



Methodological Strategies for Prospective Harmonization of Studies: Application to 10 Distinct Outcomes Studies of Preventive Interventions Targeting Opioid Misuse

Ty A. Ridenour⁷ · Gracelyn Cruden⁵ · Yang Yang⁹ · Erin E. Bonar¹⁰ · Anthony Rodriguez⁶ · Lissette M. Saavedra⁷ · Andrea M. Hussong¹¹ · Maureen A. Walton¹⁰ · Bethany Deeds³ · Jodi L. Ford⁴ · Danica K. Knight⁹ · Kevin P. Haggerty⁸ · Elizabeth Stormshak¹² · Terrence K. Kominsky¹ · Kym R. Ahrens⁸ · Diana Woodward² · Xin Feng⁴ · Lynn E. Fiellin¹³ · Timothy E. Wilens² · David J. Klein⁶ · Claudia-Santi Fernandes¹³

Accepted: 18 July 2022
© Society for Prevention Research 2022

Abstract

The Helping to End Addiction Long-Term (HEAL) Prevention Cooperative (HPC) is rapidly developing 10 distinct evidence-based interventions for implementation in a variety of settings to prevent opioid misuse and opioid use disorder. One HPC objective is to compare intervention impacts on opioid misuse initiation, escalation, severity, and disorder and identify whether any HPC interventions are more effective than others for types of individuals. It provides a rare opportunity to prospectively harmonize measures across distinct outcomes studies. This paper describes the needs, opportunities, strategies, and processes that were used to harmonize HPC data. They are illustrated with a strategy to measure opioid use that spans the spectrum of opioid use experiences (termed involvement) and is composed of common “anchor items” ranging from initiation to symptoms of opioid use disorder. The limitations and opportunities anticipated from this approach to data harmonization are reviewed. Lastly, implications for future research cooperatives and the broader HEAL data ecosystem are discussed.

Keywords Opioid misuse · Opioid use disorder · Prevention · Harmonization · Integrative data analysis · Adolescents · Young adults

The ongoing Helping to End Addiction Long-Term (HEAL) Prevention Cooperative (HPC) is funded by the National Institutes of Health, administered by the National Institute on Drug Abuse, and created to rapidly develop a diverse set of evidence-based interventions for implementation in a

variety of settings to prevent opioid misuse. Each of the 10 HPC research projects has distinct characteristics, including intervention strategies, location along the Centers for Disease Control and Prevention’s (CDC’s) Continuum of Care, underlying theoretical models, settings, targeted age

✉ Ty A. Ridenour
tridenour@rti.org

¹ Emory University and the Cherokee Nation, Atlanta, USA

² Massachusetts General Hospital, Boston, USA

³ National Institute On Drug Abuse, North Bethesda, USA

⁴ The Ohio State University, Columbus, USA

⁵ Oregon Social Learning Center, Eugene, USA

⁶ RAND Corporation, & University of California, Los Angeles, USA

⁷ RTI International, 3040 E. Cornwallis Rd., PO Box 12194, 326 Cox Bldg, Research Triangle Park, 27709-2194 NC, USA

⁸ Seattle Children’s Hospital & University of Washington, Seattle, USA

⁹ Texas Christian University, Fort Worth, USA

¹⁰ University of Michigan, Ann Arbor, USA

¹¹ University of North Carolina, Chapel Hill, USA

¹² University of Oregon, Eugene, USA

¹³ Yale University, New Haven, USA

groups and populations, level of participants' opioid use or misuse, and even measurement of the common outcome: opioid misuse. The 10 HPC studies included intervention programs or strategies adapted from evidence-based interventions that had been developed and/or researched by their respective study teams. One HPC objective is to generate evidence about how the interventions compare regarding impacts on opioid misuse initiation, escalation, severity, and disorder. A second HPC objective is to identify whether any HPC interventions are more effective than others for some individuals. Summarizing and interpreting evidence that is accumulated across distinct HPC research projects exemplifies challenges that are faced in prevention and social sciences more broadly. However, unlike retrospective efforts to aggregate evidence across studies, HPC offered an opportunity to prospectively harmonize across studies to maximize cross-study comparability and investigate research questions that the individual studies cannot address alone.¹

This paper describes the opportunities, challenges, strategies, and process involved with harmonizing the HPC data. First is a brief overview of the scientific utility of harmonizing varied features among HPC research projects with the same focus (e.g., an outcome of opioid misuse). Next are brief comparisons among the HPC research projects followed by our strategies to harmonize key measures to support cross-intervention investigations. Finally, planned research questions illustrate how harmonization will be used to leverage the anticipated HPC opportunities. Implications of these harmonization strategies for future research cooperatives will be considered. For brevity, we refer to each research project by its primary grantee institution. We also acknowledge that this research would not be possible without the essential community partners who collaborated throughout each research project; their crucial roles are described in greater detail in Graham et al. (this issue).

Needs for and Opportunities That Can Arise from *Harmonizing* Methods Among Prevention Studies

Research has revealed genetic, neurological, individual, and contextual risk and protective factors associated with patterns and progression of substance use and consequent disorder (Stanis & Andersen, 2014). These findings have informed prevention interventions to reduce the risk of substance experimentation and progression, resulting in a variety of evidence-based programs that are most effective when matched to their target populations' level of risk (LeNoue &

Riggs, 2016; SAMHSA, 2019). Given the heterogeneity of prevention interventions (e.g., scope, content, target populations), a critical component of evaluating their effectiveness involves consideration of target sample characteristics, level of intervention intensity, and the context of intervention implementation (Brown et al., 2018).

Extant literature on these topics has critical limitations (Tanner-Smith & Lipsey, 2015). First, individual prevention studies tend to report findings of a selection of variables derived from regionally distributed participants, limiting generalizability. Second, studies usually do not have sufficient statistical power to explore potential mechanisms underlying intervention effectiveness or possible moderators of intervention outcomes. Third, meta-analyses and reviews are often limited by publication biases and selective outcomes (Chan et al., 2004). Additional data analysis (e.g., moderation) is also bounded by the availability of data and inferential equivalence of variables (e.g., variables are operationalized distinctively or cannot be meaningfully harmonized) (Fortier et al., 2011b).

Increasingly, researchers recognize the need for harmonization to address these limitations. Yet, its use in prevention studies remains uncommon (Brown et al., 2018). Harmonization is a systematic approach, which if done prospectively, involves development of and adherence to study protocols (e.g., common variables, standardized timeline for data collection) and data management plans (e.g., standardized data forms) across studies to allow for advanced statistical methods that can greatly expand the research questions that can be tested. Harmonization allows for more efficient building of the evidence base while capitalizing on between-study heterogeneity. Harmonization can be facilitated through alignment of study designs (e.g., comparison groups, target outcomes) or evaluation approaches (e.g., measures, time-points) and often involves utilizing the "lowest common denominators" in study designs such as a shared outcome measurement frequency (e.g., baseline, 3 months).

