School Psychology

Understanding Adolescent Mental Health Symptom Progression in School-Based Settings: The Substance Use and Risk Factors (SURF) Longitudinal Survey

Meghan A. Costello, Michael Pascale, Kevin Potter, Sarah J. Knoll, Alec Bodolay, Jasmeen Kaur, Rosie Du, Lauren Greenspan, Caroline A. Gray, Joseph McIntyre, and Randi M. Schuster

Online First Publication, February 10, 2025. https://dx.doi.org/10.1037/spq0000681

School Psychology

ISSN: 2578-4218
https://doi.org/10.1037/spq0000681

Understanding Adolescent Mental Health Symptom Progression in School-Based Settings: The Substance Use and Risk Factors (SURF) Longitudinal Survey

Meghan A. Costello[1, 2], Michael Pascale[1], Kevin Potter[1], Sarah J. Knoll[1], Alec Bodolay[1], Jasmeen Kaur[1], Rosie Du[1], Lauren Greenspan[1], Caroline A. Gray[1], Joseph McIntyre[3], and Randi M. Schuster[1, 2]

[1] Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, United States
[2] Harvard Medical School, Harvard University
[3] Harvard Graduate School of Education, Harvard University

The adolescent mental health crisis has prompted a need for an improved understanding of developmental trajectories of psychopathology to promote understanding of risk and protective factors and bolster prevention and intervention efforts. The present study describes the Substance Use and Risk Factors Survey, a universal screener of mental health, substance use, and school and contextual factors administered in public middle and high schools in Massachusetts ($N_{2020} = 3{,}522$, $N_{2021} = 6{,}484$, $N_{2022} = 23{,}915$). Using a seven-question linking code approach, a subset of students were longitudinally linked across administrations ($N_{2020–2022} = 563$, 58% female sex, 30.4% minoritized racial/ethnic identity, $M_{age} = 13.2$ years; $N_{2021–2022} = 1{,}545$, 51% female sex, 28.9% minoritized racial identity, $M_{age} = 14.3$ years). Linking using minimally invasive questions such as the ones presented here may reduce risk, increase privacy, and offer a low-burden opportunity to link observations across time. This work aims to characterize longitudinal trajectories of mental health including substance use in large, community-based samples, as well as the individual-, school-, and community-level risk and protective factors that may modulate the expression of mental health symptoms over time.

Impact and Implications

This study supports the use of a set of nonidentifying linking questions to track adolescent responses to surveys over time without attaching survey responses to their names. This method will allow for school-based studies to look for change in adolescent mental health over time and help us understand what puts adolescents at risk for poorer mental health and what helps them thrive. Ultimately, by providing this information to school psychologists, we can respond to students' needs in real time and support their mental health and well-being.

Keywords: adolescence, mental health, substance use, longitudinal development, school-based screening

Supplemental materials: https://doi.org/10.1037/spq0000681.supp

Adolescence is increasingly recognized as a sensitive period for psychological symptom progression, and rates of adolescent distress have continued to grow over the past several decades (Racine et al., 2021). Longitudinal linkage of community-based survey data would offer an underutilized opportunity to understand adolescent mental health symptom trajectories and risk factors to inform prevention and intervention efforts for this vulnerable population. To facilitate investigations of adolescent psychological development in school-based samples, the present study validates the use of seven minimally invasive survey questions and a simple algorithm for record linkage, demonstrates real usage in large school-based surveys, and provides a package of materials for future use.

The global youth mental health crisis has prompted calls for action, with an emphasis on the need for scalable, evidence-based prevention efforts (Benton et al., 2021). Adolescents have endorsed multifold increases in mental health symptom prevalence and severity over the

Matthew J. Mayer served as action editor.

Meghan A. Costello https://orcid.org/0000-0002-9520-4676

The authors have no conflicts of interest to declare that are relevant to the content of this article. Funds for this study are provided by the Massachusetts Department of Public Health, Office of Youth and Young Adult Services' federal award by the Substance Abuse and Mental Health Services Administration (Grant INTF2400H78500224455; principal investigator: Randi M. Schuster), as well as the Patient-Centered Outcomes Research Institute (Grant AU-2022C1-26355; principal investigator: Randi M. Schuster). Funding support for Meghan A. Costello was provided by the Intramural Research Program, National Institute on Drug Abuse (Grant 5K12DA043490-07).

Correspondence concerning this article should be addressed to Randi M. Schuster, Department of Psychiatry, Massachusetts General Hospital, 60 Staniford Street, Boston, MA 02114, United States. Email: rschuster@mgh.harvard.edu

past decade (Ahn-Horst & Bourgeois, 2024). Alongside the rise in reported symptoms, there has been an increase in rates of emergency department visits following suicide attempts (Twenge et al., 2019). Most adolescents who need mental health treatment do not receive it, despite the recognition that mental health concerns in adolescence are a robust predictor of mental health concerns across the lifespan (Uhlhaas et al., 2023). Understanding the progression of increasing mental health symptoms and associated risk factors is a necessary public health task. The present study seeks to provide a streamlined, accessible protocol and materials for use by researchers and schools to comprehensively survey their student population and link student responses over time without requiring students' records to be identified.

Surveillance of youth mental health has largely been accomplished through anonymous, cross-sectional population studies (e.g., Youth Risk Behavior Survey, Monitoring the Future; Miech et al., 2023). These studies offer large-scale, representative health information, which is key to documenting prevalence rates, monitoring for emerging trends, and defining current patterns of comorbidity. However, cross-sectional surveys are limited in their ability to investigate early risk factors and longitudinal patterns of mental health symptom progression. Well-characterized cohort studies, in contrast, allow for deep longitudinal phenotyping, often of clinical samples around treatment transitions (e.g., Blanco et al., 2017; Lawson et al., 2020). However, findings from clinical cohort studies may not generalize to youth with pre- or subclinical symptoms (Blanco et al., 2017), who are prime candidates for early prevention and intervention efforts through schools.

Linking data in community cohorts increases the ability to define symptom trajectories in generalizable samples (Shek & Ng, 2016). By linking data, investigators can characterize symptom trajectories and glean a better understanding of how psychological problems develop over time (Shek & Ng, 2016). Early identification of risk factors for symptom progression may be used as indicators for targeted intervention efforts *before* clinical symptoms expound into full-blown clinical concerns. Schools offer a rich opportunity to characterize adolescent mental health trajectories and identify points for early intervention. Over 17 million adolescents were enrolled in public or private schools in the United States in fall 2021 (Fabina et al., 2023). Adolescents in the United States have far more consistent contact with schools than any other system, making them the ideal setting for studying the naturalistic development of mental health distress. Despite benefits, longitudinal linking has been difficult to achieve in school-based surveys because linking can be burdensome at the point of survey administration. Linking often requires a cipher, meaning that a member of school or study staff must individually assign a survey with an ID to a specific student at each administration timepoint (increasingly difficult as the number of students participating in the survey grows). Additionally, the process of assigning and tracking IDs may require linking ID numbers with student names, which decreases perceptions of anonymity and willingness to participate. Previous work has generated creative solutions to these challenges, by using survey metadata or participant-generated ID codes (Qian et al., 2015; Vacek et al., 2017). More recently, researchers have moved toward scaffolding ID generation through a series of questions dubbed a "secret code" (Ripper et al., 2017; Schnell et al., 2010), but, to our knowledge, this approach has not been applied to large, school-based data collections. Therefore, the present study evaluates the feasibility and validity of applying the secret code-linking approach to a large sample of adolescents taking surveys in schools.

To develop early intervention targets, we must understand common trajectories of mental health symptom progression, especially in minoritized populations, and during developmental windows of heightened sensitivity. This article uses repeated survey assessments of substance use and mental health in middle and high schools across Massachusetts to characterize symptom trajectories and put forth a protocol for linking records across survey years using a set of minimally identifiable questions. Specifically, in this article, we set out to accomplish the following aims:

1. Describe the Substance Use and Risk Factors (SURF) Survey, including samples of students surveyed and study procedures across 3 years of data collection in Massachusetts schools from 2020 to 2022.
2. Establish the validity of a question-based longitudinal record linking code in the SURF Survey.
3. Apply the linking code and describe the linked subsample as a proof of concept.

Method

Participating Schools

The SURF Survey is an annual survey administered to middle and high school students (Grades 6–12) in Massachusetts that queries demographics, substance use, co-occurring other mental health symptoms, and associated risk and protective factors. The SURF Survey began in 2016 with an initial survey administered to one school. Linking questions were added to the survey in 2020, the SURF Survey longitudinal baseline. The present study describes sample and methods from the 2020 (six schools, 3,522 students), 2021 (14 schools, 6,484 students), and 2022 (60 schools, 23,915 students) administrations.

