

Directory of Public Datasets for Youth Mental Health: Enhancing Research through Data, Accessibility, and Artificial Intelligence

Hua Min, Xia Jing, Cui Tao, Joel E. Williams, Sarah F Griffinb, Christianne Esposito-Smythers, Bruce Chorpita

Submitted to: JMIR Mental Health
on: March 13, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	4
Supplementary Files.....	30
Figures	31
Figure 1.....	32
Figure 2.....	33



Directory of Public Datasets for Youth Mental Health: Enhancing Research through Data, Accessibility, and Artificial Intelligence

Hua Min¹ PhD; Xia Jing² PhD, MD; Cui Tao³ PhD; Joel E. Williams² PhD, MPH; Sarah F Griffin² PhD, MPH; Christianne Esposito-Smythers⁴ PhD; Bruce Chorpita⁵ PhD

¹Department of Health Administration and Policy College of Public Health George Mason University Fairfax US

²College of Behavioral, Social and Health Sciences Clemson University Clemson US

³Department of Artificial Intelligence and Informatics Mayo Clinic Jacksonville US

⁴4400 University Dr. Department of Psychology George Mason University Fairfax US

⁵University of California, Los Angeles Los Angeles US

Corresponding Author:

Hua Min PhD

Department of Health Administration and Policy

College of Public Health

George Mason University

4400 University Dr.

Fairfax

US

Abstract

This paper introduces a curated directory of publicly available datasets focused on youth mental health (under 18 years old). Youth mental health issues have been recognized as a pressing crisis in recent years in the USA. The directory is designed to serve as a critical infrastructure to enhance research, inform policy-making, and support the application of artificial intelligence (AI) and machine learning (ML) in youth mental health research. Unlike a systematic review, this paper offers a brief review of open data recourses, addressing the challenges posed by the growing volume of fragmented health data sources in youth mental health. By centralizing these datasets, the directory enables researchers to easily discover and access relevant datasets, fostering secondary data analysis and linked data analysis. This directory has the potential to reveal deeper insights into factors like school health policies, youth mental health outcomes, and cross-country comparisons. By improving data integration and accessibility, the directory encourages interdisciplinary collaboration and the development of evidence-based interventions for youth mental health issues. The paper also discusses the potential for automated data curation approaches, enhancing efficiency and supporting AI and ML in youth mental health research, ultimately contributing to more effective, data-driven policies, preventions, and interventions for youth mental health.

(JMIR Preprints 13/03/2025:73852)

DOI: <https://doi.org/10.2196/preprints.73852>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

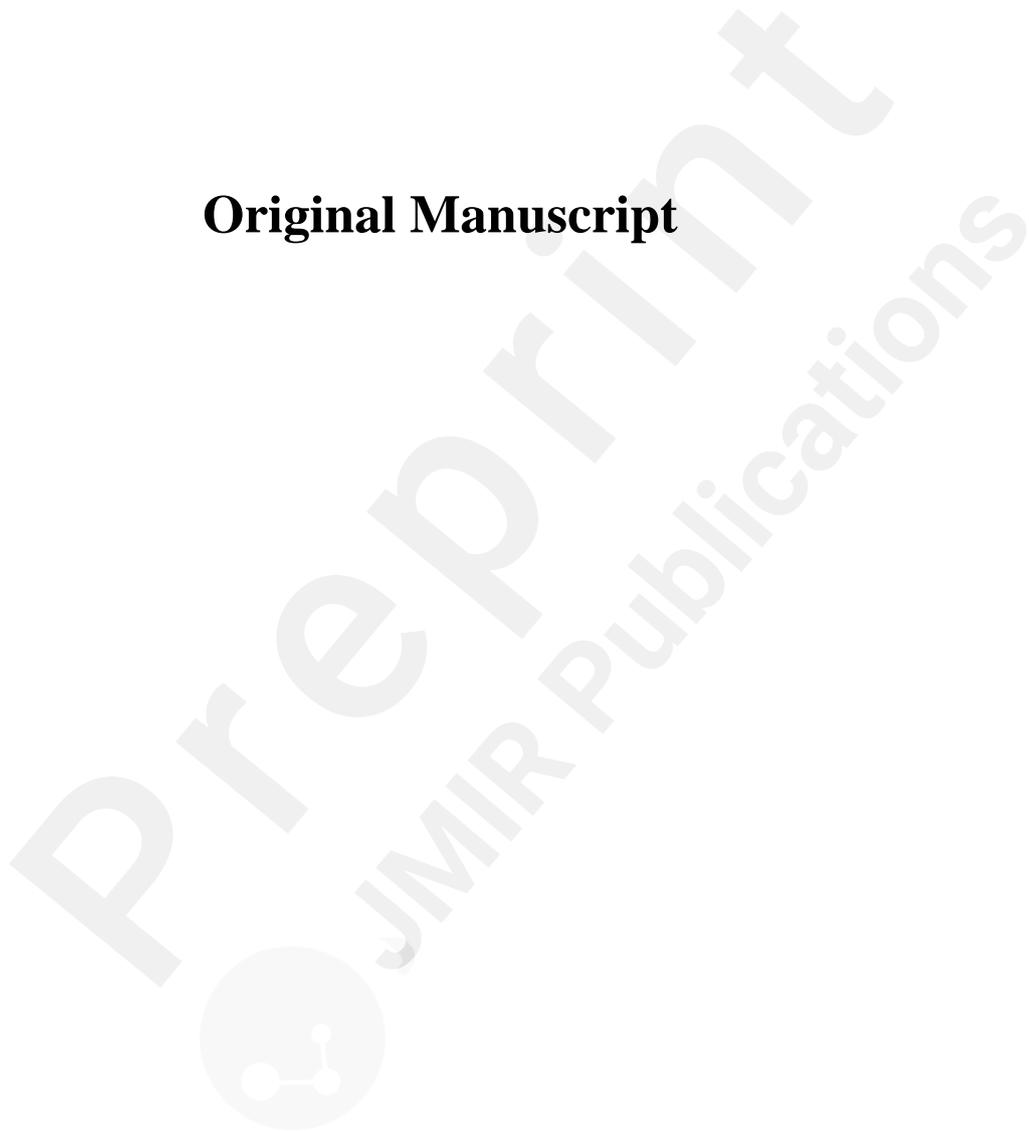
Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in a JMIR journal, I will be able to make the full manuscript available to all users.

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript



Directory of Public Datasets for Youth Mental Health: Enhancing Research through Data, Accessibility, and Artificial Intelligence

Hua Min^{a*}, Xia Jing^b, Cui Tao^c, Joel E. Williams^b, Sarah F Griffin^b, Christianne Esposito-Smythers^d, Bruce Chorpita^e

^aDepartment of Health Administration and Policy, College of Public Health, George Mason University, Fairfax, Virginia 22030, USA; ^bCollege of Behavioral, Social and Health Sciences, Clemson University, Clemson, South Carolina 29634, USA; ^cDepartment of Artificial Intelligence and Informatics, Mayo Clinic, Jacksonville, Florida, USA; ^dDepartment of Psychology, George Mason University, Fairfax, Virginia 22030, USA; ^eDepartment of Psychology, University of California Los Angeles, Los Angeles, CA 90095

*Corresponding author

Hua Min, Ph.D.

Department of Health Administration and Policy

College of Public Health

George Mason University

4400 University Dr.

Fairfax, VA 22030

Telephone: 703-993-5648

Fax: 703-993-1953

Email: hmin3@gmu.edu

Directory of Public Datasets for Youth Mental Health: Enhancing Research through Data, Accessibility, and Artificial Intelligence

This paper introduces a curated directory of publicly available datasets focused on youth mental health (under 18 years old). Youth mental health issues have been recognized as a pressing crisis in recent years in the USA. The directory is designed to serve as a critical infrastructure to enhance research, inform policy-making, and support the application of artificial intelligence (AI) and machine learning (ML) in youth mental health research. Unlike a systematic review, this paper offers a brief review of open data recourses, addressing the challenges posed by the growing volume of fragmented health data sources in youth mental health. By centralizing these datasets, the directory enables researchers to easily discover and access relevant datasets, fostering secondary data analysis and linked data analysis. This directory has the potential to reveal deeper insights into factors like school health policies, youth mental health outcomes, and cross-country comparisons. By improving data integration and accessibility, the directory encourages interdisciplinary collaboration and the development of evidence-based interventions for youth mental health issues. The paper also discusses the potential for automated data curation approaches, enhancing efficiency and supporting AI and ML in youth mental health research, ultimately contributing to more effective, data-driven policies, preventions, and interventions for youth mental health.