Needs for and Opportunities That Can Arise from *Varying* Methods Among Prevention Studies

Despite these benefits to harmonization across studies, expanding evidence across the prevention continuum requires variability in how studies are designed and intervention strategies are utilized, targeted populations and settings, and testing impacts on proximal and distal outcomes (Williams, 2016). Thus, harmonizing evaluation is often not straightforward. For example, important diversity across studies exists with respect to key population characteristics such as age, cultural heritage, and economic background. Harmonization of such factors is not only

¹ HPC data collection is ongoing and data analyses have not yet occurred.

unnecessary or impractical, but antithetical to principles of scientific equity necessitating that diverse populations be engaged in prevention research to avoid the creation or exacerbation of health disparities (e.g., through the study of only certain privileged groups) (Perrino et al., 2015). Although study population characteristics and other potential moderators such as mental health often cannot be harmonized, measures of these moderators can be (Williams, 2016). Importantly, as is discussed later, measures of proximal (i.e., mediators) and distal targets can also be partially harmonized to accommodate between-study differences. Harmonization with respect to measurement can include requiring only some items of a measure to be identical among studies, statistically placing distinct items onto the same scale, and slight altering of an item's content across studies to align with the population being assessed (e.g., "Your school" vs. "Your hospital") (Siddique et al., 2018). However, even seemingly minor measurement differences such as self-report versus parent report of the same construct (e.g., youth depression) can result in meaningful statistical and measurement differences among studies (Makol et al., 2019). By balancing study-specific measurement tools while standardizing measurement elements across studies, the integrity to necessary study diversity can be maintained while still harmonizing studies.

Available Methodologies to Aggregate Data Across Distinct Prevention Studies

Historically, the primary approaches to synthesize knowledge were literature reviews and systematic reviews. In the last 15 years, researchers developed and refined methods and analytic techniques to conduct meta-analyses, which aggregate summative data across multiple studies (e.g., treatment effect sizes from clinical trials) to more accurately quantify a population-level effect (e.g., Curran & Hussong, 2009; Rothstein et al., 2005). A more recent advance is Integrative Data Analysis (IDA), which allows far greater sophistication in accumulation of knowledge across datasets. Similar to meta-analysis, IDA is a framework to simultaneously analyze data from multiple studies (Hussong et al., 2013). IDA pools individual-level raw data from studies and allows for harmonizing across distinct measures to overcome several barriers that limited researchers from integrating data in the past. Specifically, IDA entails advanced scale scoring methods (e.g., Moderated Non-linear Factor Analysis [MNLFA], Item Response Theory [IRT]) to create comparability in measures across studies that may be assessing the same construct with some variation in item content (Bauer & Hussong, 2009; Saavedra et al., 2021). IDA offers clear advantages over single studies for HPC investigations involving multiple research projects including pooling larger

sample sizes, increasing statistical power, and greater subgroup coverage to enable analyses of low-base rate behaviors and subpopulations.

There are at least two main types of IDA studies. Retrospective IDA integrates data that were collected without consideration for potential IDA (e.g., Brown et al., 2018). Prospective IDA involves determination of common research design features and measures prior to data collection with the explicit intent of pooling raw individual participant data from ongoing and new studies, including randomized trials. Prospective IDA allows researchers to update and monitor pooled participant data sets as data are collected. Researchers can collaborate to delineate measures of key constructs to use across their individual studies to ultimately produce a pool of common and unique items that are supported by a clear set of hypotheses-driven guidelines (see Hussong et al., 2013).

Diversity Among HEAL Prevention Cooperative Studies

The 10 HPC research projects are characterized far more by variability than commonality. The funder required each research project to develop and to test a program to (1) prevent initiation of opioid misuse and opioid use disorder (OUD), (2) target recipients who are at risk of opioid misuse and range in age from 16 to 30, and (3) be delivered via an existing system where at-risk adolescents and young adults can be engaged such as health care, justice, child welfare, or school (NIH, 2018). The funded HPC research project programs target a variety of intervention recipients (e.g., individuals experiencing high risk, their parents, their communities) and employ varied study designs, timelines, and measures (Table 1). In fact, the only methodological or theoretical feature common to all 10 studies was their intent to decrease or prevent opioid misuse. Yet, even their originally proposed measures of opioid misuse outcomes varied considerably, as did their targeted levels of opioid misuse (e.g., initiation of misuse, onset of a disorder).

Given the diversity of HPC research project features, there will be limited opportunities to use traditional cross-study aggregation techniques to make quantitative comparisons among the studies. For example, meta-analysis to aggregate outcome effect sizes across studies cannot be used because none of the HPC studies are testing the same treatment. To illustrate, the University of Michigan study engages adolescent and young adult emergency department patients with varying eligibility criteria: (1) past 12-month misuse of prescription or illicit opioids or (2) past 12-month use of prescription opioids in combination with another risk factor for opioid misuse (e.g., other substance use, depression, suicidality) (Bonar et al., 2021). The behavioral

Table 1 Features of the HEAL Prevention Cooperative Intervention Outcome Studies

Grantee Institution	Intervention(s)	Target Population and Age Range	Setting(s)	Study design	Follow-up timepoints	Targeted level of opioid use	Theorized mechanism(s) of change	Opioid misuse measure(s)*
Emory University and Cherokee Nation	Integrated multilevel school, family, and community intervention	Rural American Indian and other youth living in the Cherokee Nation, aged 15–17 years at baseline to 18–20 years at follow-up	20 small rural towns and high schools in the 14 counties that partially or fully fall within the Cherokee Nation in northeastern Oklahoma	Cluster randomized trial with multilevel intervention vs. delayed-intervention	6, 12, 18, 24, 30, and 36 months	Initiation and escalation	Multilevel risk and protective factors targeting demand and supply	Involvement with legally and illegally manufactured opioids core items
Massachusetts General Hospital	N/A	Patients aged 16–30	Behavioral health and substance use disorder treatment centers in Massachusetts	Naturalistic, longitudinal	Every 6 months for duration of study	Misuse and disorder	Improved mental health and substance use	SAGE-SR, BAM
Ohio State University	Housing First, opioid and related risk prevention services (i.e., strengths-based outreach and advocacy, HIV prevention, and motivational interviewing)	Youth, aged 18–24, who meet the criteria for homelessness	Community-based, including drop-in centers, the streets, and other locations where youth are found	2-arm randomized trial of housing + preventive services vs. preventive services alone	3, 6, 9, and 12 months	Frequency of misuse, time until disorder	Accessing social resources, enhanced self-efficacy and resilience, improved stress response system	SCID-5 for disorder; total number of days, in 90 days prior to last use, age at first use, lifetime weeks of use
Oregon Social Learning Center	Families Actively Improving Relationships for Prevention (PRE-FAIR)	Parents aged 16–30	Outpatient community health and substance use clinics in Oregon serving parents involved with child welfare and self-sufficiency	Hybrid Type I, randomized controlled trial of PRE-FAIR vs. standard services	4, 6, 12, 18, and 24 months; monthly assessments during months 1–18	Initiation and escalation	Engagement; Improved relationships with their children, family, and collateral service providers	ASI: 30-day use, lifetime use, and OUD; urinalysis
RAND/UCLA	Traditions and Connections for Urban Native Americans (TACUNA)	Native American, emerging adults, ages 18–25	Nationwide US urban residents with internet connectivity	2-arm randomized controlled trial of 3 workshops + wellness gathering vs. one workshop	3, 6, and 12 months		Native culture attachment, less risky social network	MTF: use in lifetime, past 30 days, past 3 months