Survey Consent

Student participation occurred through an opt-out consent procedure (distributed in students' preferred language based on reported school need), in which caregivers were given the option to withdraw their child from the optional school-wide survey (see Supplemental Material S1 for sample opt-out letter). If caregivers did not contact study staff prior to the survey date, passive consent for survey participation was assumed. The SURF Survey was voluntary, with student assent implied by completing the survey. Participating schools and districts were aware of and in agreement with the opt-out and passive consent process.

Survey Administration

Surveys were completed electronically during school periods (e.g., homerooms, advisory). Proctoring teachers introduced the survey using an institutional review board-approved script describing the purpose of the study and informing students that participation was voluntary (see Supplemental Material S2 for full teacher script). Students were also provided a description of the survey in writing which communicated the purpose of the survey, that the survey is voluntary, and that their responses would be kept confidential, and to

communicate to the study team with any questions (see Supplemental Material S3 for the full survey). Students entered responses directly into REDCap, a Health Insurance Portability and Accountability Act-compliant platform for electronic data capture hosted on internal hospital servers. SURF Survey and accompanying opt-out consent documents were available in Amharic, Arabic, Bengali, Brazilian Portuguese, Chinese Mandarin (simplified), Dari, English, Farsi, Filipino, French, Gujarati, Haitian Creole, Russian, Spanish, Turkish, Ukrainian, and Vietnamese. Starting in 2022, using a multilanguage feature of REDCap, students could select their preferred language.

Each school accessed the survey using a unique link. School IDs assigned by the Massachusetts Department of Elementary and Secondary Education (DESE) were collated into the SURF Survey database to facilitate linkage back to DESE reports as well as census and other public data sets. Data were kept confidential and stored in a secure, coded data set. To align with the core principles of community-based research, aggregated summary reports of survey data were shared back with the school within a few weeks. Each school received a report tailored to them and did not receive information about any other schools beyond state-wide aggregated data. Reports provided summaries of survey data on reported student demographic information, substance use, and other psychological symptoms aggregated at the level of the target school and the state (see Supplemental Material S3 for a sample de-identified school report from 2022).

Measures

SURF Survey measures include a combination of single-item questions, validated symptom screeners, research-based assessments, and adapted questionnaires of interest. Items not part of psychometrically validated measures were modeled after similar national and regional surveys that assess adolescent symptoms and substance use, such as the Youth Risk Behavior Survey (Mpofu et al., 2023), Monitoring the Future (Miech et al., 2023), MetroWest (Schneider et al., 2012), Population Assessment of Tobacco and Health (Hyland et al., 2017), and National Survey on Drug Use and Health (Hedden et al., 2012). Branching logic was applied to query specific information where relevant (e.g., past 30-day substance use only queried among those who endorsed lifetime use of that substance). A summary of survey components by year is presented in Table 1 (see Supplemental Material S4 for a sample survey).

Table 1
Summary Substance Use and Risk Factors Survey Measures (2020–2022)

Survey measure	Year administered		
	2020	2021	2022
Demographics			
• Grade[a]	×	×	×
• Age[a]	×	×	×
• Sex[a]	×	×	×
• Gender identity[a]	×	×	×
• Sexual orientation[a]		×	×
• Race[a]	×	×	×
• Ethnicity[a]	×	×	×
• Country of birth[a]			×
• Adoption status[a]			×
Substance use			
• Lifetime substance use	×	×	×
• Past 30-day frequency of alcohol/cannabis/nicotine	×	×	×
• Intent to quit alcohol/cannabis/nicotine[a]			×
• Craving for cannabis/nicotine[a]			×
• Psychotic experiences during cannabis use[a]			×
• Methods of cannabis use[a]			×
• Use of nicotine flavors[a]			×
Other behavioral and psychological functioning			
• Psychoticlike symptoms	×	×	×
• Emotional reactivity	×	×	×
• Symptoms of depression and anxiety	×	×	×
• Suicidal thoughts and behaviors			×
• Symptoms of inattention and hyperactivity[a]			×
• Risk taking[a]			×
• Impact of COVID-19 on mental health[a]	×		
• Physical activity[a]			×
• Psychiatric medication use[a]			×
• Formal and informal help seeking[a]	×	×	×
School and other contextual factors			
• Racial/ethnic discrimination and associated distress			×
• Experiences of discrimination from other aspects of identity			×
• School experiences (e.g., sports, school performance, discipline, individualized education plan)			×
Linking, data quality, and follow-up			
• Attention checks			×
• Linking questions		×	×
• Interest in follow-up	×	×	×

[a] Described in Supplemental Material S5.

Demographic Characteristics

Participants reported on grade, age, sex, gender identity, sexual orientation, race, and ethnicity. Adoption status and country of birth were queried in 2022. See Supplemental Material S5 for detailed item descriptions.

Substance Use

Lifetime Substance Use. Students responded yes or no to the question "Have you ever used …" with regard to the following substances: "at least one full drink of alcohol," "marijuana (e.g., pot, weed, cannabis, THC, dab pens, edibles)," "a vape for nicotine or flavors (vapes include e-cigarettes, vape pens, e-cigars, e-hookahs, hookah pens, mods, and other electronic vapor products [e.g., Puff Bar, JUUL, SMOK, Suorin, Vuse, and blu])," "smoking a cigarette, smoking a cigar, cigarillo, or little cigar (e.g., Black and Mild, Swisher Sweet, Phillies, Backwoods)," and "smokeless tobacco (e.g., chewing tobacco, snuff dip, snus, or dissolvable tobacco products)." Follow-up questions regarding use of these most commonly used substances were administered to students who indicated lifetime use of alcohol, cannabis, and nicotine.

Students were also queried about their lifetime use of the following substances (check all that apply): "prescription drugs not used as prescribed (e.g., Adderall, Ritalin, Xanax, oxycodone, codeine)," "hallucinogens or dissociative drugs (e.g., mushrooms, LSD, Acid, PCP, ketamine, DXM)," "club drugs (e.g., Ecstasy, MDMA, Molly, GHB), cocaine (e.g., powder, crack, or freebase)," "Methamphetamine (also called speed, crystal meth, crank, ice, or meth)," "heroin or fentanyl (e.g., smack, junk, or China White),"

"inhalants (e.g., whippets, sniffed glue, breathed the contents or aerosol spray cans, or inhaled any paints or sprays)," and "anabolic steroids (e.g., juice, pumpers, tren, dbol, roids)." Students also had the option to indicate "I have never tried any of the above drugs."

Past 30-Day Frequency. If students indicated "yes" to lifetime use of alcohol, cannabis, or nicotine, detailed follow-up questions were administered regarding past 30-day use of the relevant substance because these are the most used substances in adolescence (Miech et al., 2023). Students were asked "In the past 4 weeks, (on average) how often did you use [substance]" and were instructed to respond using the scale: *0 times*, *1 time*, *less than once per week*, *at least once per week*, *2–3 days per week*, *4–6 days per week*, or *every day*. Follow-up questions were asked of students who indicated past 30-day use (including intent to quit, craving, psychotic experiences during cannabis use, methods of cannabis use, and use of nicotine flavors; measures described in Supplemental Material S5).

Other Behavioral and Psychological Functioning

Psychoticlike Symptoms. Students completed the Adolescent Psychotic-Like Symptom Screener (APSS; Kelleher et al., 2011), which is a seven-item screener that assesses subsyndromal psychotic symptoms. Items query visual and auditory hallucinations, paranoia, grandiosity, delusions of control, mind reading, and delusions of reference. Responses to each item are scored 0 (*no, never*), 0.5 (*maybe*), and 1 (*yes, definitely*). Scores on each item are summed to a total score (range = 0–7, clinical cutoff of 2). The APSS has been validated and sensitive to psychotic symptoms in general adolescent populations (Kelleher et al., 2011). Internal consistency in this sample was acceptable (Cronbach's αs = .79, .79, and .81 in 2020, 2021, and 2022, respectively).

Emotional Reactivity. Emotional reactivity, or the extent of an individual's emotional sensitivity, intensity, and persistence, was assessed using the Emotional Reactivity Scale (ERS; Nock et al., 2008). Students responded to 21 items by endorsing the extent to which the phenomena described sounds "like them," using the scale 0 (*not at all like me*), 1 (*a little like me*), 2 (*somewhat like me*), 3 (*a lot like me*), and 4 (*completely like me*). Higher total sum scores indicate higher levels of emotion reactivity (range = 0–84). The ERS yields three subscales: persistence (four items, such as "When something happens that upsets me, it's all I can think about for a long time."), sensitivity (10 items, such as "I tend to get emotional very easily."), and intensity of arousal (seven items, such as "My moods are very strong and powerful."). This questionnaire has demonstrated strong internal consistency and validity with adolescent samples (Nock et al., 2008). Internal consistency for this sample was good to excellent; Cronbach's αs = .96, .96, and .97 for the full scale; .81, .83, and .86 for the persistence subscale; .91, .92, and .94 for the sensitivity subscale; and .90, .91, and .93 for the intensity of arousal subscale.

