-3033	Clinical and Counseling Psychologists.
29-1223	Psychiatrists.
21-1013	Marriage and Family Therapists.
21-1018	Substance Abuse, Behavioral Disorder, and Mental Health Counselors.
21-1023	Mental Health and Substance Abuse Social Workers.
population	July 2021 Population Estimate.

Table 6. Variable availability across years, (OEWS)

Variable name	2021
FIPS	X
19-3033	X
29-1223	X
21-1013	X
21-1018	X
21-1023	X
population	X