Keywords: youth mental health; publicly available datasets; data accessibility; data integration; artificial intelligence, evidence-based research

Introduction

Many mental health conditions develop during youth and adolescence, making it a global public concern[1–3]. The number of young people experiencing mental health issues continues to rise. According to the Centers for Disease Control and Prevention (CDC), the percentage of high school students reporting feelings of sadness and hopelessness increased significantly over the past decade, rising from 28% in 2011 to 40% in 2023[4]. This marked increase highlights a growing mental health crisis among adolescents, emphasizing the urgent need for prevention and early intervention.

Adolescence is a period of significant psychological, emotional, and social development, making it a critical window for addressing mental health concerns[4]. When left unaddressed, these issues can lead to long-term worse consequences, including impaired educational attainment, strained relationships, and adverse health outcomes in adulthood, far beyond adolescence. In 2021, the U.S. Office of the Surgeon General released an advisory titled "Protecting Youth Mental Health," emphasizing the urgent need to address the mental health challenges faced by young people[5].

Young individuals with substance use disorders are at an elevated risk of developing co-occurring mental health challenges, including suicidal behaviors, which can exacerbate lifelong difficulties, contribute to social issues, and result in poorer treatment outcomes[6,7]. Preventable mental health issues like adolescent suicide[8,9] and substance use[10,11] significantly contribute to youth disease, disability, and mortality, highlighting the importance of proactive measures in urgent need to save lives and improve the well-being of young people. Suicide, one of the leading causes of death among young people, is often preventable through the timely identification of warning signs, access to supportive networks, and evidence-based interventions[9]. Similarly, substance use disorders, which frequently emerge during adolescence, can be reduced through early preventive strategies that focus on education, community support, and behavioral health services[11]. Investing in preventive mental health care during adolescence yields far-reaching benefits[12], not only for affected individuals at the time, but also for them later in life, their families, communities, and societies. Effective prevention programs can enhance resilience, reduce stigma, and promote mental well-being, ultimately fostering healthier and more productive generations overall.

A significant gap persists between the mental health needs of children and adolescents in the United States and the availability of services to meet those needs[13]. Most young people with mental health conditions do not seek or receive treatment[14]. Barriers include limited awareness of available services, inaccessibility, insurance coverage issues, insufficient coordinated care, a shortage of specialized providers, unstable living conditions, concerns about confidentiality, and fear of

stigmatization[14,15]. Schools have been identified as critical settings for screening, preventing, and treating youth mental health issues, as students spend the majority of their daily time in these environments[4,16]. To address this critical issue, it is essential to strengthen youth-focused mental health services, integrate mental health education into school curricula, and equip educators, parents, and peers with tools to recognize and respond to early signs of distress[4]. By prioritizing prevention and early treatment, we can disrupt the cycle of untreated mental health issues and cultivate supportive environments where young people can thrive.

In alignment with these efforts, evidence-based medicine continues to drive advancements in mental health care[17]. Evidence-based approaches are crucial for addressing youth mental health challenges[18]. To ensure high-quality, impactful research and promote its translation into clinical practice, service delivery, and policy, the National Institute of Mental Health (NIMH) developed its Strategic Plan for Research aiming at achieving these goals [19]. This plan addresses emerging challenges and opportunities by leveraging scientific advances to improve mental health outcomes. Among its four goals, Goal 2—"Examine Mental Illness Trajectories Across the Lifespan"—closely aligns with the collection and analysis of mental health data using diverse datasets.

Mental health conditions are multifaceted[20], influenced by a combination of biological factors (e.g., brain trauma, genetics), psychological factors (e.g., stress, traumatic events), and environmental, social, and economic determinants. Understanding these complexities demands the collection and integration of diverse datasets to capture the full spectrum of influences on mental health, especially for youth. By assembling data that captures biological, psychological, social, and environmental factors, researchers can reveal a more comprehensive picture of mental health conditions, discover patterns, identify at-risk populations, and design targeted interventions that address the unique needs of young people. Notable existing data sources include national surveys like the Youth Risk Behavior Surveillance System (YRBSS)[21], the National Survey on Drug Use and Health (NSDUH)[22], and the National Survey of Children's Health (NSCH)[23], large-scale

longitudinal studies such as the Adolescent Brain and Cognitive Development (ABCD) Study[24], and targeted clinical trials[25] exploring the efficacy of early interventions and prevention strategies.

The growing volume of biomedical data generated by these studies provides a solid foundation for data mining and knowledge discovery to extract meaningful insights across diverse data types[26–29]. The new emerging AI techniques[27], such as ML, natural language processing (NLP), and especially generative AI[28], are increasingly employed to analyze large-scale datasets, identify early indicators of mental health issues, and develop predictive models for conditions like depression, anxiety, suicide, and substance use disorders. For example, AI is being used to process real-time data from wearable devices[30], social media activities[31], and electronic health records (EHR) [27] to gain insights into behavioral patterns and risk factors among adolescents.

One significant challenge in data science and AI is the lack of accessible healthcare data. EHR data is often restricted due to privacy and security regulations, such as HIPAA. This limitation highlights the critical role publicly available datasets play in enabling big data analysis and advancing research. These datasets are particularly vital in advancing youth mental health research, where access to comprehensive data is crucial for revealing meaningful insights and discovering new interventions.

With the NIH data sharing policy, an increasing amount of research data is now available for secondary data reuse, enabling a broader range of studies to build upon existing findings. Datasets such as the YRBSS, NSDUH, NSCH, and ABCD study provide researchers with open access to comprehensive data on behavioral, social, and biological factors that influence mental health. Effectively integrating these datasets will be crucial in deepening our understanding of mental health and developing innovative solutions for prevention and treatment. By leveraging these resources, researchers can collaborate across institutions, validate findings, and accelerate the discovery of new insights. Furthermore, publicly available data facilitates the development of AI-powered tools to identify trends, predict risks, and personalize mental health interventions. This approach can lead to

scalable, equitable solutions to address the growing mental health needs of young people worldwide.

This paper aims to review existing publicly available datasets related to youth mental health and provide a detailed directory as a reference tool for researchers interested in integrating diverse datasets for advanced data analysis or AI-assisted discovery across different aspects of youth mental health. By highlighting the scope and potential of these datasets, this paper seeks to guide future research efforts in leveraging diverse data sources to better understand and address the complex challenges facing youth mental health.

Method

In this brief review, we employed a multi-faceted approach to identify publicly available datasets relevant to youth mental health. Figure 1 presents the overall workflow.

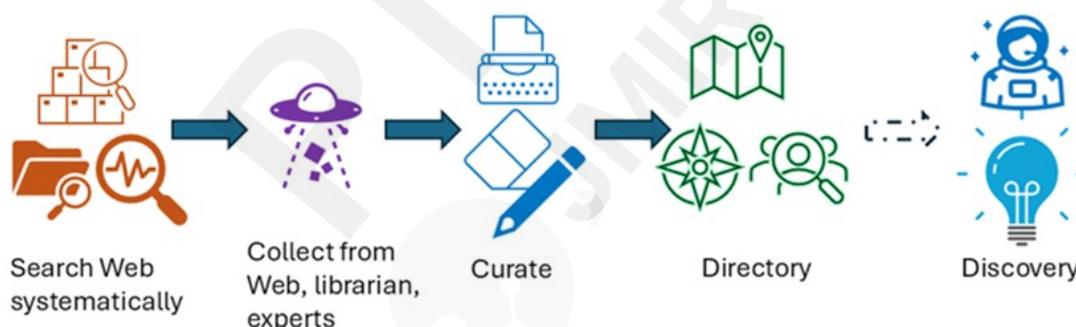


Figure 1. Overall workflow of curating the youth mental health data directory (dashed arrow indicates anticipated usage of the directory).

The structured search included three main approaches:

- (1) Targeted Search: We conducted focused searches of reputable websites of health organizations, institutions, and government agencies (e.g., National Institutes of Health

(NIH), CDC) using keywords such as "youth or adolescence mental health datasets," "publicly available data," and "mental health surveys." Our search focused primarily on U.S.-based resources.

- (2) **Librarian Consultation:** We collaborated with professional librarians specializing in health and social sciences to identify and verify datasets not readily found in publicly available search engines.
- (3) **Expert Knowledge:** Leveraging our team's expertise and prior familiarity with youth mental health research, we included datasets we had previously worked with or identified as significant in the field.