Symptoms of Depression and Anxiety. Symptoms of depression and anxiety were screened using the Patient Health Questionnaire–Four Item Version, commonly used in primary care and pediatrics (Kroenke et al., 2009). This screener includes a two-item depression screener, the Patient Health Questionnaire–Two Item and two-item anxiety screener, the Generalized Anxiety Disorder–Two Item. Students were asked how often they had been bothered by symptoms of depression and anxiety over the past 2 weeks. Students responded on a scale describing the frequency of experiencing symptoms over the prior 2 weeks, where 0 = *not at all*, 1 = *several days*, 2 = *more than half the days*, and 3 = *nearly every day*. Scores can be calculated for overall distress (all items), total depression (Patient Health Questionnaire–Two Item), and total anxiety (Generalized Anxiety Disorder–Two Item). This measure has been validated with adolescents, with total score thresholds of scores less than 2 = "normal," 2–5 = "mild," 5–8 = "moderate," and 8–12 = "severe" (Löwe et al., 2010). The Patient Health Questionnaire–Two Item and Generalized Anxiety Disorder–Two Item are typically administered with clinical cutoff scores of 3 (Plummer et al., 2016). Internal consistency in this sample was acceptable to good; Cronbach's αs = .84, .85, and .88 for the full scale; .86, .86, and .87 for the anxiety symptom subscale; and .70, .74, and .79 for the depression symptom subscale.

Suicidal Thoughts and Behaviors. Students answered four questions on past-year suicidal thoughts (ideation, suicidal plan), suicidal attempts, and nonsuicidal self-injury. Students responded yes or no to the following questions: "Did you ever have thoughts about killing yourself (ending your life)?" "Did you think about how you would kill yourself?" "Did you try to kill yourself?" "Did you hurt yourself on purpose without trying to kill yourself?" All students, regardless of their responses, received the following information about resources:

> If you are having thoughts or feelings of suicide, please know there are people who can help. Here are some numbers you can use 24 hours a day, 7 days a week to talk to someone who cares. Crisis Text Line: 741741. National Suicide Prevention Lifeline: 9-8-8. Please also make sure to talk to a trusted adult if you ever feel like you cannot keep yourself safe or need extra support.

School and Other Contextual Factors

Racial/Ethnic Discrimination and Associated Distress. Starting in 2022, students who were in at least ninth grade reported on exposure to race/ethnic-based discrimination and distress associated with those exposures using the Adolescent Discrimination Distress Index (Fisher et al., 2000). Students responded to the following question:

> After each statement, tell us if you have experienced each of the following types of discrimination because of your race or ethnicity. Remember, for these questions, we are only interested in occasions when racial–ethnic discrimination was at least partly responsible for your experience.

Students responded to 15 items describing discriminatory experiences (e.g., "You were wrongly disciplined or given after-school detention because of your race or ethnicity"). For each discriminatory experience a student reported, they were then asked to rate how upsetting that experience was, on a scale of 1 (*not at all*), 2 (*slightly*), 3 (*moderately*), 4 (*considerably*), and 5 (*extremely*). Scores were then averaged to generate an overall discrimination distress score, with higher scores indicating more discrimination distress. This measure is comprised of three subscales that can be calculated to assess the degree of institutional discrimination (six items), educational discrimination (four items), and peer discrimination (five items). Internal consistency for this sample was good for the full scale and acceptable for most subscales; Cronbach's α for full scale = .86, for institutional discrimination = .78, for educational discrimination = .56, and for peer discrimination = .69.

Experiences of Discrimination Based on Aspects of Identity Not Related to Race and/or Ethnicity. Students were asked: "In the past 12 months, have you experienced discrimination related to any of the following aspects of who you are? Check all that apply."

Students had the option to report on discrimination related to gender, sexual orientation, religion disability, socioeconomic status, and "other" personal identities; they also had the option to indicate that they had not experienced discrimination in the past 12 months. If "other" was selected, participants were presented with a free response text field to describe the identities for which they experienced discrimination over the past year. After students completed this scale, responses were scored dichotomously as 0 (*no attribution of discrimination to the listed identity*) and 1 (*attribution of discrimination to the listed identity*).

School Experiences. Students responded yes or no to "During the past 12 months, did you play on a school sports team?" Students responded to the question: "During the past 12 months, how would you describe your grades in school?" using the scale *mostly As*, *mostly Bs*, *mostly Cs*, *mostly Ds*, *mostly Fs*, *mixed grades that range from As and Bs to Cs and Ds*, *none of these grades*, and *not sure*.

Students indicated whether they have an individualized education program (IEP) in school by responding "yes," "no," or "not sure" to the following question: "Do you have an IEP (Individualized Education Program) at school? An IEP is a plan or program developed by school for extra supports or services to help students who have difficulties that sometimes make learning in school challenging." When students indicated "yes" that they have an IEP, they were asked "If you have an IEP, what is it for? Check all that apply": "ADD/ADHD," "other learning difficulty (e.g., for math, writing, reading/dyslexia)," "autism spectrum disorder," "emotional concerns," "I am not sure what my IEP is for," and/or "something else not listed here."

Regarding discipline and substance use in school, students indicated "yes" or "no" to the following questions. First, students were asked: "During the past 12 months, were you suspended from school for any reason?" If students indicated that they had been suspended in the past year, they were asked: "Were you suspended from school for something having to do with alcohol, vaping, or other drugs?" Students who endorsed lifetime substance use were asked: "Have you used alcohol, marijuana, nicotine (vapes, cigarettes, etc.), or other drugs on school property?"

Linking, Data Quality, and Follow-Up Questions

Attention Checks. Beginning in 2022, two attention check questions were included: "Please choose [specific answer] for this question. This is just to see if you are paying attention." In this study, records were excluded if the student failed both attention checks.

Completion Checks. Records were flagged based on the degree of completion of nonbranched questions (i.e., asked of the entire sample). Beginning 2022, records were excluded if less than 60% of nonbranched questions were complete.

Interest in Follow-up. For a subset of schools across all survey years, students were invited to provide study staff with their contact information if they were interested in being included in a research recruitment database. Contact information provided in response to this question was also used to create a subset of 94 confirmed linked pairs from 2021 to 2022, based on matching information provided at the two time points. This confirmed linked data set was used to validate the linking procedure (described below).

Linking Questions. The SURF Survey included a set of questions intended to link records across years without using their name or other personal identifiers (see Table 2). Questions were selected because responses were likely to be static over time and easily understood by students.

Longitudinal Record Linking Procedure

Records were matched across assessment years by cohort within the school district to restrict the number of possible longitudinal linkages to those that were most likely. Cohort was defined as the projected year of graduating Grade 12 plus or minus 1 year to account for students who skip or repeat a grade. Restricting linkages to school district allows for record linkage even if students move schools within a district. Dissimilarity scores were calculated based on the total number of linking questions minus the number of the exact matches between records for the questions across assessment years (i.e., 0 indicated all responses matched, 1–7 indicated a degree of mismatch from one question different to seven questions). Records with a dissimilarity score of 0 were deemed linked and assigned a participant ID. In the case of a "tie" (i.e., multiple records have identical responses), the first records listed are assigned to be linked. For all records that remained unlinked using all seven linking questions, this process was repeated using different combinations of the subsets of six (i.e., "one-missing" subsets) of linking questions, assigning unique ID numbers to sets of newly linked records at each step. See Figure 1 for a schematic of the linking process, Supplemental