Table 1 (continued)

Grantee Institution	Intervention(s)	Target Population and Age Range	Setting(s)	Study design	Follow-up timepoints	Targeted level of opioid use	Theorized mechanism(s) of change	Opioid misuse measure(s)*
Seattle Children's Hospital	3 different-intensity interventions based on a combination of Adolescent Community Reinforcement Approach, Assertive Continuing Care, Trauma Affect Regulation, Guide for Education and Therapy, Motivational Interviewing	Youth, aged 15–25, close to reentry from justice residential facilities without moderate or severe opioid use disorder	All Washington State supported Juvenile Rehabilitation Facilities funded through Department of Children, Youth, and Families	Sequential, Multiple Assessment Randomized Trial (SMART)	3 and 6 months (substance use and related outcomes); 12 months (recidivism)	Initiation, frequency of use, misuse	Relationships, skill building, and connection to resources to make non-use more rewarding than use	CRAFT
Texas Christian University	Trust Based Relational Intervention to enhance youths' relationship with caregivers, empowering caregivers to identify and address youths' physiological and emotional needs, and attachment with a safe adult	Youth, aged 15–18, close to being released from post-adjudication facilities	Juvenile justice reentry programs of Illinois and Texas	Effectiveness/Implementation, Hybrid Type I	3, 6, 12, and 18 months	Initiation and escalation	Supportive, responsive adult; youth emotion self-regulation, youth-caregiver relationships	TCU Drug Screen, PhenX items for use, TLFB, Urinalysis
University of Michigan	Health coach session or portal-based messaging	Patients, ages 16–30 with past-year opioid use + at least one other risk factor or past-year opioid misuse	Emergency department	Factorial	3, 6, and 12 months	Misuse	Self-efficacy	ASSIST for severity of opioid misuse, NSDUH for screening past-year use, past 30-day number of days based on ASI
University of Oregon	Family Check-Up Online	Young adult parents with histories of substance misuse and children aged 18 months to 5 years	Rural Oregon early childhood agencies	Randomized Controlled Trial	3, 6, and 12 months	Frequency of misuse	Parenting skills, parent wellness, healthy behavioral routines	PhenX use and misuse frequency, administrative records for disorder

Table 1 (continued)

Grantee Institution	Intervention(s)	Target Population and Age Range	Setting(s)	Study design	Follow-up timepoints	Targeted level of opioid use	Theorized mechanism(s) of change	Opioid misuse measure(s)*
Yale University	Videogame	Students, ages 16–19, with 30-day non-opioid substance use and risk for anxiety or depression	School-based health centers	Randomized Controlled Trial	Week 6, 3, 6, and 12 months	Initiation	Increase perceived risk of harm	MTF lifetime opioid use, 30-day opioid use

*All projects include the Involvement with Legally Manufactured Opioids measure (Table 2) and a corresponding measure of involvement with illegal opioids. *ASI*, Addiction Severity Index; *BAM*, Brief Addiction Monitor; *CTC*, Communities that Care Youth Survey; *DAST*, Drug Abuse Screen Test; *Form 90*, Form 90 Substance Use Interview; *K-SADS*, Schedule for Affective Disorders and Schizophrenia for School-Aged Children; *MTF*, Monitoring the Future; *NSDUH*, National Survey on Drug Use and Health; *PhenX*, consensus measures for Phenotypes and eXposures; *SCID-5*, Structured Clinical Interview for DSM-5; *TLFB*, Timeline Follow Back

interventions being tested are delivered virtually and tailored to each participant's opioid use or misuse history or risk factors. In contrast, the Emory University study of rural high school sophomores tests a community-level intervention that includes (1) community-level campaigns and programs to raise awareness of the dangers of opioid use, decrease access to opioids and other drugs, and strengthen protective factors and opportunities that are incompatible with opioid use and (2) in-school screening, brief intervention, and referral to treatment for substance use.

Certain commonalities among subsets of HPC research projects, or subsamples within them, can be leveraged by aggregating data from only those subsamples. For example, perhaps a sufficient number of female students have depression or anxiety in the Emory University study who could be compared to similar participants in the Yale University evaluation of outcomes from a videogame to prevent the initiation of opioid misuse in at-risk high school students (based on at least mild symptoms of depression or anxiety and prior use of cigarettes, e-cigarettes, or illicit drugs other than opioids). The videogame delivers relatable storylines and skill-building activities with the purpose of increasing students' perceived risk of harm from opioids, reducing their intentions to misuse opioids, enhancing refusal skills and mental health coping strategies, and normalizing help-seeking. Perhaps the two most comparable HPC studies are testing different interventions that were both designed for youths or young adults transitioning from juvenile justice residential settings to reentry into their home communities (Texas Christian University and Seattle Children's Hospital). Additionally, two studies of parenting interventions are delivered to young parents who reside in proximal areas of rural counties (Oregon Social Learning Center and University of Oregon).

However, the remaining studies are unique in terms of their settings or intervention recipients, including two health care-based studies designed for emergency departments (University of Michigan) or mental health and substance use disorder clinics (Massachusetts General Hospital). Two other interventions are designed for American Indian/Alaska Native populations but otherwise differ in intervention strategies, ages of samples, mechanisms of change, and methodologies (RAND/UCLA, Emory University).

Methods Selected to Facilitate HPC Study Harmonization

Despite these challenges to harmonization, certain commonalities among HPC research project designs will facilitate important cross-study research opportunities. Most studies included 3-, 6-, and 12-month outcomes, while two of the three research teams that originally had different follow-up timelines agreed to instead use 3-, 6-, and 12-month

outcomes to align with other studies. The University of Oregon's study was supported by a supplemental funding mechanism (the others were funded under a UG3/UH3 phased mechanism), which restricted its timeline and precluded long-term follow-ups. Fortunately, the harmonization of follow-ups provides junctures at which the follow-up timelines among these studies will be comparable (Table 1).

HPC Measures Standardization

Because each HPC research project team separately designed its methods to test its own intervention, the setting in which it is delivered, intervention recipients, mechanisms of change, and other unique aspects of each program, the primary HPC strategy to enhance harmonization across studies was to identify measures that would maximize the scientific knowledge garnered within and across studies. Identical measures that were used across HPC research projects were termed "Standardized Measures" (or identical measures, not to be confused with "standardized assessments" that are normed to national populations) (Reynolds et al., 2021) to distinguish this strategy from statistically "Harmonized Measures" (see HPC Harmonization Strategies below). For example, our standardized opioid misuse measures were in part designed to maximize the number of ways in which intervention impacts could be detected. They also were designed to maximize comparability of HPC data to other prominent studies such as the concurrent HEAL Initiatives (NIH, 2021); national surveillance surveys and epidemiology studies (e.g., Monitoring the Future or National Survey of Drug Use and Health); and individual studies such as the Adolescent Brain Cognitive Development study (DHHS, 2021). Ultimately, the HPC identified 26 common measures composed of 94 items, some of which were used in only a subset of HPC studies.