Next, we curated and categorized the identified datasets, providing a detailed and consistent description of each, including the following elements:

- **Dataset Name:** The title of the dataset and the organizations responsible for its creation.
- **Description:** The primary research goals and questions the dataset is designed to address.
- **Components:** Key variables and types of data collected, such as biological, psychological, social, or economic data.
- **Downloadable Site:** Direct URL link to download the dataset, with an emphasis on datasets that are freely available and do not require restrictive licensing or complex application processes.
- **Data Format:** Information about the format of the data (e.g., CSV, SAS, SPSS, etc.).
- **Data Categories:** The type of study or data collection method, such as national surveys, administrative records, or longitudinal cohort studies.

Results

A list of publicly available datasets addressing youth mental health and school policies was compiled in the Table 1. While this is not a comprehensive list of all available datasets, it includes key

resources that are relevant to youth mental health. The CDC initiates, conducts, and supports many national surveys to understand youth mental health, including the YRBSS, School Health Profiles (SHP), School Health Policies and Practices Study (SHPPS), National Health Interview Survey (NHIS), and the National Youth Tobacco Survey (NYTS). Another important agency is the Substance Abuse and Mental Health Services Administration (SAMHSA)[32] which collects data related to substance use and mental health services, including National Survey on Drug Use and Health (NSDUH), National Substance Use and Mental Health Services Survey (N-SUMHSS), Mental Health Client-Level Data (MH-CLD), and Treatment Episode Data Set (TEDS). The U.S. Census Bureau also conducts an annual household survey, the National Survey of Children's Health (NSCH), which provides national and state-level data on the health and healthcare needs of children aged 0–17, as well as their families and communities. The National Institute of Mental Health (NIMH) Data Archive is a valuable resource for mental health data, including studies focusing on youth mental health. One such study is the ABCD Study, which tracks over 11,000 children starting at ages 9-10 to examine how factors like substance use and genetics influence brain development. Another is the Treatment for Adolescents with Depression Study (TADS), along with its follow-up, Substance Use and Other Outcomes Following Treatment for Adolescent Depression (SOFTAD), evaluates treatments for adolescent depression and their long-term effects on substance use and other outcomes. Additionally, the Agency for Healthcare Research and Quality (AHRQ) conducts the annual Medical Expenditure Panel Survey (MEPS) to collect data on healthcare use, costs, insurance coverage, and health status from individuals, families, medical providers, and employers across the U.S. The National Data Archive on Child Abuse and Neglect (NDACAN) curates a wide range of datasets related to child maltreatment, including administrative records, survey results, and state and local-level reports. These datasets provide valuable insights into the prevalence, characteristics, and outcomes of child abuse and neglect, as well as the effectiveness of related systems and services.

Like the ABCD Study, there are two other local neuroimaging studies: the NKI Rockland

Sample (NKI-RS)[33] and the Healthy Brain Network (HBN)[34]. The NKI-RS focuses on understanding brain function and structure across a wide range of conditions, while HBN aims to improve the diagnosis and treatment of mental health and learning disorders by identifying biological markers, such as brainwave signals and imaging data. Both studies provide valuable neuroimaging data to advance research in mental health and cognitive development. Image data from both studies are stored on Amazon Web Services (AWS) S3, providing accessible and scalable storage for large datasets.

Table 1. Overview of Publicly Available Datasets Addressing Youth Mental Health and School Policies.

Name	Administered By	Description	Components	Downloadable Site	File Format	Category
Youth Risk Behavior Surveillance System (YRBSS)	CDC's Division of Adolescent and School Health (DASH)	Since 1991, the YRBS has been administered during the spring of odd-numbered years to students in grades 9–12 enrolled in U.S. public and private schools.	Focuses on youth health behaviors, experiences, and conditions, including topics such as sexual activity, injury and violence, bullying, diet, physical activity, obesity, mental health indicators, suicide-related behaviors, and substance use.	https://www.cdc.gov/yrbs/data/index.html	Access and ASCII	National survey
School Health Profiles (SHP)	CDC's Division of Adolescent and School Health (DASH)	Since 1996, the survey has been conducted every even-numbered year to assess school health policies and practices in states, large urban school districts, and U.S. territories.	Topics include Health Education, Physical Education and Physical Activity, Tobacco Use Prevention, Nutrition Services, School Health Policies, Family and Community Involvement, School-Based Health Services, Professional Development.	Request data at https://www.cdc.gov/school-health-profiles/contact/index.html#cdc_generic_section_2-request-data	ASCII, SAS, or Excel	National survey
School Health Policies and Practices Study (SHPPS)	CDC	Since 1994, the survey has been conducted every six years at the state, district, school, and classroom levels for grades K–12.	It has 8 components of school health at the state and district levels: health education, physical education and activity, health services, mental health and social services, nutrition services, healthy and safe school environment, faculty and staff health promotion, and family and community involvement	https://archive.cdc.gov/www_cdc.gov/healthyouth/data/shpps/data.htm	ASCII, Access, SAS, SPSS	National survey
National Health Interview Survey (NHIS)	CDC through the National Center for Health Statistics (NCHS)	Since 1957, the survey has been conducted annually on the civilian noninstitutionalized population, selecting one adult (Sample Adult) and one child (Sample Child) randomly from each family.	Annual core includes chronic condition, functioning and disability, insurance, health care access and use, health-related behaviors, and demographic. Rotating core services utilization, mental health assessment, chronic pain, preventive service, industry and occupation, injuries.	https://www.cdc.gov/nchs/nhis/documentation/?CDC_AAref_Val=https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm	ASCII, CSV, SAS, SPSS, STATA	National survey

National Youth Tobacco Survey (NYTS)	CDC	Since 1999, the NYTS has been conducted annually as a voluntary, school-based, self-administered survey targeting U.S. middle and high school students.	The NYTS measures items related to tobacco use, like demographics, youth access to tobacco, and exposure to secondhand smoke.	https://www.cdc.gov/tobacco/about-data/surveys/historical-nyts-data-and-documentation.html	SAS, Access, and Excel	National survey
National Survey of Children's Health (NSCH)	Health Resources and Services Administration's (HRSA) Maternal and Child Health Bureau (MCHB) and U.S. Census Bureau	Conducted annually as a household survey, it is the largest national and state-level survey on the health and healthcare needs of children aged 0–17, their families, and their communities.	Topics include Family health and activities, Health conditions and functional difficulties, Health insurance status, type, and adequacy, Health care service use and access, Impact of child's health on family, Medical home, Parental health, Parental perceptions of neighborhood characteristics, Physical and mental health, Preventive and specialty care, School readiness, and Transition to adult health care	https://www.census.gov/programs-surveys/nsch/data/datasets.html	SAS, STATA	National survey
National Survey on Drug Use and Health (NSDUH)	Substance Abuse and Mental Health Services Administration (SAMHSA)	Conducted annually and provides representative of persons aged 12 and older in the civilian noninstitutionalized population of the United States	Topics include the use of tobacco, alcohol, and drugs; substance use disorders; mental health issues; and receipt of substance use and mental health treatment	https://www.samhsa.gov/data/data-we-collect/nsduh/datafiles	SAS, SPSS, Stata, Delimited, R	National survey
National Substance Use and Mental Health Services Survey (N-SUMHSS)	Substance Abuse and Mental Health Services Administration (SAMHSA)	Annual survey of all active substance use and mental health facilities across the United States	Provides number, location, and characteristics of public and private-owned mental health and substance use facilities in the United States	https://www.samhsa.gov/data/data-we-collect/n-sumhss/datafiles	SAS, SPSS, Stata, Delimited, R	National survey
Mental Health Client-Level Data (MH-CLD)	Substance Abuse and Mental Health Services Administration (SAMHSA)	Since 2011, MH-CLD captures patients treated in facilities that are either operated by or receive block grant funding through a state mental health agency.	Collects demographic and mental health characteristics data on approximately 5 to 6 million patients receiving care in outpatient, hospital inpatient, and residential services each year	https://www.samhsa.gov/data/data-we-collect/mh-cld/datafiles	SAS, SPSS, Stata, Delimited, R	Administrative Records Data
Treatment Episode Data Set (TEDS)	Substance Abuse and Mental Health Services Administration (SAMHSA)	TEDS is a national data system of annual admissions to and discharges from substance use treatment facilities that are licensed or certified by Single State Agencies (SSAs) to provide substance use treatment services.	Contains demographic, substance use, mental health, clinical, legal, and socioeconomic characteristics of all admissions and discharges aged 12 and over who are receiving publicly funded substance use treatment services.	https://www.samhsa.gov/data/data-we-collect/teds/datafiles	SAS, SPSS, Stata, Delimited, R	Administrative Records Data
Adolescent Brain	NIMH Data	It is a prospective cohort study that enrolled 11,876	Collects data on brain development, physical health,	https://nda.nih.gov/		Ongoing