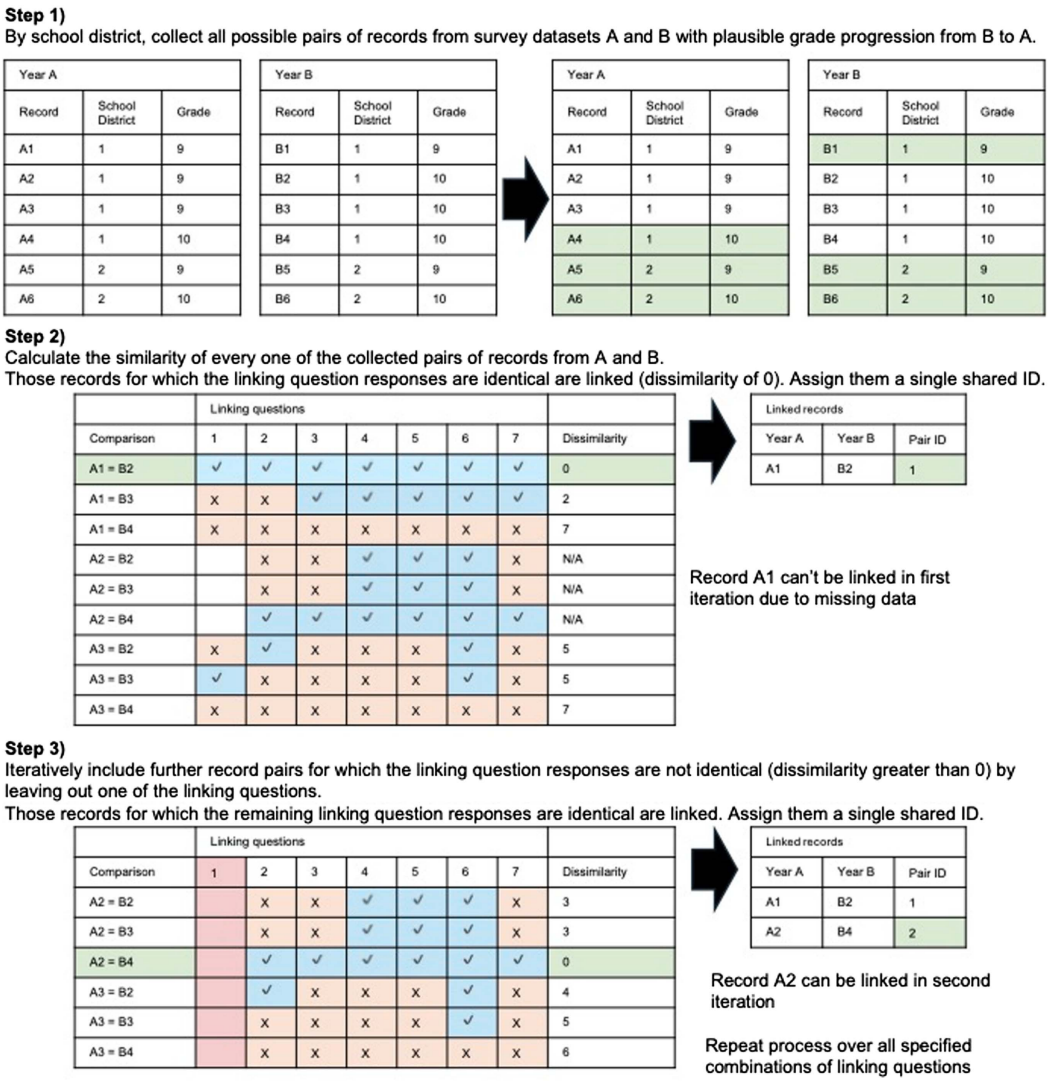
Table 2
Substance Use and Risk Factors (SURF) Linking Code Items and Responses

Linking question	Possible response
1. What month were you born in?	(January–December)
2. What is your sex? This question is asking about your sex assigned at birth.	(Male, female)
3. How many older siblings do you have? a. (if >0) What month was your oldest sibling born in?	(0–3+) (January–December)
4. What is the third letter of your first name?	(a–z)
5. Do you have a middle name? a. (if yes) What is the first letter of your middle name?	(Yes, no) (a–z)
6. What best describes the color of your eyes?	(Black, brown, blue, green, hazel, gray)
7. What are the first three letters of the street you lived on in the beginning of fifth grade?	(a–z)

Note. Question phrasing changed slightly as surveys were developed. Questions presented above are planned to be used for all subsequent administrations after 2022.

Figure 1

Linking Algorithm Schematic



Note. NA = not available; ID = participant identification number. See the online article for the color version of this figure.

Material S6 for pseudocode algorithm, and Supplemental Material S7 for R syntax.

Analytic Approach: Evaluation of Linking Accuracy

To validate linking code performance, a series of simulation studies were run. First, to evaluate response consistency in the linking code questions across assessment time points, we used a subset of 94 confirmed linked pairs (188 records) from high schoolers in 2021–2022, henceforth referenced as the "confirmed, linked sample." These 94 record pairs were identified based on voluntarily provided student contact information that was consistent across 2021 and 2022 (see the Interest in Follow-Up section measure description).

Data were simulated for two separate assessment years via resampling, using as a basis actual data from 9th- to 11th-grade students from the 2021 survey ($N = 16{,}011$ records). Students were randomly assigned with a 50% probability to have either records that should be linked across years (true positives, duplicates of confirmed linked records) or records that should not be linked (true negatives, random samples from all possible response options). Missingness was simulated for each record, commensurate with observed missingness (93.4% missing zero questions, 5.45% missing one, 0.67% missing two, 0.26% missing three, 0.16% missing four, 0.04% missing five, and 0.02% missing six). Partial records were generated to represent the introduction of new students at the 2022 data collection year. Record linkage was then attempted using the stepwise approach described above to evaluate the sensitivity of different combinations of linking questions. These steps were repeated 30 times, with new data each time, to obtain estimates of the variability in the true positive and true negative rates.

Results

Aim 1: Describe the Complete Surveyed Sample Across 3 Years of Data Collection in Massachusetts High Schools From 2020 to 2022

Large samples of students in Massachusetts high schools completed the SURF Survey each year (N_{2020} = 3,522; N_{2021} = 6,484; N_{2022} = 23,915; see Table 3). Beginning in 2022, records were filtered for at least 60% completion who passed attention checks. As a result, in 2022, a total of 4,855 records were dropped (4,085 with 0% completion, likely reflecting survey connectivity challenges). Surveyed schools included 24 middle schools (sixth to eighth grade), 34 high schools (ninth to 12th grade), and three combined (seventh to 12th grade). Schools included four charter schools, two vocational–technical schools, one therapeutic day school, one arts-focused magnet school, and one private college preparatory school. Mean percent of total school enrollment surveyed was 75.5% (SD = 11.8%) in 2020, 83.3% (SD = 11.0%) in 2021, and 72.7% (SD = 16.7%) in 2022. Opt-out rates within each school were relatively small (2020: M = 1.0% [SD = 0.7%]; 2021: M = 1.0% [SD = 0.4%]; 2022: M = 3.7% [SD = 3.5%]). Students completed the survey in a median time of 10 min in 2020, 11 min in 2021, and 14 min in 2022.

The mean age of the sample was about 14 years old each year, reflecting the capture of students in sixth through 12th grades ($M_{\text{age 2020}}$ = 14.1 years, SD = 2.0 years; $M_{\text{age 2021}}$ = 14.3 years, SD = 2.0 years; $M_{\text{age 2022}}$ = 14.7 years, SD = 1.9 years). The sample is comprised of primarily boys (45.6%–47.2%) and girls (46.2%–51.1%) and includes students who identified with a minoritized gender identity, including trans boy/man (0.2%–0.8%), trans girl/woman (0.2%–0.3%), and another gender (0%–3.1%). Each year, the sample consisted primarily of white students (63.4%–68.6%) and included students who identified as Asian (5.9%–18.9%), Black/Haitian/African American (4.1%–5.9%), holding multiple racial identities (0%–8.3%), and other racial identities (3.9%–8.0%). A small percentage of students who identified as American Indian/Alaskan Native (0.5%–1.1%), Hawaiian/Pacific Islander (0.1%–.2%), and Middle Eastern/North African (0%–1.5%) were represented most years. The sample increasingly identified as ethnically Hispanic/Latino/a (7.2% in 2020, 20.6% in 2022).

In 2021 and 2022, most students identified as straight (74%–76.2%); the remaining students identified primarily as bisexual (7.8%–8.4%), with small subsets of the sample identifying as gay or lesbian (2.8%–3.2%), unsure (4.4%–5.7%), asexual (0%–1.1%), something else (4.2%–3.6%), or noting that they do not want to say (2.2%–3.6%). In 2022, most students reported they were born in the United States (88.8%) and that they were not adopted (94.3%).

Follow-up analyses generated three sets of post-stratification weights that were created by raking the sample data to match population data from DESE. For all weights, we raked on student race and grade level, which were the only variables included on both the survey and in DESE's data. Weights match the sample to (a) the population of students in Massachusetts, (b) the population of students in the districts sampled, and (c) the population of the county in which their district was located (this final set of weights has not been adjusted to account for varying sizes of counties and is only appropriate for within-county analyses). Survey weights will be distributed, alongside survey data, to analytic collaborators.

Aim 2: Establish the Validity of a Question-Based Longitudinal Record Linkage Code in the SURF Annual School Survey

Rates of Consistent Responses in Known Linked Records

Table 4 reports the proportion of consistent responding per each linking question for the confirmed, linked sample (i.e., the 94 students who provided records in 2021 and 2022 that could be linked with confidence based on voluntarily provided names and contact information).

The least consistent responding occurred for a number of siblings, eye color, and the first three letters of a student's fifth-grade street address. Based on these results, linking accuracy was examined based on (a) excluding the items for older siblings and eye color and (b) excluding the items for older siblings, eye color, and street address. Question 4 ("What is the 3rd letter of your first name?") was added in 2022 and is therefore excluded from linking analyses in this article.

Confidence Linking Records Based on Linking Questions

Table 5 reports the percentages and 95% confidence intervals (based on the β distribution) for the true positive and true negative rates over 30 simulations of data per the defined resampling approach from the 2022 SURF survey data set.