Standardizing measurement of opioid misuse led to development of a new outcome measure to span the full continuum of involvement, from opioid misuse initiation through symptoms of disorder, composed of common items used in prevention and treatment research that stand as single-item measures of outcomes (Table 2). Many existing research and clinical measures are relevant to individual HPC research project settings, but they vary in relevance to the targeted outcomes of other HPC research projects. While aiming for common items across research projects, it was important to also take advantage of additional information that could be culled from the unique instruments used in each research project. Thus, standardized "common items" were identified to serve as anchors to delineate the underlying construct of opioid use involvement onto which study-specific items could further delineate individual differences on a commensurate scale across the HPC studies (Curran et al., 2014). To illustrate this latter strength, on the one end of the spectrum is the Massachusetts General Hospital study

with participants from inpatient psychiatric departments and treatment centers for substance use disorder which entails complete detailed measures of opioid misuse and OUD. In contrast, the Emory University study of high school students focused measurement on opioid misuse initiation and light use in anticipation that OUD would be rare.

Process to Identify Standardized Common Items

To select "common items," consensus agreement among all HPC Principal Investigators was required to ensure that their study design priorities were met or at least not impinged upon, their primary measurement interests were accounted for, and vetting of candidate items was thorough. This required months of at least biweekly meetings with significant compromises made by some projects (such as adding items to a survey protocol that was restricted to 20 min).

Six steps guided our identification of standardized constructs and measures (illustrated with the Involvement with Legally Manufactured Opioids [ILO] measure, Table 2). First, a measurement expert (with 25 years' experience in psychometrics and psychological test development) who was not a member of any HPC research project identified the objectives in each study for measuring opioid consumption (e.g., whether interventions targeted opioid misuse initiation, escalation, or disorder) and the measures that were originally planned. Next, that individual identified convergences and differences among study objectives and measures. Third, one or more specific instruments or items to measure opioid misuse were proposed to a workgroup composed of HPC Principal Investigators and measurement experts. Workgroup members debated the adequacy and utility of the items or tools, at times resulting in a search for alternatives to better meet all studies' objectives. Fourth, members voted to approve an item or tool or determine if further negotiation was needed (with steps three and four sometimes repeated). The next step was to determine if nuances in the wording or presentation of an item were needed for one or more HPC research projects. Only sufficiently minor modifications (e.g., permitting regional-specific slang terms for opioids, or the optional additions to item ILO1 in Table 2) were considered to yield data that would be comparable to the other HPC research projects. Lastly, these study-specific adjustments were approved by consensus vote.

Opioid Misuse Involvement: Misuse of Legally Manufactured Opioids

As mentioned, a key objective for the HPC opioid involvement measures was to quantify the full continuum of opioid misuse experiences. As the HPC workgroup developed measures of opioid misuse, they found the most comparable resource was by Saha and colleagues (2012), which anchored

Table 2 Involvement with legally manufactured opioids**Introduction:**

This next item asks about using prescription opioids for pain relief or treatment (e.g., Vicodin, Norco, Fentanyl, Hydrocodone, Oxycodone, Percocet, Tramadol, Tylenol with Codeine 3 or 4, Dilaudid, Methadone, Buprenorphine or Bupe, Suboxone) in any way a doctor or medical provider did not tell you to use them. This includes:

- Using without a prescription of your own (for example, someone else's medicine)
- Using more or for longer than you were told to take it
- Using for reasons other than pain (such as to get high, to sleep, or for anxiety)

NOTE: Do not report your use of "over-the-counter" pain relievers such as aspirin, Tylenol, Advil, or Aleve.

Note: local terms for legal opioids can be added to the list of examples.

Note: for younger participants who may not know the term "opioids," this term can be replaced with an alternative for the lead-in

Item	Level of involvement	Question	Response options
ILO1	Initiation, amount used	During your life, how many times have you used a prescription opioid without a doctor's prescription or differently than how a doctor or medical provider told you to use it? ^a	0 times, 1 or 2 times, 3–9 times, 10–19 times, 20–39 times, 40–99 times, 100 or more times
ILO1a	Initiation	(Optional lead-in): Have you ever used a prescription opioid without a doctor's prescription or differently than how a doctor or medical provider told you to use it?	Yes, No
ILO1b	Initiation	(Optional follow-up IF ILO1 = 1 + times): Do you have a current prescription for a prescription opioid from a doctor or medical provider?	Yes, No
ILO2	Initiation	(IF ILO1 = Used 1 or more times): How old were you the first time that you used a prescription opioid?	9 or younger, 10 years old, 11 years old, etc., up to the oldest age of study participants
ILO3	Amount used (regular use)	(IF ILO1 = Used 1 or more times -OR- used 3 or more times): Have you ever used a prescription opioid at least once a month for 3 months in a row?	Yes, No
ILO4	Amount used (regular use)	(IF "Yes" to ILO3): How old were you the first time you used a prescription opioid at least once a month for 3 months in a row?	9 or younger, 10 years old, 11 years old, etc., up to the oldest age of study participants
ILO5	Amount used (frequency)	(IF ILO1 = Used 1 or more times): During the past 30 days, how many days did you use a prescription opioid?	0 to 30 days
ILO5a	Amount used (frequency)	(Optional follow-up IF ILO1 = Used 1 or more times): On the days that you used a prescription opioid, how many times did you use the prescription opioid on average?	1 time, 2 times, 3 times, 4 times, 5 times, 6 or more times
ILO6	Problematic or disordered use	(IF ILO3 = Yes): During the past 3 months, have you ever missed school, work, or other obligations [other examples may be added] because of using a prescription opioid? Note: an alternative wording may ask: During the past 3 months, how many times have you missed school, work, or other obligations because of using a prescription opioid? Note: an alternative wording may ask: When was the last time that you missed school, work, or other obligations because of using a prescription opioid?	Yes, No Never, 1 time, 2 times, 3–5 times, 6–9 times, 10–19 times, 20 or more times (Note: additional amounts may be added.) Past month, 2–3 months ago, 4–12 months ago, 1 + years ago, Never (Note: time lags may vary, but must include a threshold at 3 months ago.)