Cognitive Development Study (ABCD)	Archive	children aged 9 to 11 years, recruited from 21 study sites in the baseline year (2016-2018). This ongoing study will follow participants until they are approximately 19–20 years old.	behavioral patterns, and mental health, including neuroimaging, substance use, cognitive assessments, and psychosocial factors. It also gathers genetic and epigenetic data to explore how environmental and genetic influences affect adolescent development over time.	general-query.html?q=query=featured-datasets:Adolescent%20Brain%20Cognitive%20Development%20Study%20(ABCD)		longitudinal cohort study
Treatment for Adolescents with Depression Study (TADS) / Substance Use and Other Outcomes Following Treatment for Adolescent Depression (SOFTAD)	NIMH Data Archive	TADS included 439 participants ages 12 to 17 from various geographic regions in the US who were diagnosed with major depression. The SOFTAD study recruited 196 of these 439 adolescents and followed them for an additional 3.5 years to explore whether successful treatment of depression reduces the risk of developing Substance Use Disorders (SUDs) and other outcomes.	Examines the short- and long-term effectiveness of the antidepressant medication fluoxetine (Prozac), Cognitive Behavioral Therapy (CBT) alone, and their combination for treating depression in adolescents.	Request data at https://www.nimhgenetics.org/download-tool/DP	SAS, SPSS, CSV	Multi-site clinical trial
Medical Expenditure Panel Survey (MEPS)	AHRQ	Started in 1996, it is an annual survey of families, individuals, medical providers (doctors, hospitals, pharmacies, etc.), and employers across the U.S. It includes two main components: Household Component and Insurance Component.	Collects data on the specific health services that Americans use, how frequently they use them, the cost of these services, and how they are paid for, as well as data on the cost, scope, and breadth of health insurance held by and available to U.S. workers.	https://meps.ahrq.gov/data_stats/download_data_files.jsp	SAS, Stata, R, ASCII	National survey
National Data Archive on Child Abuse and Neglect (NDACAN)	U.S. Department of Health and Human Services	Since 1988, NDACAN promotes scholarly exchange among researchers in the child maltreatment field. The archive provides access to national, state, and local-level data.	Its primary aim is to provide access to national, state, and local-level data, primarily aimed at tracking the volume and nature of child maltreatment reporting each year in the US. It supports research to inform policies and practices that improve child welfare and protect children.	Request data at https://www.ndacan.acf.hhs.gov/datasets/datasets-list.cfm	SAS, SPSS, and STATA	Various types of data
NKI Rockland Sample (NKI-RS)	NIH and the NY State Office of Mental Health	Approximately 1,500 individuals aged 6 to 85 years were recruited from the Rockland County community in New York.	The NKI-Rockland Sample provides a rich neuroimaging and phenotypic resource to characterize lifespan normative brain-behavior relationships. Data include psychiatric diagnostic, medical, behavioral, and cognitive phenotyping, as well as multimodal brain imaging (resting fMRI, diffusion MRI, morphometric MRI, arterial spin labeling), genetics, and actigraphy.	Neuroimaging data via AWS S3 buckets at s3://fcp-indi/data/Projects/RocklandSample/RawDataBIDS/Latest . Request Phenotypical data at https://fcon_1000.projects.nitrc.org/indi/enhanced/phenotypicdat	BIDS and CSV	A longitudinal study

Healthy Brain Network (HBN)	Child Mind Institute	HBN creates a Biobank from a community sample of 10,000 children and adolescents (ages 5–21) residing in the New York City area	The HBN collects a wide array of data, including neuroimaging (MRI), genetics, behavioral assessments, and clinical data, all aimed at improving the understanding of brain health and mental illness across different age groups.	a.html Request phenotypic data at http://data.healthbrainnetwork.org/ . All EEG, eye tracking and imaging data can be accessed at http://fcon_1000.projects.nitrc.org/indi/ctmi/healthy_brain_network	BIDS and CSV	Cross-sectional study with biobanking
-----------------------------	----------------------	---	--	--	--------------	---------------------------------------

The Inter-university Consortium for Political and Social Research (ICPSR)[35], based at the University of Michigan, is one of the world's largest archives of social science data. It offers extensive datasets for research and education across social, behavioral, and health sciences. These datasets include both publicly available and restricted-access data from single studies, U.S. national surveys, and international studies and surveys. A search for "adolescent or youth mental health" on ICPSR returns 4,453 studies. Table 2 shows ten randomly selected youth mental health datasets from the ICPSR. For example, Monitoring the Future: A Continuing Study of American Youth, funded by National Institute on Drug Abuse, is an ongoing study of the behaviors, attitudes, and values of Americans from adolescence through adulthood. To date, 133 data files are hosted at ICPSR.

Table 2. Ten randomly selected Youth Mental Health Datasets from ICPSR.

Name	Description	Components	Downloadable Site	Data Format	Category
National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2018	Adolescents in grades 7 through 12 were enrolled during the 1994-1995 school year. They were followed into young adulthood through four in-home interviews, with the most recent conducted in 2008, when participants were aged 24-32.	Collected data on social, economic, psychological, and physical well-being, alongside contextual information about family, neighborhood, community, school, friendships, peer groups, and romantic relationships.	https://www.icpsr.umich.edu/web/ICPSR/studies/21600	SAS, SPSS, STATA, R, ASCII, Delimited	Longitudinal cohort study
Family cumulative risk and mental health in Chinese adolescents	Examine the developmental cascades between family cumulative risk, life satisfaction, and anxiety and depression among Chinese adolescents.	Family cumulative risk and mental health	https://www.openicpsr.org/openicpsr/project/194861/version/V1/view	SPSS	Study

Resilience and Mental Health among Juveniles	The first study involved 201 juveniles, the second involved 253 juveniles.	Study 1 is to analyze the relationship between resilience and the broadly-understood mental health of juveniles admitted to youth education centers. Study 2 aimed to understand the direct and indirect role of resilience in shaping the mental health of juveniles.	https://www.openicpsr.org/project/120001/version/V1/view	SPSS	Study
Adolescent Depression	This dataset focuses on self-compassion, self-efficacy, and trait resilience as mediators between insecure attachment (specifically attachment anxiety and attachment avoidance) and depression.	Anxiety and depression	https://www.openicpsr.org/project/209446/version/V1/view	SPSS	Study
Health Behavior in School-Aged Children (HBSC), 2009-2010	Since 1982, the WHO Regional Office for Europe has sponsored the HBSC study, surveying the health behaviors of young people every four years in over 40 countries. The data available here are from the results of the United States survey conducted during the 2009-2010 school year.	The HBSC survey includes questions about various substances (tobacco, alcohol, marijuana, etc.), family composition, physical health, and health behaviors such as eating habits, dieting, physical activity, body image, health issues, and bullying. School administrators also complete a survey on school programs, policies, and health course content.	https://www.icpsr.umich.edu/web/ICPSR/studies/34792	SAS, SPSS, STATA, R, ASCII, Delimited	Multiple-year survey
National Youth Survey (NYS) Series	NYS is a longitudinal study of delinquency among American youth during the period 1976 through 1980. For this series, parents and youth were interviewed about events and behavior of the preceding year to gain a better understanding of both conventional and deviant types of behavior by youths.	Data include demographics, socioeconomic status, family disruptions, neighborhood issues, parental expectations, labeling, family and peer influences, attitudes toward deviance, parental discipline, community engagement, substance use, victimization, pregnancy, depression, outpatient services, domestic violence, and sexual activity.	https://www.icpsr.umich.edu/web/ICPSR/series/88	SAS, SPSS, STATA, ASCII, Delimited	Longitudinal survey
National Survey of Children: Wave I, 1976, Wave II, 1981, and Wave III, 1987 (ICPSR 8670)	The survey aims to assess the physical, social, and psychological well-being of American children, create a national profile of their living conditions, analyze the relationship between these conditions and child development, and examine the effects of marital disruption on children and family	Information is provided on the child's well-being, family, experiences with family disruption, behavior, physical health, and mental health.	https://www.icpsr.umich.edu/web/ICPSR/studies/8670	ASCII	National survey
Child Abuse, Neglect, and Violent Criminal Behavior in a Midwestern Metropolitan Area of the United	The study examines the link between childhood abuse or neglect and later criminal behavior, focusing on whether early victimization leads to criminal offending in adolescence or adulthood, and its association with juvenile and adult arrests, particularly for violent offenses.	Part 1: Demographics (age, race, sex, birthdate). Part 2: Abuse/neglect details (type, duration, removal, placement, survival). Part 3: Family info (household, disruptions, reporters) and perpetrator data (relation, age, race, sex). Part 4: Adult arrest charges (occasion,	https://www.icpsr.umich.edu/web/ICPSR/studies/38156	SAS, SPSS, STATA, ASCII	A prospective cohort study