Linking only on the full set of seven linking questions results in an extremely high true negative rate (99.9%–100.0%), but a noticeably lower true positive rate (68.2%–70.3%). Iteratively incorporating records linked using all possible subsets of six items (one-missing) substantially improved the true positive rate (93.1%–94.4%). Iteratively incorporating records linked using the subset of five items without the two items with the highest inconsistency rate (older siblings and eye color) had no discernible impact on rates. By contrast, iteratively incorporating records linked using the subset of five items without eye color and street improved the true positive rate (94.2%–95.3%) and had a statistically significant but small reduction of the true negative rate (99.6%–99.8%). Finally, iteratively incorporating records linked using the subset of four items excluding the three with the highest inconsistency rate (older siblings, eye color, and street address) improved the true positive even further (95.8%–96.8%) but at the expense of a sizeable reduction in the true negative rate (92.3%–93.3%).

Aim 3: Apply the Linking Code and Describe the Linked Subsample as a Proof of Concept

The linking procedure identified students with repeated observations (see Table 6).

When compared with nonlinked students in their baseline year (2020 or 2021, excluding 12th graders because they were not eligible for linking), in 2020 linked students were on average slightly younger (t_{2020} = 8.4, $p < .001$, $M_{\text{age linked 2020}}$ = 13.2,

Table 3
Sample Demographic Characteristics

Demographic characteristic	2020 (N = 3,522, six schools)		2021 (N = 6,484, 13 schools)		2022 (N = 23,915, 60 schools)	
	N	%	*N*	%	*N*	%
Age	8 years: 0	0.0	8 years: 1	0.0	8 years: 1	0.0
	9 years: 4	0.1	9 years: 2	0.0	9 years: 1	0.0
	10 years: 3	0.1	10 years: 15	0.2	10 years: 3	0.0
	11 years: 389	9.8	11 years: 465	7.2	11 years: 1,219	5.1
	12 years: 540	13.6	12 years: 981	15.1	12 years: 2,453	10.3
	13 years: 567	14.3	13 years: 1,015	15.7	13 years: 2,836	11.9
	14 years: 512	12.9	14 years: 972	15.0	14 years: 3,985	16.7
	15 years: 513	12.9	15 years: 1,035	16.0	15 years: 4,326	18.1
	16 years: 449	11.3	16 years: 951	14.7	16 years: 3,872	16.2
	17 years: 437	11.0	17 years: 809	12.5	17 years: 3,547	14.8
	18 years: 86	2.2	18 years: 178	2.8	18 years: 1,297	5.4
	19 years: 4	0.1	19 years: 4	0.1	19 years: 78	0.3
	20 years: 1	0.0	20 years: 1	0.0	20 years: 8	0.0
	21 years: 0	0.0	21 years: 1	0.0	21 years: 9	0.0
	22 years: 2	0.1	22 years: 2	0.0	22 years: 19	0.1
	Missing: 15	*0.4*	*Missing: 51*	*0.8*	*Missing: 261*	*1.1*
Grade	Fifth: 0	0.0	Fifth: 21	0.3	Fifth: 0	0.0
	Sixth: 500	12.4	Sixth: 641	9.9	Sixth: 2,077	8.7
	Seventh: 568	14.1	Seventh: 1,004	15.5	Seventh: 2,717	11.4
	Eighth: 559	13.9	Eighth: 1,023	15.8	Eighth: 3,027	12.7
	Ninth: 501	12.5	Ninth: 1,004	15.5	Ninth: 4,659	19.5
	10th: 501	12.5	10th: 998	15.4	10th: 4,139	17.3
	11th: 450	11.2	11th: 945	14.6	11th: 3,823	16.0
	12th: 427	10.7	12th: 802	12.4	12th: 3,390	14.2
	Missing: 16	*0.4*	*Missing: 46*	*0.7*	*Missing: 83*	*0.0*
Gender	Boy/man/male: 1,637	46.5	Boy/man/male: 3,061	47.2	Boy/man/male: 10,911	45.6
	Girl/woman/female: 1,801	51.1	Girl/woman/female: 2,997	46.2	Girl/woman/female: 11,415	47.7
	Trans boy/man: 7	0.2	Trans boy/man: 47	0.7	Trans boy/man: 180	0.8
	Trans girl/woman: 10	0.3	Trans girl/woman: 13	0.2	Trans girl/woman: 77	0.0
	Another gender: 0	0.0	Another gender: 66	1.0	Another gender: 745	3.1
	Don't want to say: 0	0.0	Don't want to say: 40	0.6	Don't want to say: 243	1.0
	Not sure: 0	0.0	Not sure: 97	1.5	Not sure: 256	1.1
	Missing: 67	*1.9*	*Missing: 163*	*2.5*	*Missing: 88*	*0.4*
Sexual orientation	Not assessed		Straight: 2,891	76.2	Straight: 21,299	74.0
			Bisexual: 319	8.4	Bisexual: 2,256	7.8
			Gay or lesbian: 123	3.2	Gay or lesbian: 799	2.8
			Asexual: 0	0.0	Asexual: 303	1.1
			Something else: 138	3.6	Something else: 1,197	4.2
			Unsure: 167	4.4	Unsure: 1,629	5.7
			I don't want to say: 85	2.2	I don't want to say: 1,030	3.6
			Missing: 72	*1.9*	*Missing:* 257	*0.0*
			Not queried (below ninth grade): 2,689			
Race	American Indian/Alaskan Native: 39	1.1	American Indian/Alaskan Native: 48	0.7	American Indian/Alaskan Native: 124	0.5
	Asian: 666	18.9	Asian: 892	13.8	Asian: 1,398	5.9
	Haitian/Black/African American: 143	4.1	Haitian/Black/African American: 381	5.9	Haitian/Black/African American: 1,391	5.8
	Hawaiian/Pacific Islander: 5	0.1	Hawaiian/Pacific Islander: 5	0.1	Hawaiian/Pacific Islander: 36	0.2
	Middle Eastern/Northern African: 0	0.0	Middle Eastern/Northern African: 85	1.3	Middle Eastern/Northern African: 346	1.5
	Multiple: 0	0.0	Multiple: 500	7.7	Multiple: 1,993	8.3
	Other: 137	3.9	Other: 277	4.3	Other: 1,903	8.0
	White: 2,338	66.4	White: 4,112	63.4	White: 16,398	68.6
	Missing: 194	*5.5*	*Missing: 184*	*2.8*	*Missing: 326*	*1.4*
Ethnicity	Hispanic/Latino/a: 252	7.2	Hispanic/Latino/a: 675	10.4	Hispanic/Latino/a: 4,934	20.6
	Not Hispanic/Latino/a: 3,242	92.1	Not Hispanic/Latino/a: 5,663	87.3	Not Hispanic/Latino/a: 18,705	78.2
	Missing: 28	*0.8*	*Missing: 136*	*2.1*	*Missing: 276*	*1.2*
Country of birth					United States: 21,231	88.8
					Other: 2,533	10.6
					Missing: 151	*0.6*
Adoption status					Adopted: 492	2.5
					Not adopted: 18,480	94.3
					Not sure: 317	1.6
					Not queried (below ninth grade): 9,170	

Note. Sexual orientation only assessed in students in or above ninth grade in 2021. Adoption status only in students in or above ninth grade in 2022. Percentages reflect only students who assessed, which may not sum to 100 due to rounding. Beginning in 2022, records were filtered for at least 60% completion.

Table 4
Linking Code Responses in Confirmed Linked Records (2021–2022)

Item	% consistent (*n*/*N*)
1. What month were you born in?	99 (93/94)
2. What is your sex? This question is asking about your sex assigned at birth.	100 (94/94)
3. How many older siblings do you have? (if >0) 3a. What month was your oldest sibling born in?	85 (80/94)
4. What is the third letter of your first name?	Only administered in 2022
5. Do you have a middle name? (if yes) 5a. What is the first letter of your middle name?	94 (88/94)
6. What best describes the color of your eyes?	89 (84/94)
7. What are the first three letters of the street you lived on in the beginning of fifth grade?	90 (85/94)

$M_{\text{age nonlinked 2020}} = 13.8$) and in lower grades ($t_{2020} = 8.0$, $p < .001$, $M_{\text{grade linked 2020}} = 8.0$, $M_{\text{grade nonlinked 2020}} = 8.5$). In 2021, linked students were slightly older ($t_{2021} = -5.5$, $p < .001$, $M_{\text{age linked 2021}} = 14.0$, $M_{\text{age nonlinked 2021}} = 13.7$) and in higher grades ($t_{2021} = -6.9$, $p < .001$, $M_{\text{grade linked 2021}} = 8.9$, $M_{\text{grade nonlinked 2021}} = 8.5$). Minoritized students were less likely to be linked than nonminoritized counterparts regarding racial identity ($t_{2020} = 1.8$, $p = .06$; $t_{2021} = 8.4$, $p < .001$), Hispanic/Latino(a) ethnic identity ($t_{2020} = 4.6$, $p < .001$; $t_{2021} = 6.8$, $p < .001$), and gender identity ($t_{2020} = 3.0$, $p < .001$; $t_{2021} = 2.8$, $p < .01$). No differences were found regarding sexual orientation.