Table 2 (continued)

Item	Level of involvement	Question	Response options
ILO7	Problematic or disordered use	<p><i>(If ILO3 = Yes): During the past 3 months, have your relationships with friends, partners, or family [other examples may be added] ever been affected negatively because of using a prescription opioid?</i></p> <p><i>Note: an alternative wording may ask: During the past 3 months, how many times have your relationships with friends, partners, or family been affected negatively because of using a prescription opioid?</i></p> <p><i>Note: an alternative wording may ask: When was the last time that your relationships with friends, partners, or family were affected negatively because of using a prescription opioid?</i></p>	<p>Yes, No</p> <p>Never, 1 time, 2 times, 3–5 times, 6–9 times, 10–19 times, 20 or more times</p> <p><i>(Note: additional amounts may be added.)</i></p> <p>Past month, 2–3 months ago, 4–12 months ago, 1 + years ago, Never <i>(Note: time lags may vary, but must include a threshold at 3 months ago.)</i></p>

Bold items are included in all studies. Italicized text indicates alternative wordings, questions, or response options that could be used among studies from which the common base harmonized quantification can be derived

^aThe time frame for questions can vary from wave to wave. For the baseline survey, lifetime will be the frame of reference. For follow-up waves, the frame of reference will be since the last time data were collected from a participant

OUD criteria to a common severity continuum using data from the National Epidemiologic Survey on Alcohol and Related Conditions (Fig. 1). However, their analyses omitted indicators of less opioid involvement that were expected for many HPC study participants, as illustrated in Fig. 1. Thus, in developing opioid misuse involvement measures, emphasis was placed on a broader range of opioid involvement.

Strengths and nuances of HPC opioid use outcome measures are illustrated by the Involvement with Legally Manufactured Opioids (ILO) measure (Table 2) which was developed through the working group's activities prior to data collection (steps 3 to 5 described earlier). The standardized items selected for ILO were selected to stand as single-item measures (i.e., outcomes in their own right) and in part designed to track progressions in distinct phases in individuals' "involvement." The speed in progressing from a stage of lesser (e.g., initiation) to greater (e.g., regular use) involvement correlates with the construct of "addictive liability" used in animal studies (Ridenour et al., 2005, 2006). Thus, if a prevention program reduces the rate or speed of progression from lesser to greater involvement, it would be shown that the intervention reduces addictive liability. Moreover, seven anchor items were selected rather than the more common two to four anchor items in case not all items reflect a single latent construct of "involvement" (and could be omitted from the involvement measure).

Of course, the first stage of misusing an opioid is initiation (ILO1), although it could be argued that earlier "involvement" could consist of contemplating initiation. In fact, several HPC studies measure pre-initiation constructs such as perceived harms of misuse or use of opioids as prescribed. Item ILO1 queries whether a respondent has ever misused a legal opioid and how much use had occurred prior to an HPC baseline assessment. The wording for ILO1 was modeled after a similar item in the 2019 Youth Risk Behavior Survey (CDC, 2021) which asks, "...how many times have you used...?" to assess both whether a respondent misused an opioid and the amount of past use. One approach used in some research projects was to add ILO1a as a screening item to avoid any risk that ILO1 may convey a perception that opioid misuse is normative (e.g., youth in Emory University study). A second nuance was to add ILO1b in HPC studies with participants who are prescribed opioids and thus important to track (e.g., University of Michigan's health care patients). If a participant reported having never misused a legally manufactured opioid, the additional questions about involvement with legally manufactured opioids could be skipped.

Follow-up questions about misuse of prescription opioids start with asking age of onset of initiation (ILO2). This item provides the first age for testing time-to-event outcomes and for testing progressions in staged misuse of prescription opioids. The next two questions ask if (and age of onset) an opioid has ever been misused regularly, defined as misuse at least

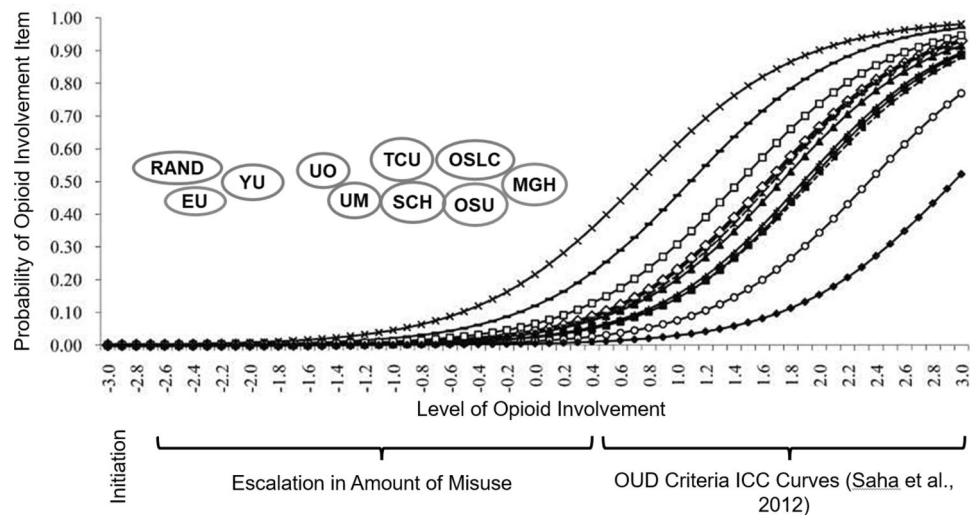


Fig. 1 Conceptual distribution of the average level of involvement with opioid misuse that is anticipated for HPC studies participants superimposed on item response theory results for opioid use disorder criteria. Note. Each oval indicates the level of opioid involvement that is anticipated for most participants of an HPC study sample at 12-month follow-ups. ICC, item characteristic curves of opioid use disorder criteria reported by Saha et al. (2012) (estimated using an item response theory two-parameter model). OUD, opioid use dis-

order. EU, Emory University/Cherokee Nation study. MGH, Massachusetts General Hospital study. OSU, Ohio State University study. OSLC, Oregon Social Learning Center study. RAND, RAND/UCLA study. SCH, Seattle Children's Hospital/University of Washington study. TCU, Texas Christian University study. UM, University of Michigan study. UO, University of Oregon study. YU, Yale University study

once per month for 3 months in a row. Item ILO5 asks about frequency of misusing prescription opioids (with the optional ILO5a to measure finer gradations of frequency).

The final two standardized questions about prescription opioid involvement query two DSM-5 and ICD-10 opioid use disorder criteria (Saunders, 2017). These two criteria were selected because they frequently have the earliest onset and greatest prevalence among adolescents and young adults, based on clinical experience of HPC members who provide treatment for opioid use disorder. Only two OUD criteria were included to accommodate the time constraints of two HPC studies and because the studies focusing on OUD outcomes use more comprehensive diagnostic measures. Tools to assess OUD criteria vary in their response options (yes/no, frequency of occurrence, recency of occurrence); thus, three alternatives to querying OUD criteria were permitted (see ILO6 and ILO7 in Table 2) to accommodate HPC studies' planned analyses. The "common denominator" response option that could be derived from these alternative questions was whether a participant had (not) experienced the OUD criterion.