States, 1967-1988 (ICPSR 9480)		counts, year, location, offense). Part 5: Juvenile arrest charges (year, number of arrests, offense). Parts 1-3 focus on individuals under age 11, Part 4 on adult charges, and Part 5 on juvenile charges.			
Monitoring the Future: A Continuing Study of American Youth (12th-Grade Survey), 2020 (ICPSR 38156)	This survey of 12th-grade students examines changes in values, behaviors, and lifestyle orientations of American youth. Students complete one of six randomly assigned questionnaires, each with unique topical questions but all including core questions on demographics and drug use.	The survey covers drugs like tobacco, alcohol, marijuana, prescription medications, LSD, cocaine, ecstasy, heroin, and more. It also includes topics such as attitudes toward religion, women's roles, educational aspirations, self-esteem, drug education, and exposure to violence and crime.	https://www.icpsr.umich.edu/web/ICPSR/studies/38156	SAS, SPSS, STATA, R, ASCII, Delimited	Multiple-year survey
Youth, Education, and Society Supplement: School Health Policies and Practices Survey (YES), 2006-2014 (ICPSR 36350)	YES surveyed secondary schools in the Monitoring the Future study and a larger supplementary sample, conducting annual surveys of school administrators from 2006-2007 to 2013-2014 school year.	Topics include school characteristics, school nutrition and physical education policies, school lunch programs, and school vending machines, stores and snack bars.	https://www.icpsr.umich.edu/web/ICPSR/studies/36350	SAS, SPSS, STATA, ASCII	Multiple-year survey

Other central repositories for open data from the U.S. government include Data.gov, which are presented in Table 3. The Department of Health and Human Services (HHS) publishes datasets as part of its Open Data program. The primary resource for public access to health-related data is Healthdata.gov, with additional specialized sites available for specific topics, including Data.CDC.gov, OpenFDA, and Data.CMS.gov. Figure 2 summarizes the results.

Table 3. Important data sources from reputable organizations that include youth mental health data.

Name	URL	Note	No. of Datasets
ICPSR	https://www.icpsr.umich.edu/web/pages/ICPSR/index.html	Provides access to a wide range of social science datasets	more than 350,000
Data.gov	https://www.data.gov	A central repository for all U.S. government data	306,135
Healthdata.gov	https://www.healthdata.gov	The U.S. Department of Health and Human Services' platform for accessing health-related datasets	3,239
Data.CDC.gov	https://data.cdc.gov	The CDC's open data platform	1,077
OpenFDA	https://open.fda.gov	Provides access to FDA-related datasets	152
Data.CMS.gov	https://data.cms.gov	Offers datasets from the Centers for Medicare & Medicaid Services	160



Figure 2. The summary view of the youth mental health data directory.

In addition to the raw data collected, researchers often identify, code, and summarize publicly available research reports and publications within a particular domain (e.g., children and adolescent psychotherapy), categorizing them as researcher assertions rather than data assertions[36]. Such literature data, similar to systematic literature reviews, represent another critical form of publicly accessible information. Despite their descriptive nature, they can be utilized for meta-analyses, understanding the current state of the field, identifying potential research gaps, and conducting other aggregate evidence-based analyses. Common and reliable sources for identifying data or studies include PubMed, Cochrane Library, PsycINFO, Embase, ClinicalTrials.gov, and NIH RePORT for research portfolios.

Discussions

Summary of results

Our findings highlight several major national survey series, conducted by organizations such as the CDC, SAMHSA, and the U.S. Census Bureau, that focus on youth and adolescent mental health and substance use. For instance, the YRBSS monitors behaviors that may negatively impact the health of high school students. The NSCH provides extensive data on various aspects of children's lives, including physical and mental health, healthcare access and quality, and their family, neighborhood, school, and social environments. Similarly, the NSDUH offers estimates of substance use and mental illness at the national, state, and local levels, helping to identify prevalence across different groups,

track trends over time, and assess treatment needs.

We also collect data on state and school health policies, which provide data with different scopes and granularities. Since youth spend the majority of their daily time at school, these policies play a crucial role in their mental health. The CDC's Division of Adolescent and School Health (DASH) has developed an action guide to promote mental health and well-being in schools[37]. It also collects data on youth and school health policies and practices through initiatives such as the School Health Profiles (SHP). Similarly, the CDC has conducted the School Health Policies and Practices Study (SHPPS) to provide additional insights. These two datasets, SHP and SHPPS, are primarily focused on school policy and practice, rather than on individual student data or a sample of student data. However, they are important for researchers who want to study the impact of policy on student mental health, which is why we have also included them in our directory.

The WHO conducts a similar survey, the Global School-based Student Health Survey (GSHS), to help countries measure and assess behavioral risk and protective factors among young people aged 13 to 17, related to the leading causes of morbidity and mortality among children and adults worldwide.

The ABCD Study, NKI-RS, and HBN are three influential neuroimaging studies that are contributing valuable insights into brain development and mental health. A key aspect of their success is the use of advanced storage solutions like AWS S3, which enables them to handle large and complex datasets efficiently. By using AWS S3, these studies improve data management and provide easier access to large-scale neuroimaging data, advancing mental health research and supporting personalized, data-driven treatment approaches.

ICPSR, Data.gov, Healthdata.gov, Data.CDC.gov, OpenFDA, and Data.CMS.gov are valuable health data repositories that host a wide range of research data, including surveys, longitudinal studies, and individual research projects. These platforms provide extensive datasets that are crucial for public health research, policy analysis, and decision-making. They offer access to data

on various health topics, such as disease prevalence, healthcare utilization, medication usage, and health behaviors, allowing researchers, policymakers, and practitioners to make informed decisions and conduct comprehensive analyses to improve public health outcomes.

Importance of publicly available data for youth mental health research

Adolescence is a pivotal stage of development where mental health challenges, such as anxiety, depression, and trauma-related disorders, often emerge. However, these challenges remain under-researched due to limited access to comprehensive, high-quality datasets. Furthermore, considering youths are at the early stage of one's life, effective management and treatment of mental health conditions will have profound impacts on the rest of their lives, which distinguishes the crucial period and critical nature of studying youth mental health.

Access to publicly available data enables researchers to identify trends and patterns in mental health conditions, risk factors, and service utilization among diverse populations of youth. Such data also allow for the evaluation of the impact of school health policies on mental health outcomes, shedding light on how these policies can support students' well-being and promote mental health education. By uncovering key insights into the prevalence and effects of mental health challenges, these data guide future research and inform intervention strategies. Furthermore, they provide a foundation for developing evidence-based interventions tailored to the unique needs and characteristics of different youth groups, promoting more effective and targeted outcomes.

Publicly available data also helps to address health disparities by providing a means to explore differences in mental health outcomes based on factors such as socioeconomic status, race, and geographic location. This, in turn, can inform more equitable solutions and ensure that mental health services reach those who need them most. Furthermore, open access to data fosters cross-disciplinary collaboration, bringing together experts from various fields such as psychology, public health, and education to tackle complex mental health issues from multiple angles.

Moreover, publicly available datasets empower community organizations, schools, families,

and mental health providers by providing actionable insights that help shape programs, allocate resources more effectively, and create supportive environments for youth.

Significance of curating a list of publicly available data and resources for youth mental health research

With the growing number of studies, datasets, and reports being published across various platforms—each designed to address specific aspects of health and often collecting data independently without integration—the volume of fragmented data is increasing. This highlights the urgent need to create a centralized list of health data repositories to streamline access, reduce redundancy, and enhance their overall utility, thereby accelerating advancements in youth mental health research.

Compiling a list of available data also encourages data reuse, allowing researchers to explore secondary research questions and make the most of existing datasets. While there are similar efforts, such as ICPSR, its scope is broader and not specifically focused on youth mental health. In contrast, our directory focuses exclusively on youth mental health, helping users save time by eliminating the need to search for datasets individually. Moreover, ICPSR does not provide a comprehensive list of youth mental health datasets, omitting key resources such as the NYTS and MH-CLD.