Exploratory Data Visualization Using Linking Code

As a purely exploratory follow-up, and to illustrate the potential application of linked data to identify individual trajectories, a series of plots were generated using the 2021–2022 linked data set (N = 1,545; see Figure 2).

This visualization demonstrates the degree of variation in mental health and substance use trajectories across 1 year for this subset of the sample. Although the present study does not offer any hypotheses or statistical analysis of the data, future work will pursue characterizations of these trajectories and of surrounding factors that predict stability, increase, or decrease in endorsed symptomatology.

Discussion

To promote adolescent mental health, a better understanding of individual, naturalistic mental health trajectories in community-based samples is needed. This study described the protocol for ongoing school-based survey procedures and demonstrated the validity and feasibility of longitudinal linking using seven minimally invasive questions. This study illustrates that in a school-based sample, stability, increases, and decreases in reported mental health and substance use can be tracked over time. This work provides the foundation for ongoing efforts to characterize individual, school, and sample-wide trajectories of mental health symptoms for large, community-based groups of adolescents.

Application of the Linking Protocol

The linking protocol successfully identified consecutive survey responses across 2 and 3 years. The linking algorithm functioned conservatively, with very few false positives in a synthetic data set. Sequentially adding linked records based on subsets of six questions bolstered the true positive rate while remaining conservative about true negatives. This allows for confidence in the linking method's ability to tolerate typical data inconsistencies, in that it allows for an individual to skip, miss, or differently answer linking questions from year to year and still be appropriately linked to their other data.

Regarding the performance of the linking questions, while concordance among known linked records was generally good, the question regarding the number of older siblings only matched in 80 of the 94 measured instances (85% of the time). Although these questions were designed to be static, they are also designed to be as easy to remember as possible. This requires a balance to be struck: It is true that family structure could be more prone to change than a more stable fact (such as the first three letters of one's residential street address in fifth grade), and it is also true that some students may find the number of siblings to be a fact that they remember more readily (or, are more likely to know in the first place). Despite

Table 5
True Positive and True Negative Rates for Linking Algorithm Over Different Subsets of Linking Questions in Simulated Data Set

Linking question Subset	% [95% CI]	
	True positive rate	True negative rate
All seven questions	69.3 [68.2, 70.3]	100.0 [99.9, 100.0]
All seven questions + All subsets of six questions	93.7 [93.1, 94.4]	99.9 [99.8, 99.9]
All seven questions + All subsets of six questions + Subset without siblings/eye color	93.7 [93.1, 94.4]	99.9 [99.8, 99.9]
All seven questions + All subsets of six questions + Subset without eye color/street address	94.8 [94.2, 95.3]	99.7 [99.6, 99.8]
All seven questions + All subsets of six questions + Subset without siblings/eye color/street	96.3 [95.8, 96.8]	92.8 [92.3, 93.3]

Note. CI = confidence interval.

Table 6
Linked Sample Demographic Characteristics

Demographic characteristic	3-year linked sample 2020–2022 (N = 563)		2-year linked sample 2021–2022 (N = 1,545)	
	N	%	*N*	%
Age	10 years: 1	0.0	10 years: 1	0.1
	11 years: 86	15.3	11 years: 128	8.3
	12 years: 122	21.7	12 years: 199	12.9
	13 years: 113	20.1	13 years: 206	13.3
	14 years: 109	19.4	14 years: 240	22.0
	15 years: 115	20.4	15 years: 238	21.2
	16 years: 0	0.0	16 years: 297	19.2
	17 years: 17	3.0	17 years: 46	3.0
Grade	Sixth: 106	18.8	Sixth: 165	10.7
	Seventh: 127	22.6	Seventh: 202	13.1
	Eighth: 102	18.1	Eighth: 210	3.6
	Ninth: 131	23.3	Ninth: 356	23.0
	10th: 97	17.2	10th: 331	21.4
	11th: 0	0.0	11th: 281	18.2
Gender	Boy/man/male: 231	41.0	Boy/man/male: 676	43.8
	Girl/woman/female: 327	58.1	Girl/woman/female: 790	51.1
	Trans boy/man: 1	0.7	Trans boy/man: 10	0.7
	Trans girl/woman: 0	0.0	Trans girl/woman: 1	0.1
	Another gender: 0	0.0	Another gender: 19	1.2
	Don't want to say: 0	0.0	Don't want to say: 4	0.3
	Not sure: 0	0.0	Not sure: 22	1.4
	Missing: 4	*0.2*	*Missing: 23*	*1.5*
Sexual orientation			Straight: 747	77.2
			Bisexual: 90	9.3
			Gay or lesbian: 31	3.2
			Something else: 37	3.8
			Unsure: 45	4.7
			Don't want to say: 15	1.6
			Missing: 3	0.3
			Below ninth grade: 577	
Race	American Indian/Alaskan Native: 8	1.4	American Indian/Alaskan Native: 13	0.8
	Asian: 104	18.5	Asian: 233	15.1
	Haitian/Black/African American: 10	1.8	Haitian/Black/African American: 44	2.9
	Hawaiian/Pacific Islander: 0	0.0	Hawaiian/Pacific Islander: 0	0.0
	Middle Eastern/Northern African: 0	0.0	Middle Eastern/Northern African: 9	0.6
	Multiple: 0	0.0	Multiple: 111	7.2
	Other: 13	2.3	Other: 26	1.7
	White: 392	69.6	White: 1,099	71.1
	Missing: 36	*6.4*	*Missing: 10*	*0.7*
Ethnicity	Hispanic/Latino/a: 21	3.7	Hispanic/Latino/a: 103	6.7
	Not Hispanic/Latino/a: 541	96.1	Not Hispanic/Latino/a: 1,430	92.6
	Missing: 1	*0.2*	*Missing: 12*	*0.8*

Note. Only demographic characteristics for which baseline data are available are included in the table above. Sexual orientation only assessed in students in or above ninth grade in 2021. Percentages reflect only students who were assessed, which may not sum to 100 due to rounding.

variations in response consistency, the linking algorithm performed relatively well, with few false positives, which may speak to the value of having many linking questions that range in difficulty and content.

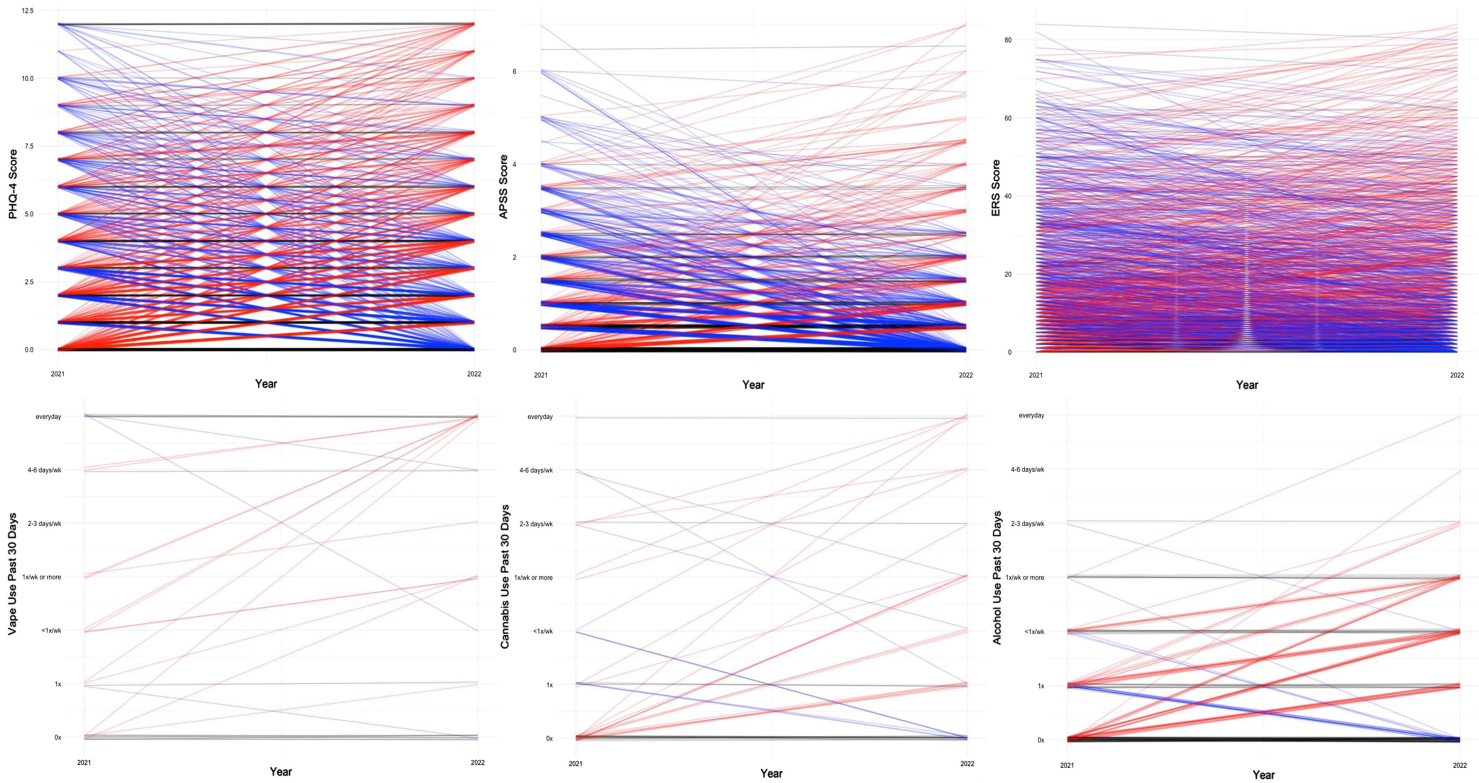
Logistically, this linking method offers confidence beyond other forms of linking (i.e., relying on participant recall of an ID) and the ability to link records without collecting private identifiers. This minimizes the risk of individual survey identification, allowing for a simpler consent process and increased honesty in responding and mitigating risk to participant confidentiality (Lothen-Kline et al., 2003). When linking requires explicit identifiers, participants must consent to the associated risk. For minors, this usually means requiring direct consent from a guardian, limiting adolescents' autonomy and privacy (Faden et al., 2013). Previous studies noted that requiring explicit parental consent resulted in over 90% white samples, suggesting an additional impact on equity and generalizability (Schleider et al., 2022). The linking method provided by this survey minimizes risks to privacy at every step (including collection, storage, analysis, and sharing), and schools were able to administer the survey on their own without an on-site study team.