Analytic Strategies for Testing Opioid Misuse Involvement Outcomes

By design, each item of the opioid misuse involvement measures could serve as an important outcome. Some binary indicators of opioid misuse such as initiation or OUD symptoms

are traditionally tested using categorical analyses (e.g., χ^2 tests, logistic regression). Continuous measures of opioid misuse (e.g., frequency) will likely yield greater statistical power compared to binary outcomes. To illustrate, one test of HPC intervention efficacy could be whether it delays the age at which progression to regular opioid misuse occurs using time-to-event analyses such as Cox regression. For studies in which many participants misuse opioids, outcomes might be analyzed based on normally distributed (Gaussian) outcomes using linear regression but may require transformed outcomes if only slightly skewed (e.g., log-linear transformation). Even in HPC studies in which skewed distributions are expected (e.g., frequency of use), outcomes could be tested using Poisson or negative binomial regression or the zero-inflated versions of these models (Preisser et al., 2016; Zaninotto & Falaschetti, 2011).

Because the latent opioid use involvement constructs use anchor items across all HPC research projects, scores on these measures provide opportunities for analyses that include all HPC studies, taking advantage of much larger samples because all HPC participants will have a value. ILO scores will likely be right skewed and require analysis using a form of regression that aligns with the severity and shape of the distribution. Such analyses also will require statistical accounting of clustering of participants within a study (e.g., using multilevel regression).

Each HPC intervention has a potential to reduce progressions for multiple levels of opioid use involvement. A change

from a lesser to a greater stage of involvement indicates a progression that can be analyzed using methods such as logistic regression or configural frequency analysis (Ridenour et al., 2005). Speed at which such transitions occur can be analyzed using time-to-event analyses such as Cox regression (Ridenour et al., 2005). Regression models of the rate or speed of transitions can incorporate multiple predictors and interaction terms that include study arm (Ridenour et al., 2006).

HPC Facets That Could Not Be Standardized

Despite the expanded research opportunities afforded by harmonization of HPC measures, diversity in research project designs and theoretical models limited standardization of assessments. For example, given increased risk of suicidal thoughts and behaviors among patients presenting to an emergency department (Hedegaard et al., 2020) and association between suicidal thoughts and behaviors and opioid misuse (Baiden et al., 2019), the University of Michigan study is screening for suicide risk. However, this construct was not standardized because it is rare among some research project samples (e.g., younger school students) or limited personnel are available to conduct ethically required risk management for those screening positive. Instead, other mental health constructs (i.e., depression, anxiety symptoms) were standardized given scientific focus across HPC projects and the availability of brief, reliable, and valid assessments. Moreover, some research projects included only adolescents whereas others spanned adolescent and adult age ranges, posing challenges to the validity of measures to be standardized since many scales are not developed and evaluated for the full age range. Where possible, age-specific items were included for parallel measurement by age (e.g., PROMIS pain items) or scales that are validated across the broad age range were used (e.g., PHQ-2; GAD-2).

Feasibility constraints precluded standardization of constructs that were less commonly studied across the HPC research projects. For example, several research projects require very brief assessments based on logistics of the study setting that require avoiding interruptions of core activities (e.g., education of students in schools, care of patients in the fast-paced emergency department). Similarly, only two items assessing opioid use consequences could be standardized across all HPC research projects, whereas some studies are using multi-item scales based on their intervention and study aims.

Finally, standardization of measures (as well as harmonization more broadly) in consortia such as the HPC increases burden to projects, which is a critical consideration when planning startup activities and timelines. Standardization and harmonization of measures took the HPC over 6 months, which resulted in some projects having to submit

IRB applications and initiate piloting activities (e.g., programming online surveys) prior to completion of measure standardization. In turn, these projects had to repeat some pre-study activities (e.g., IRB amendments, creation of new online surveys) given the extent of measure changes (e.g., editing or eliminating some original measures, accommodating HPC measures). Future consortia-driven research could be improved by planning for staff and time to complete these tasks, as standardization and harmonization of measures is clearly scientifically important to advance the field, providing key opportunities for cross-study collaboration. At the same time, allowing flexibility for each project's unique assessments based on the proposed sample, setting, and science is critical as well, with the ideal balance something we hope the HPC achieved in part using harmonization strategies other than standardization.

HPC Harmonization Strategies for Integrative Data Analysis

Analysis of data from multiple HPC research projects will require commensurate data from those studies. When standardized measures (i.e., identical measures across studies) are not feasible, statistical harmonization is a suitable alternative. Statistical harmonization involves a series of steps to (1) statistically anchor items of different measures of a construct onto the same underlying latent characteristic and (2) derive scores for individuals on the latent characteristic.

For HPC investigations, integrative data analysis (IDA) offers particularly salient advantages over alternative data pooling methods such as Individual Patient Data and meta-analysis. For instance, as noted, central to HPC is the collection of data capturing ILO to span initiation, regular use, escalating use (frequency), and problematic use. Given the frequently low base rate of opioid misuse in any individual sample, the ability to pool across several research projects for analytic purposes may be required to answer questions pertaining to severe opioid misuse. Moreover, IDA facilitates statistical harmonization to integrate scores from different research project measures of a construct onto a single latent variable, while accounting for measurement differences among the research projects.

Two broad approaches to combining measures that differ across studies include logical harmonization and psychometric (or analytic) harmonization. Logical harmonization refers to identifying like items across datasets used for pooling (such as HPC Standardized Items). For retrospective IDA, logical harmonization is often guided by content expert ratings, a crosswalk, or harmonization grid to determine one-to-one correspondence across items that query the same subconstruct on a factor of interest. However, decades of psychometric studies demonstrate that even the same

measure administered in different ways, in different locations, or to different samples may not perform equivalently (e.g., high school students may experience relationship problems because of opioid misuse more readily than adults who live alone) or subgroups of a sample may respond to some items in a systematically different way than other subgroups (e.g., persons without a job may not experience reduced fulfillment of obligations due to opioid misuse as easily as working individuals).

Psychometric harmonization techniques using differential item functioning (of item response theory, IRT), measurement (in)equivalence (of confirmatory factor analysis, CFA), and newer models that blend the two will test for, quantify, and account for measurement differences among HPC research projects. Psychometric harmonization with IDA equates measures across studies not at the item level but at the factor level. By accounting for how individual items differentially relate to their common underlying factor across studies or subgroups, commensurate measures of the factor can be created across studies.

Among the variety of approaches to psychometric harmonization, Moderated Non-linear Factor Analysis (MNLFA; Bauer & Hussong, 2009) is a highly flexible tool that combines the IRT and CFA traditions. MNLFA tests whether influences on items of a measure differ across participants in the pooled studies with respect to the factor mean and variance (assessing impact in the IRT tradition). MNLFA also tests whether item responses and factor loadings vary among participants of different studies even after controlling for impact. After quantifying such item differences among HPC research projects, the differences across studies are incorporated into an iterative test to psychometrically model them (parallel to approaches for evaluating invariance in the CFA tradition). Given the presence of some items that share meaning across samples (i.e., anchor items that are invariant in mean and loading across research projects), results of MNLFA may then be used to score measures in ways that take into account study differences, patterns of item endorsement (rather than simply how many items are endorsed), and other factors that contribute to differential item functioning. These factor scores may then be used in subsequent analyses to test substantive hypotheses (see Hussong et al., 2011).