Additionally, a centralized health data repository helps policymakers by providing easy access to relevant data, empowering them to make informed, data-driven decisions to improve youth mental health outcomes. Furthermore, such a resource supports interdisciplinary collaboration by enabling researchers from diverse fields—such as public health, psychology, pediatrics, education, and social work—to access and utilize datasets relevant to their work. This collaborative potential can accelerate innovation, inform policy, and drive the development of targeted interventions to address complex health challenges such as youth mental health effectively.

Benefits of linked datasets

The list of publicly available data and resources not only promotes secondary data analysis but also provides a foundation for more advanced linked data analysis. By integrating datasets from different

sources, such as school health policies and youth mental health outcomes, researchers can gain deeper insights into the relationships between various factors influencing youth well-being. For example, Foti et al.[38] investigated how state and local agencies use the YRBSS and SHP in various ways to monitor and address issues related to adolescent and school health. Gould et al.[39] used two national surveys to assess the impact of the BP Deepwater Horizon oil rig explosion on mental health, substance use disorders, chronic health conditions, and behavioral health service use among local residents. The first survey was the NSDUH Gulf Coast Oversample (GCO), part of the 2011 NSDUH, and the second was the CDC's Gulf States Population Survey (GSPS). Linked data allowed researchers to conduct more comprehensive analyses that can inform more effective interventions and policies, both locally and globally.

However, several challenges exist in linking youth mental health datasets. First, the number of well-known and widely available datasets in this field is limited. This paper aims to introduce existing datasets by compiling them from multiple sources into a centralized directory. Second, the lack of standardized frameworks hinders the integration of youth mental health data. Data collected and archived by various organizations and studies result in inconsistencies in terminology, variations in data structures, and differences in how key concepts—such as mental health outcomes, demographic information, and treatment interventions—are represented across datasets. One widely recognized solution to this challenge is the use of ontologies, which have been successfully applied in the medical field. Ontologies provide a structured framework to harmonize these core concepts, facilitating more efficient data sharing, comparison, and integration[33]. While some ontologies exist in the behavioral and mental health domain, such as the Behaviour Change Intervention Ontology (BCIO)[40], Addiction Ontology[41], Mental Disease Ontology[42], Mental Health Management Ontology[43], and Social Determinants of Health Ontology (SDoHO)[44], they remain insufficient and lack detailed coverage specific to youth mental health. The absence of comprehensive standardization creates significant barriers to effective data integration and analysis, further

underscoring the urgent need for ontologies in the behavioral sciences[45].

Limitations

While the availability of publicly accessible health data is invaluable for advancing research on youth mental health, there are several limitations that must be addressed. First, our list is not comprehensive, but it serves as a starting point for researchers to discover and access relevant health data repositories. Second, there are inconsistencies across datasets. Datasets across different platforms may use varying definitions, classifications, and measurement tools for mental health conditions, making it difficult to compare or combine data from multiple sources. This lack of standardization can introduce biases and complicate analyses. Third, many publicly available datasets are cross-sectional, capturing a snapshot of youth mental health at a particular point in time. However, understanding mental health requires a longitudinal approach, tracking changes over time. The lack of long-term data restricts the ability to assess causal relationships or predict future mental health outcomes. Lastly, some datasets may not adequately represent certain geographic regions or demographic groups, particularly marginalized communities. This lack of representation can hinder the understanding of mental health disparities and limit the applicability of research findings to diverse populations.

Future work

Despite these limitations, several avenues for future work can improve the utility and impact of publicly available data for youth mental health research. This paper represents an initial step in creating a dedicated directory for publicly available youth mental health data recourses. In the future, we aim to develop a Youth Mental Health Ontology (YMHO) based on the datasets in this directory, harmonizing the data with established standards (e.g., Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)[46]) and YMHO itself. By compiling the high-quality datasets, based on defined criteria, to advance the understanding of youth mental health.

There is a need for standardized definitions, measures, and data collection methods across datasets to facilitate data sharing and integration. Such standardization would allow researchers to more effectively compare and combine data from different sources, leading to more robust findings.

Continued advances in ML and AI are expected to drive breakthroughs across a wide range of research fields, including mental health, in the coming years. However, these advancements depend on large volumes of well-curated, FAIR (Findable, Accessible, Interoperable, and Reusable) data. Effective and automatic curation processes are crucial for data integration and harmonization, encompassing the characterization, annotation, management, and preservation of digital datasets. These processes will further improve the efficiency and accessibility of these valuable resources and also meet the needs of a large number of curated data for ML and AI to apply.

Building collaborative platforms that integrate multiple data sources and encourage interdisciplinary collaboration will promote innovative solutions to youth mental health issues. These platforms could also facilitate real-time data sharing, allowing for quicker responses to emerging mental health trends and crises.

Conclusions

The availability of publicly accessible health data is a cornerstone for improving youth mental health outcomes. Compiling an extensive list of such data and resources not only streamlines access but also enhances the impact of research, interventions, and policies for youth mental health. This initiative is vital for fostering collaboration, addressing disparities, and advancing the collective goal of healthier, more resilient adolescents.

References:

- [1] Wang R, Zhu B, Yu X, et al. Childhood violence exposure and anxiety and depression of children and adolescents. *J Affect Disord* [Internet]. 2025 [cited 2024 Dec 27];369:608–614.
- [2] Depa N, Desai S, Patel S, et al. Mental health disparities amongst sexual-minority adolescents of the US – A national survey study of YRBSS-CDC. *Psychiatry Res* [Internet]. 2022 [cited 2024 Dec 27];314:114635.
- [3] Blakemore S-J. Adolescence and mental health. *The Lancet* [Internet]. 2019 [cited 2024 Dec 27];393(10185):2030–2031.
- [4] Demissie Z, Brener N. Demographic Differences in District-Level Policies Related to School Mental Health and Social Services—United States, 2012. *J Sch Health* [Internet]. 2017 [cited 2024 Dec 27];87(4):227–235.
- [5] Protecting Youth Mental Health [Internet]. Office of the Surgeon General, U.S. Department of Health and Human Services; 2021 [cited 2025 Jan 1]. Available from: <https://www.hhs.gov/surgeongeneral/priorities/youth-mental-health/index.html>.
- [6] Erskine HE, Moffitt TE, Copeland WE, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. *Psychol Med* [Internet]. 2015 [cited 2025 Jan 2];45(7):1551–1563.
- [7] Richert T, Anderberg M, Dahlberg M. Mental health problems among young people in substance abuse treatment in Sweden. *Subst Abuse Treat Prev Policy* [Internet]. 2020 [cited 2025 Jan 2];15(1):43.
- [8] Moreno MA. Preventing Adolescent Suicide. *JAMA Pediatr* [Internet]. 2016 [cited 2024 Dec 27];170(10):1032.
- [9] Hink AB, Killings X, Bhatt A, et al. Adolescent Suicide—Understanding Unique Risks and Opportunities for Trauma Centers to Recognize, Intervene, and Prevent a Leading Cause of Death. *Curr Trauma Rep* [Internet]. 2022 [cited 2024 Dec 29];8(2):41–53.
- [10] Skiba D, Monroe J, Wodarski JS. Adolescent Substance Use: Reviewing the Effectiveness of Prevention Strategies. *Soc Work* [Internet]. 2004 [cited 2024 Dec 27];49(3):343–353.
- [11] Stanis JJ, Andersen SL. Reducing substance use during adolescence: a translational framework for prevention. *Psychopharmacology (Berl)* [Internet]. 2014 [cited 2024 Dec 29];231(8):1437–1453.
- [12] Das JK, Salam RA, Lassi ZS, et al. Interventions for Adolescent Mental Health: An Overview of Systematic Reviews. *J Adolesc Health Off Publ Soc Adolesc Med*. 2016;59(4S):S49–S60.
- [13] Bringewatt EH, Gershoff ET. Falling through the cracks: Gaps and barriers in the mental health system for America's disadvantaged children. *Child Youth Serv Rev*. 2010;32(10):1291–1299.
- [14] Brener ND, Weist M, Adelman H, et al. Mental Health and Social Services: Results From the School Health Policies and Programs Study 2006. *J Sch Health* [Internet]. 2007 [cited 2024 Dec