Lessons for Future Survey Administration

Designing and implementing this longitudinal, linked survey protocol has offered insights into some important considerations for future implementation of this and other similar surveys. First, we

Figure 2

Within-Person Trajectory From 2021 to 2022 on the PHQ-4, APSS, ERS, and Reported Days of Use of Alcohol, Cannabis, and Vaped Nicotine Out of the Past 30 Days



Note. Red lines indicate increases in the corresponding construct, blue indicates decrease, and black indicates stability. Slight jitter (±.05) and transparency have been included for visual ease. PHQ-4 = Patient Health Questionnaire–Four Item Version; APSS = Adolescent Psychotic-Like Symptom Screener; ERS = Emotional Reactivity Scale. See the online article for the color version of this figure.

highlight the importance of prioritizing collaborative, purposeful, and communicative relationships between the research team and school partners. To build trust and communication, study staff may invest time presenting at school board and faculty meetings to increase understanding and buy-in. After data collection, it is critical to follow up with data reports and supporting interpretation, dissemination, and action planning in response to students' needs. In the present study, we took the approach of providing schools with personalized reports on their school's survey results as well as aggregated state-wide data and worked with schools in deciding responsive steps for rehabilitation and student support where the reports indicate student needs. School-wide administration of surveys across the state would not be possible without school partners. Neglecting these relationships and researcher's responsibility to support communities involved in research can quickly undermine the success of what is otherwise procedurally simple. Trust and collaboration are particularly important in communities that have historically been overlooked and harmed by research practices throughout the entirety of the research process. Establishing a longitudinal, trusting, and consistent presence in the school and among students may facilitate honest responding by students that increases over time, as the study staff remains true to their word that privacy and confidentiality will be maintained. School cultural shifts in response to student needs may also support students' interest in engaging with surveys, if they believe that the study staff and school care to take input and enact change accordingly. Study staff should prioritize collaboration with school partners (e.g., by meeting with faculty to increase understanding). It is also key to support schools' follow-up by providing data reports and scaffolding interpretation, dissemination, and action planning in response to students' needs. To facilitate translatable research efforts, it is key that research actively partner with communities to facilitate a dialogue about the research process and outcomes. The relationship-building strategies described here successfully supported interdisciplinary collaboration among a team of hospital-based clinical psychologists and staff with the school-based teams of teachers, administrators, nurses, counselors, students, and other involved staff members.

Additional lessons regard logistics surrounding the consent opt-out procedure. In 2022, opt-outs were performed via a REDCap form (rather than email to study staff). In general, providing a secure, online format for parents to enter opt-out information directly reduces privacy concerns, opportunities for error in tracking opt-outs, and burden on study personnel.

Another set of lessons consider survey content. First, selecting widely used survey measures can facilitate comparisons across studies and cohorts. Additionally, iterative survey administration allows for researchers to make responsive changes to survey content (e.g., assessing reactions to COVID-19 in 2020). Research and school teams must collaboratively make decisions about the core constructs and assessment tools used to construct a thorough but not burdensome survey that can be completed within the school day. With regard to more specific wording: When looking to interpret the meaning of a nonresponse, it is necessary to be able to distinguish between a "passive no" (skipping a question) from an "active no" (when the answer to the question is truly no). Relatedly, survey administrators must make decisions regarding thresholds for tolerating incompleteness in records. As this survey grew longer and incorporated branching logic, it became clear that students with higher rates of substance use were more likely to be noncompleters because they were presented with more follow-up questions due to survey branching logic. Additionally, beginning in 2022, SURF now tracks degree of completeness (0%–100%), so that completeness can be considered differently based on the needs of data analysts.

Limitations

Limitations of this study warrant consideration and are closely linked to some of the challenges presented by this approach. For example, when records are intentionally not identified, it is not possible to follow up with individuals whose data may be missing, incongruent, or nonsensical. As with any self-report survey, it is possible that students incorrectly report their mental health symptoms or substance use. To aid in interpretation, future iterations of this survey will include an item asking students to report on whether they were honest in the survey. It is our intent to promote adolescent research autonomy, which includes empowering them with the capacity to decide for themselves whether they would like to participate in the study *and* whether they would like to honestly respond to survey questions. Of course, additionally, incorrect responses early in the branching logic, whether intentional or not, could cascade forward and limit the ability to gather any follow-up information on branched questions. Additionally, records are not linked if a student moves school districts, potentially limiting representativeness of the linked sample if, for example, residential mobility is higher for families from minoritized or marginalized racial or ethnic backgrounds (Rumberger, 2003) and with greater financial strain (Clark, 2018). The inability to track moves outside of one school district could theoretically contribute to systematic exclusion of students who hold a minoritized racial or ethnic identity and for students from a lower socioeconomic status background. Although survey weights can address *some* of this disparity in representation, they are limited by the restricted number of students in some racial identity categories (e.g., some districts had no students from a racial group complete the survey, even though they report that students from that racial group are present in the district).

Regarding the survey weights, specifically, it is only possible to post-stratify the sample based on variables that could be matched to population-level DESE data. Although these variables, grade and race, are important for many analyses, they are not the only variables that may be relevant. Therefore, although the post-stratification weights match the sample to known population totals on certain characteristics, they will not for others. Moreover, note that post-stratification is based on student reports of gender and grade, and it is possible that some students did not answer these items in a way that reflects the school's understanding of their race and grade. In future iterations, administrative data (which were not available for this study) could be used to address potential differences in student categorization.

Future work could also investigate whether the false-positive rate of links is significantly improved when restricting by district (vs. not) and evaluate the point at which allowing all possible linked candidates to be considered is worthwhile and when it becomes computationally overwhelming. Additionally, to perform school-level analyses that properly acknowledge the nesting of the data, larger samples are needed: At present, the samples at individual schools are still quite variable (the 2021–2022 sample is divided between 11 schools with 20–307 records per school).

Future Directions

Survey administration is continuing and expanding to schools throughout Massachusetts. This will eventually allow for analyses that account for individual, school, and sample-wide trajectories of pathology, the application of survey population weights in analyses, and the extension of linking across more years. These trajectory analyses may be used to identify risk markers for developing pathology, as well as developmental points at which prevention and intervention efforts may be most fruitful. Relatedly, future iterations of this survey will incorporate an option for students to indicate interest in follow-up school-based mental health referral. In this way, research can function to support students while also characterizing developing mental health and substance use concerns. It will be key to replicate the linking procedure described here and evaluate the extent to which substance use and mental health may relate to students' likelihood of being linked across time. With a continuously growing sample size, incoming data can be used to identify risk factors for substance use and disconnection from survey efforts, describe comorbidities among mental health and substance use, and clarify the role of contextual and identity factors in these processes.

Conclusion

To address the ongoing youth mental health crisis in the United States, it is key to better characterize how symptoms develop across time. Integrating linking code protocols with widespread screeners offers a pivotal opportunity to bolster our understanding of prevention and intervention needs. Confidentiality afforded by linking codes may facilitate community-level research while protecting adolescents' privacy, a key ingredient to their willing and honest participation. Approaches such as the one presented here may ultimately bolster opportunities for community-collaborative research, in the name of understanding symptom progression for a range of difficult-to-reach populations.

References

Ahn-Horst, R. Y., & Bourgeois, F. T. (2024). Mental health-related outpatient visits among adolescents and young adults, 2006–2019. *JAMA Network Open*, *7*(3), Article e241468. https://doi.org/10.1001/jamanetworkopen.2024.1468

Benton, T. D., Boyd, R. C., & Njoroge, W. F. M. (2021). Addressing the global crisis of child and adolescent mental health. *JAMA Pediatrics*, *175*(11), 1108–1110. https://doi.org/10.1001/jamapediatrics.2021.2479

Blanco, C., Hoertel, N., Franco, S., Olfson, M., He, J. P., López, S., González-Pinto, A., Limosin, F., & Merikangas, K. R. (2017). Generalizability of clinical trial results for adolescent major depressive disorder. *Pediatrics*, *140*(6), Article e20161701. https://doi.org/10.1542/peds.2016-1701

Clark, A. (2018). The role of residential mobility in reproducing socioeconomic stratification during the transition to adulthood. *Demographic Research*, *38*, 169–196. https://doi.org/10.4054/DemRes.2018.38.7

Fabina, J., Hernandez, E. L., & McElrath, K. (2023). *School enrollment in the United States: 2021. American Community Survey reports*. United States Census Bureau. https://www.census.gov/content/dam/Census/library/publications/2023/acs/acs-55.pdf

Faden, R., Kass, N., Whicher, D., Stewart, W., & Tunis, S. (2013). Ethics and informed consent for comparative effectiveness research with prospective electronic clinical data. *Medical Care*, *51*(8, Suppl. 3), S53–S57. https://doi.org/10.1097/MLR.0b013e31829b1e4b

Fisher, C. B., Wallace, S. A., & Fenton, R. E. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence*, *29*(6), 679–695. https://doi.org/10.1023/A:1026455906512

Hedden, S., Gfroerer, J., Barker, P., Smith, S., Pemberton, M. R., Saavedra, L. M., Forman-Hoffman, V. L., Ringeisen, H., & Novak, S. P. (2012). Comparison of NSDUH mental health data and methods with other data sources. In *CBHSQ data review* (pp. 1–19) Substance Abuse and Mental Health Services Administration.

Hyland, A., Ambrose, B. K., Conway, K. P., Borek, N., Lambert, E., Carusi, C., Taylor, K., Crosse, S., Fong, G. T., Cummings, K. M., Abrams, D., Pierce, J. P., Sargent, J., Messer, K., Bansal-Travers, M., Niaura, R., Vallone, D., Hammond, D., Hilmi, N., … Compton, W. M. (2017). Design and methods of the Population Assessment of Tobacco and Health (PATH) study. *Tobacco Control*, *26*(4), 371–378. https://doi.org/10.1136/tobaccocontrol-2016-052934

Kelleher, I., Harley, M., Murtagh, A., & Cannon, M. (2011). Are screening instruments valid for psychotic-like experiences? A validation study of screening questions for psychotic-like experiences using in-depth clinical interview. *Schizophrenia Bulletin*, *37*(2), 362–369. https://doi.org/10.1093/schbul/sbp057

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics*, *50*(6), 613–621. https://doi.org/10.1176/appi.psy.50.6.613

Lawson, L. P., Richdale, A. L., Haschek, A., Flower, R. L., Vartuli, J., Arnold, S. R., & Trollor, J. N. (2020). Cross-sectional and longitudinal predictors of quality of life in autistic individuals from adolescence to adulthood: The role of mental health and sleep quality. *Autism*, *24*(4), 954–967. https://doi.org/10.1177/1362361320908107

Lothen-Kline, C., Howard, D. E., Hamburger, E. K., Worrell, K. D., & Boekeloo, B. O. (2003). Truth and consequences: Ethics, confidentiality, and disclosure in adolescent longitudinal prevention research. *The Journal of Adolescent Health*, *33*(5), 385–394. https://doi.org/10.1016/S1054-139X(03)00184-8

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A., & Brähler, E. (2010). A four-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire–4 (PHQ-4) in the general population. *Journal of Affective Disorders*, *122*(1–2), 86–95. https://doi.org/10.1016/j.jad.2009.06.019

Miech, R. A., Johnston, L. D., Patrick, M. E., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2023). *Monitoring the future national survey results on drug use, 1975–2022: Secondary school students*. Institute for Social Research.

Mpofu, J. J., Underwood, J. M., Thornton, J. E., Brener, N. D., Rico, A., Kilmer, G., Harris, W. A., Leon-Nguyen, M., Chyen, D., Lim, C., Mbaka, C. K., Smith-Grant, J., Whittle, L., Jones, S. E., Krause, K. H., Li, J., Shanklin, S. L., McKinnon, I., Arrey, L., … Roberts, A. M. (2023). Overview and methods for the youth risk behavior surveillance system—United States, 2021. *MMWR Supplements*, *72*(1), 1–12. https://doi.org/10.15585/mmwr.su7201a1

Nock, M. K., Wedig, M. M., Holmberg, E. B., & Hooley, J. M. (2008). The emotion reactivity scale: Development, evaluation, and relation to self-injurious thoughts and behaviors. *Behavior Therapy*, *39*(2), 107–116. https://doi.org/10.1016/j.beth.2007.05.005

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic metaanalysis. *General Hospital Psychiatry*, *39*, 24–31. https://doi.org/10.1016/j.genhosppsych.2015.11.005

Qian, W., Battista, K., Bredin, C., Brown, K. S., & Leatherdale, S. T. (2015). *Assessing longitudinal data linkage results in the COMPASS study: Technical report series*. University of Waterloo. https://uwaterloo.ca/compass-system/publications#technical

Racine, N., McArthur, B. A., Cooke, J. E., Eirich, R., Zhu, J., & Madigan, S. (2021). Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: A meta-analysis. *JAMA*

Pediatrics, *175*(11), 1142–1150. https://doi.org/10.1001/jamapediatrics.2021.2482

Ripper, L., Ciaravino, S., Jones, K., Jaime, M. C. D., & Miller, E. (2017). Use of a respondent-generated personal code for matching anonymous adolescent surveys in longitudinal studies. *The Journal of Adolescent Health*, *60*(6), 751–753. https://doi.org/10.1016/j.jadohealth.2017.01.003

Rumberger, R. W. (2003). The causes and consequences of student mobility. *The Journal of Negro Education*, *72*(1), 6–21. https://doi.org/10.2307/3211287

Schleider, J. L., Mullarkey, M. C., Fox, K. R., Dobias, M. L., Shroff, A., Hart, E. A., & Roulston, C. A. (2022). A randomized trial of online single-session interventions for adolescent depression during COVID-19. *Nature Human Behaviour*, *6*(2), 258–268. https://doi.org/10.1038/s41562-021-01235-0

Schneider, S. K., O'Donnell, L., Stueve, A., & Coulter, R. W. (2012). Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health*, *102*(1), 171–177. https://doi.org/10.2105/AJPH.2011.300308

Schnell, R., Bachteler, T., & Reiher, J. (2010). Improving the use of self-generated identification codes. *Evaluation Review*, *34*(5), 391–418. https://doi.org/10.1177/0193841X10387576

Shek, D. T., & Ng, C. S. (2016). Longitudinal research design in adolescent developmental research. *International Journal on Disability and Human Development*, *15*(4), 349–358. https://doi.org/10.1515/ijdhd-2017-5002

Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E., & Binau, S. G. (2019). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology*, *128*(3), 185–199. https://doi.org/10.1037/abn0000410

Uhlhaas, P. J., Davey, C. G., Mehta, U. M., Shah, J., Torous, J., Allen, N. B., Avenevoli, S., Bella-Awusah, T., Chanen, A., Chen, E. Y. H., Correll, C. U., Do, K. Q., Fisher, H. L., Frangou, S., Hickie, I. B., Keshavan, M. S., Konrad, K., Lee, F. S., Liu, C. H., … Wood, S. J. (2023). Towards a youth mental health paradigm: A perspective and roadmap. *Molecular Psychiatry*, *28*(8), 3171–3181. https://doi.org/10.1038/s41380-023-02202-z

Vacek, J., Vonkova, H., & Gabrhelík, R. (2017). A successful strategy for linking anonymous data from students' and parents' questionnaires using self-generated identification codes. *Prevention Science*, *18*(4), 450–458. https://doi.org/10.1007/s11121-017-0772-6

Received June 7, 2024
Revision received November 7, 2024
Accepted November 18, 2024 ■