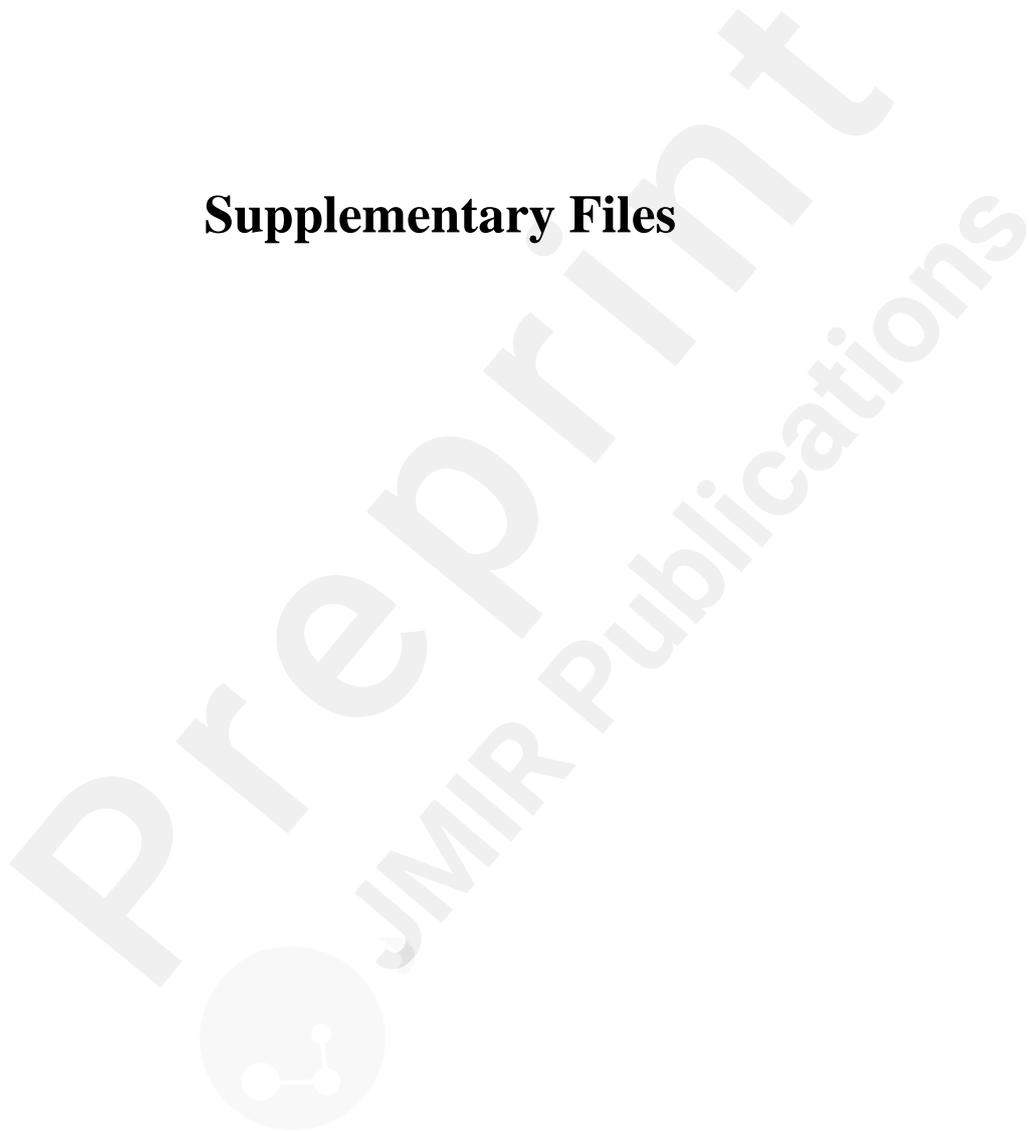
29];77(8):486–499.

- [15] Weist MD. Expanded School Mental Health Services. In: Ollendick TH, Prinz RJ, editors. *Adv Clin Child Psychol* [Internet]. Boston, MA: Springer US; 1997 [cited 2024 Dec 29]. p. 319–352. Available from: http://link.springer.com/10.1007/978-1-4757-9035-1_9.
- [16] Committee on School Health. School-Based Mental Health Services. *Pediatrics* [Internet]. 2004 [cited 2024 Dec 29];113(6):1839–1845.
- [17] Gratzer D. Improving access to evidence-based mental health care. *Can Med Assoc J* [Internet]. 2020 [cited 2024 Dec 29];192(13):E342–E343.
- [18] Graaf G, Accomazzo S, Matthews K, et al. Evidence Based Practice in Systems of Care for Children with Complex Mental Health Needs. *J Evid-Based Soc Work* [Internet]. 2021 [cited 2024 Dec 29];18(4):394–412.
- [19] The National Institute of Mental Health Strategic Plan [Internet]. The National Institute of Mental Health; [cited 2025 Jan 1]. Available from: <https://www.nimh.nih.gov/about/strategic-planning-reports>.
- [20] Remes O, Mendes JF, Templeton P. Biological, Psychological, and Social Determinants of Depression: A Review of Recent Literature. *Brain Sci* [Internet]. 2021 [cited 2024 Dec 29];11(12):1633.
- [21] Brener ND, Mpfu JJ, Krause KH, et al. Overview and Methods for the Youth Risk Behavior Surveillance System — United States, 2023. *MMWR Suppl* [Internet]. 2024 [cited 2024 Dec 29];73(4):1–12.
- [22] Behavioral Health Barometer: Region 1, Volume 7: Indicators as Measured in the 2021-2022 National Surveys on Drug Use and Health [Internet]. Rockville (MD): Substance Abuse and Mental Health Services Administration (US); 2024 [cited 2024 Dec 29]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK608627/>.
- [23] National Survey of Children’s Health Data Briefs [Internet]. Rockville (MD): Health Resources and Services Administration; 2018 [cited 2024 Dec 29]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK603024/>.
- [24] Alcohol Research: Current Reviews Editorial Staff. NIH’s Adolescent Brain Cognitive Development (ABCD) Study. *Alcohol Res Curr Rev*. 2018;39(1):97.
- [25] clinicaltrials.gov [Internet]. [cited 2025 Jan 1]. Available from: <https://clinicaltrials.gov/search?cond=Youth%20Mental%20Health>.
- [26] Tovar D, Cornejo E, Xanthopoulos P, et al. Data Mining in Psychiatric Research. In: Kobeissy FH, editor. *Psychiatr Disord* [Internet]. Totowa, NJ: Humana Press; 2012 [cited 2024 Dec 29]. p. 593–603. Available from: https://link.springer.com/10.1007/978-1-61779-458-2_37.
- [27] Graham S, Depp C, Lee EE, et al. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Curr Psychiatry Rep* [Internet]. 2019 [cited 2024 Dec 29];21(11):116.

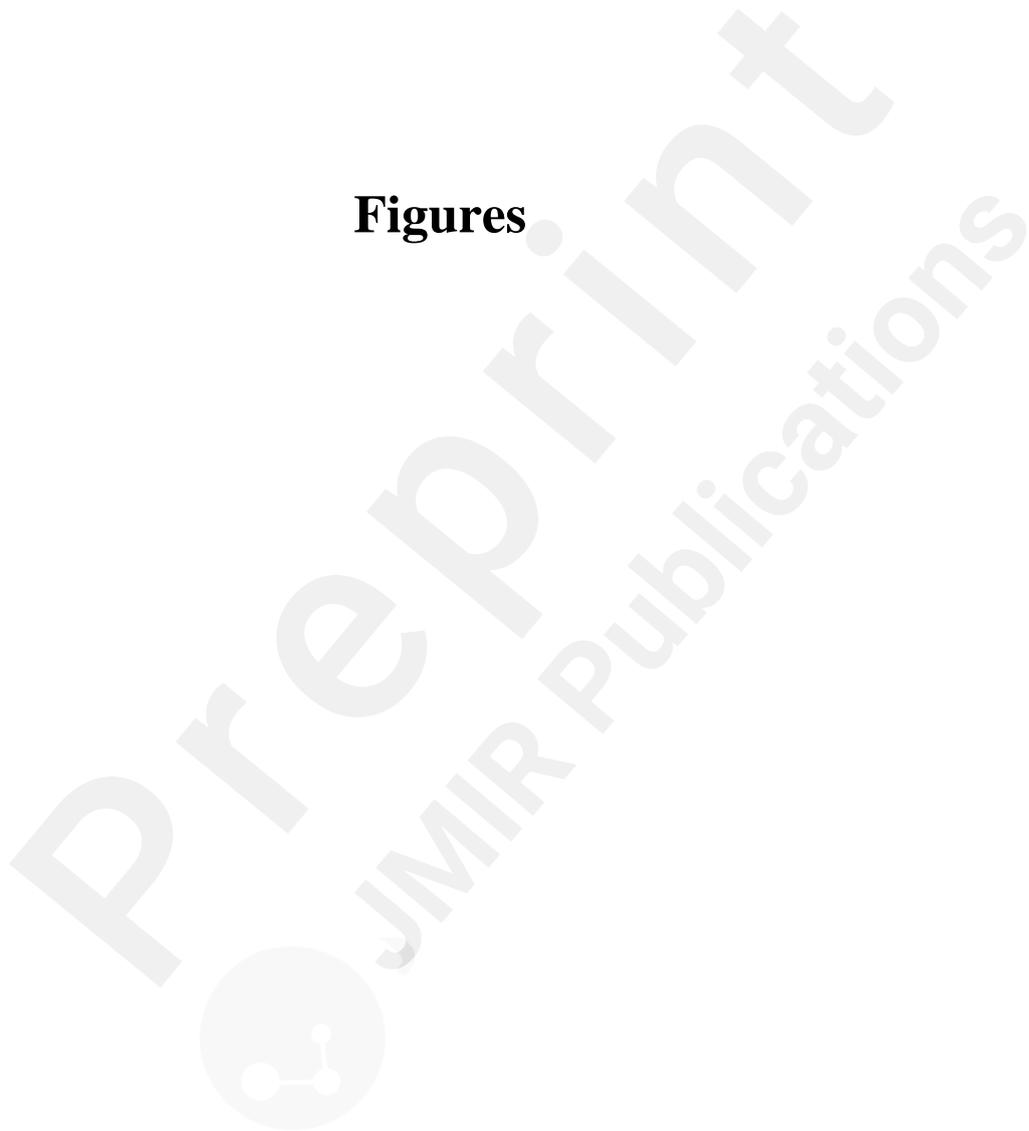
- [28] King DR, Nanda G, Stoddard J, et al. An Introduction to Generative Artificial Intelligence in Mental Health Care: Considerations and Guidance. *Curr Psychiatry Rep* [Internet]. 2023 [cited 2024 Dec 29];25(12):839–846.
- [29] Lee EE, Torous J, De Choudhury M, et al. Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging* [Internet]. 2021 [cited 2024 Dec 29];6(9):856–864.
- [30] Hickey BA, Chalmers T, Newton P, et al. Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review. *Sensors* [Internet]. 2021 [cited 2024 Dec 29];21(10):3461.
- [31] Ahmed A, Aziz S, Toro CT, et al. Machine learning models to detect anxiety and depression through social media: A scoping review. *Comput Methods Programs Biomed Update* [Internet]. 2022 [cited 2024 Dec 29];2:100066.
- [32] McCance-Katz EF. The Substance Abuse and Mental Health Services Administration (SAMHSA): New Directions. *Psychiatr Serv* [Internet]. 2018 [cited 2025 Jan 5];69(10):1046–1048.
- [33] Tobe RH, MacKay-Brandt A, Lim R, et al. A longitudinal resource for studying connectome development and its psychiatric associations during childhood. *Sci Data* [Internet]. 2022 [cited 2025 Jan 27];9(1):300.
- [34] Alexander LM, Escalera J, Ai L, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data* [Internet]. 2017 [cited 2025 Jan 27];4(1):170181.
- [35] Inter-university Consortium for Political and Social Research (ICPSR) [Internet]. [cited 2025 Mar 12]. Available from: <https://www.icpsr.umich.edu/web/pages/ICPSR/index.html>.
- [36] Chorpita BF, Daleiden EL. Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *J Consult Clin Psychol* [Internet]. 2009 [cited 2025 Feb 10];77(3):566–579.
- [37] Promoting Mental Health and Well-Being in Schools: An Action Guide for School and District Leaders [Internet]. CDC; [cited 2025 Jan 6]. Available from: https://www.cdc.gov/mental-health-action-guide/media/pdfs/DASH_MH_Action_Guide_508.pdf.
- [38] Foti K, Balaji A, Shanklin S. Uses of Youth Risk Behavior Survey and School Health Profiles Data: Applications for Improving Adolescent and School Health. *J Sch Health* [Internet]. 2011 [cited 2025 Jan 6];81(6):345–354.
- [39] Gould DW, Teich JL, Pemberton MR, et al. Behavioral Health in the Gulf Coast Region Following the Deepwater Horizon Oil Spill: Findings from Two Federal Surveys. *J Behav Health Serv Res* [Internet]. 2015 [cited 2025 Jan 6];42(1):6–22.
- [40] Wright AJ, Norris E, Finnerty AN, et al. Ontologies relevant to behaviour change interventions: a method for their development. *Wellcome Open Res* [Internet]. 2020 [cited 2025 Feb 2];5:126.

- [41] Hastings J, Cox S, West R, et al. Addiction Ontology: Applying Basic Formal Ontology in the Addiction domain. Qeios [Internet]. 2020 [cited 2023 Sep 8]; doi: 10.32388/HZHJIP.
- [42] Hastings J, Ceusters W, Jensen M, et al. Representing Mental Functioning: Ontologies for Mental Health and Disease. *Ontol Ment Funct ICBO Workshop Proceedings Third Int Conf Biomed Ontol.* 2012.
- [43] Mental Health Management Ontology [Internet]. [cited 2023 Sep 28]. Available from: <https://bioportal.bioontology.org/ontologies/MHMO>.
- [44] Dang Y, Li F, Hu X, et al. Systematic design and data-driven evaluation of social determinants of health ontology (SDoHO). *J Am Med Inform Assoc* [Internet]. 2023 [cited 2023 Sep 26];30(9):1465-1473.
- [45] Kaplan RM, Beatty AS. Ontologies in the Behavioral Sciences: Accelerating Research and the Spread of Knowledge. [Internet]. National Academies of Sciences, Engineering, and Medicine; Division of Behavioral and Social Sciences and Education; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on Accelerating Behavioral Science through Ontology Development and Use; [cited 2025 Jan 27]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK584329/>.
- [46] Reinecke I, Zoch M, Reich C, et al. The Usage of OHDSI OMOP – A Scoping Review. In: Röhrig R, Beißbarth T, König J, et al., editors. *Stud Health Technol Inform* [Internet]. IOS Press; 2021 [cited 2025 Jan 20]. Available from: <https://ebooks.iospress.nl/doi/10.3233/SHTI210546>.

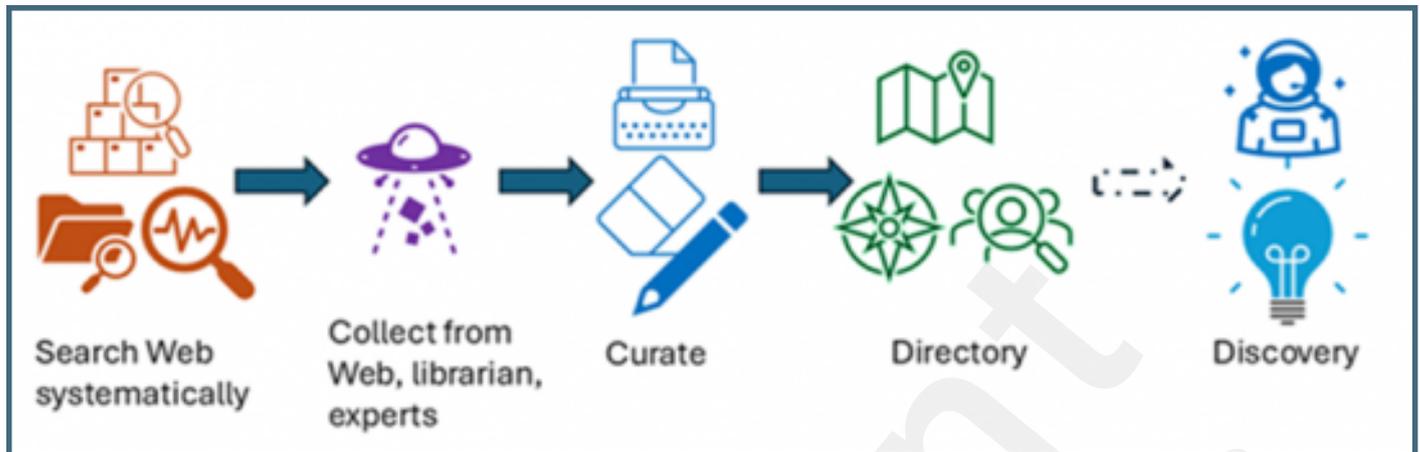
Supplementary Files



Figures



Overall workflow of curating the youth mental health data directory (dashed arrow indicates anticipated usage of the directory).



The summary view of the youth mental health data directory.