HPC Research Opportunities Expected to Arise from Harmonization

Many questions that could not otherwise be tested in individual HPC research projects will be testable by harmonizing data across them. Because HPC recruitment targets at-risk samples, their opioid misuse rates are expected to

be greater than the general population at follow-up waves (especially in control groups). HPC data will provide opportunities to prospectively delineate cross-lag associations between presumed risk factors and opioid misuse. To illustrate, depression is measured the same way in all HPC studies as it likely increases risk for opioid misuse while consequences and biophysiological changes from opioid misuse may increase risk for subsequent depression (Volkow et al., 2019).

Additionally, interactions between risk factors and exposure to prevention program(s) that were not hypothesized a priori may be tested as moderators of opioid misuse outcomes. Such studies may utilize data from multiple HPC studies even if differences in samples or settings preclude harmonization of their data. To illustrate, greater educational attainment may bolster the impact of prevention programs that present dangers of opioid misuse (Haller et al., 2010). Replicating this finding across otherwise distinct interventions (e.g., RAND/UCLA, University of Michigan, Yale University) would identify a salient attribute by treatment interaction that might be used to tailor prevention across distinct populations. A subset of HPC studies harmonized measures of emotion regulation to investigate both its etiology roles in opioid misuse and test its potential as a mechanism of change (Texas Christian University) or a moderator of intervention outcomes (Massachusetts General Hospital, Seattle Children's Hospital, and Yale University). Similarly, testing the range of prevention programs with outcomes that vary according to type or number of baseline risk factors may reveal universal characteristics that must be accounted for if opioid misuse (or related outcomes such as suicidal thoughts and behaviors) is to be prevented (Brook et al., 2011). For example, perhaps the challenges that accompany having nonbinary gender identification or familial substance use disorder history require intervention that is specifically tailored for these individuals (Capistrant & Nakash, 2019).

Summary and Implications for Future Research Cooperatives and the HEAL Data Ecosystem

During the planning phase, HPC research project sites focused on attaining consensus agreement of the highest priority constructs and measures for the cooperative's main goals. This six-step process generated a defined set of core measures, "standardized measures," that supports analysis across the 10 HPC research project sites, detection of intervention impacts, and comparability among HPC studies and other national studies. Also of great interest, however, was that the HPC used both stringent and flexible harmonization approaches prospectively (Fortier

et al., 2011a) to ensure the quality of the data for a core set of priority measures while preserving the heterogeneity needed to fulfill the individual HEAL research project site's prevention intervention aims.

Two innovative aspects of maintaining flexibility in this coordinated approach are the ability to harness systematic variability within the aggregate sample from 10 distinct research projects and incorporation of synchronized measurement by statistically harmonizing them. By nimbly balancing this tension between centralized and decentralized data, the HPC's harmonization model is uniquely poised to generate understanding of intervention impacts on the full continuum of opioid misuse experience in adolescents and young adults with a precision that has yet to be attained (Volkow et al., 2019). We encourage other collaborative research networks to consider this dynamic harmonization process to push the boundaries of what individual research projects can do in isolation and maximize federal funding investments to generate public health innovations.

Over the past decade, NIH has shown a strong commitment to improving data infrastructure, analysis, and sharing to provide a foundation for conducting research and pursuing scientific innovation. Within the substance use and addiction research field, a variety of harmonization tools and platforms have enhanced feasibility and usage in epidemiology, prevention, and services research, including the PHENotypes and eXposures Toolkit; the Seek, Test, Treat, and Retain Data Collection and Harmonization Initiative; and the Justice Community Opioid Innovation Network's common core measures (Chandler et al., 2015; Ducharme et al., 2021; Hamilton et al., 2011). In this tradition, the NIH HEAL Initiative has supported its projects to invest in creating and sharing robust datasets that will be a part of the HEAL Data Ecosystem, a cloud-based platform to search for data and findings via a web interface that is under development (<https://heal.nih.gov/data/heal-data-ecosystem>). Because of its detailed data harmonization process outlined in this paper, the HEAL HPC is well suited to generate prevention evidence to reduce opioid misuse through public data sharing and use in the HEAL Data Ecosystem.

Funding This research was supported by the National Institutes of Health through the NIH HEAL Initiative as part of the HEAL Prevention Initiative. The authors gratefully acknowledge the collaborative contributions of the National Institute on Drug Abuse (NIDA) and support from the following awards: Emory University and the Cherokee Nation (UH3DA050234; MPIs Kelli Komro, Terrence Kominsky, Juli Skinner); Massachusetts General Hospital (UH3DA050252; MPIs Timothy Wilens, Amy Yule); The Ohio State University (UH3DA050174; MPIs Natasha Slesnick, Kelly Kelleher); Oregon Social Learning Center (UH3DA050193, Lisa Saldana); RAND Corporation (UH3DA050235, Elizabeth D'Amico, Daniel Dickerson); RTI International (U24DA050182; MPIs Phillip Graham, Ty Ridenour); Seattle Children's Hospital and University of

Washington (UH3DA050189; MPIs Kym Ahrens, Kevin Haggerty); Texas Christian University (UH3DA050250; PI Danica Knight); University of Michigan (UH3DA050173; MPIs Maureen Walton, Erin Boner); University of Oregon (P50DA048756; Elizabeth Stormshak); and Yale University (UH3DA050251; PI Lynn Fiellin).

Declarations

Ethics Approval All studies that are described or referenced herein have been reviewed and approved by the Investigators' respective Institutional Review Boards. All procedures with human subjects were performed in accordance with the ethical standards of the institute and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent to Participate N/A.

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Baiden, P., Graaf, G., Zaami, M., Acolatse, C. K., & Adeku, Y. (2019). Examining the association between prescription opioid misuse and suicidal behaviors among adolescent high school students in the United States. *Journal of Psychiatric Research, 112*, 44–51.
- Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods, 14*, 101.
- Bonar, E. E., Kidwell, K. M., Bohnert, A. S. B., Bourque, C. A., Carter, P. M., Clark, S. J., Glantz, M. D., King, C. A., Losman, E. D., McCabe, S. E., Philyaw-Kotov, M. L., Prosser, L. A., Voepel-Lewis, T., Zheng, K., & Walton, M. A. (2021). Optimizing scalable, technology-supported behavioral interventions to prevent opioid misuse among adolescents and young adults in the emergency department: A randomized controlled trial protocol. *Contemporary Clinical Trials, 108*, 106,523.
- Brook, J. S., Lee, J. Y., Finch, S. J., Koppel, J., & Brook, D. W. (2011). Psychosocial factors related to cannabis use disorders. *Substance Abuse, 32*, 242–251.
- Brown, C. H., Brincks, A., Huang, S., Perrino, T., Cruden, G., Pantin, H., Howe, G., Young, J. F., Beardslee, W., Montag, S., & Sandler, I. (2018). Two-year impact of prevention programs on adolescent depression: An integrative data analysis approach. *Prevention Science, 19*, 74–94.
- Capistrant, B. D., & Nakash, O. (2019). Lesbian, gay, and bisexual adults have higher prevalence of illicit opioid use than heterosexual adults: Evidence from the National Survey on Drug Use and Health, 2015–2017. *LGBT Health, 6*, 326–330.
- Centers for Disease Control and Prevention (CDC). (2021). *1991–2019 High School Youth Risk Behavior Survey data*. Retrieved from <http://nccd.cdc.gov/youthonline/>.
- Chan, A.-W., Krleža-Jerić, K., Schmid, I., & Altman, D. G. (2004). Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Canadian Medical Association Journal, 171*, 735–740.
- Chandler, R. K., Kahana, S. Y., Fletcher, B., Jones, D., Finger, M. S., Aklin, W. M., Hamill, K., & Webb, C. (2015). Data collection and harmonization in HIV research: The Seek, Test, Treat and Retain Initiative at the National Institute on Drug Abuse. *American Journal of Public Health, 105*, 2416–2422.

- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*, 81–100. <https://doi.org/10.1037/a0015914>
- Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., & Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research, 49*, 214–231.
- Department of Health and Human Services (DHHS). (2021). *Adolescent Brain Cognitive Development: The ABCD Study is the largest long-term study of brain development and child health in the United States*. Retrieved from <https://abcdstudy.org/>.
- Ducharme, L. J., Wiley, T. R., Mulford, C. F., Su, Z. I., & Zur, J. B. (2021). Engaging the justice system to address the opioid crisis: The Justice Community Opioid Innovation Network (JCOIN). *Journal of Substance Abuse Treatment, 128*, 108307. <https://doi.org/10.1016/j.jsat.2021.108307>
- Fortier, I., Doiron, D., Burton, P., & Raina, P. (2011). Consolidating data harmonization—How to obtain quality and applicability? *American Journal of Epidemiology, 174*, 261–264.
- Fortier, I., Doiron, D., Little, J., Ferretti, V., L'Heureux, F., Stolk, R. P., Knoppers, B. M., Hudson, T. J., Burton, P. R., On behalf of the International Harmonization Initiative. (2011). Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies. *International Journal of Epidemiology, 40*, 1314–1328. <https://doi.org/10.1093/ije/dyr106>
- Haller, M., Handley, E., Chassin, L., & Bountress, K. (2010). Developmental cascades: Linking adolescent substance use, affiliation with substance use promoting peers, and academic achievement to adult substance use disorders. *Development and Psychopathology, 22*(4), 899–916.
- Hamilton, C. M., Strader, L. C., Pratt, J. G., Maiese, D., Hendershot, T., Kwok, R. K., Hammond, J. A., Huggins, W., Jackman, D., Pan, H., Nettles, D. S., Beaty, T. H., Farrer, L. A., Kraft, P., Marazita, M. L., Ordovas, J. M., Pato, C. N., Spitz, M. R., Wagener, D., & Haines, J. (2011). The PhenX Toolkit: Get the most from your measures. *American Journal of Epidemiology, 174*, 253–260.
- Hedegaard, H., Curtin, S. C., & Warner, M. (2020). Increase in suicide mortality in the United States, 1999–2018. *NCHS Data Brief, No 362*. Retrieved at www.cdc.gov/nchs/data/databriefs/db362-h.pdf
- Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology, 9*, 61–89.
- Hussong, A. M., Jones, D. J., Stein, G. L., Baucom, D. H., & Boeding, S. (2011). An internalizing pathway to alcohol use and disorder. *Psychology of Addictive Behaviors, 25*, 390.
- LeNoue, S. R., & Riggs, P. D. (2016). Substance abuse prevention. *Child and Adolescent Psychiatric Clinics of North America, 25*, 297–305. <https://doi.org/10.1016/j.chc.2015.11.007>
- Makol, B. A., De Los Reyes, A., Ostrander, R. S., & Reynolds, E. K. (2019). Parent-youth divergence (and convergence) in reports of youth internalizing problems in psychiatric inpatient care. *Journal of Abnormal Child Psychology, 47*, 1677–1689.
- National Institutes of Health (NIH). (2018). HEAL Initiative: Preventing opioid use disorder in older adolescents and young adults (ages 16–30). Accessed on October 16, 2021 at: <https://grants.nih.gov/grants/guide/rfa-files/RFA-DA-19-035.html>.
- National Institutes of Health (NIH). (2021). *All of Us Research Program: The future begins with you*. Retrieved from <https://allofus.nih.gov/>.
- Perrino, T., Beardlee, W., Bernal, G., Brincks, A., Cruden, G., Howe, G., & Brown, C. H. (2015). Toward scientific equity for the prevention of depression and depressive symptoms in vulnerable youth. *Prevention Science, 16*(5), 642–651.
- Preisser, J. S., Das, K., Long, D. L., & Divaris, K. (2016). Marginalized zero-inflated negative binomial regression with application to dental caries. *Statistics in Medicine, 35*, 1722–1735.
- Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). The problem of bias in psychological assessment. In *Mastering modern psychological testing* (pp. 573–613). Springer, Cham.
- Ridenour, T. A., Lanza, S. T., Donny, E. C., & Clark, D. B. (2006). Different lengths of times for progressions in adolescent substance involvement. *Addictive Behaviors, 31*, 962–983.
- Ridenour, T. A., Maldonado-Molina, M., Compton, W. M., Spitznagel, E. L., & Cottler, L. B. (2005). Factors associated with the transition from abuse to dependence among substance abusers: Implications for a measure of addictive liability. *Drug and Alcohol Dependence, 80*, 1–14.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley & Sons.
- Saavedra, L. M., Morgan-López, A. A., Hien, D. A., Lopez-Castro, T., Ruglass, L. M., Back, S. E., & Hamblen, J. (2021). Evaluating treatments for posttraumatic stress disorder, alcohol and other drug use disorders using meta-analysis of individual patient data: Design and methodology of a virtual clinical trial. *Contemporary Clinical Trials, 107*, 106479.
- Saha, T. D., Compton, W. M., Chou, S. P., Smith, S., Ruan, W. J., Huang, B., Pickering, R. P., & Grant, B. F. (2012). Analyses related to the development of DSM-5 criteria for substance use related disorders: 1. Toward amphetamine, cocaine and prescription drug use disorder continua using Item Response Theory. *Drug and alcohol dependence, 122*(1–2), 38–46.
- Saunders, J. B. (2017). Substance use and addictive disorders in DSM-5 and ICD 10 and the draft ICD 11. *Current Opinion in Psychiatry, 30*, 227–237.
- Siddique, J., de Chavez, P. J., Howe, G., Cruden, G., & Brown, C. H. (2018). Limitations in using multiple imputation to harmonize individual participant data for meta-analysis. *Prevention Science, 19*(1), 95–108.
- Stanis, J. J., & Andersen, S. L. (2014). Reducing substance use during adolescence: A translational framework for prevention. *Psychopharmacology (berl), 231*, 1437–1453.
- Substance Abuse and Mental Health Services Administration (SAMHSA). (2019). *Substance misuse prevention for young adults*. Publication No. PEP19-PL-Guide-1. Retrieved from <https://store.samhsa.gov/sites/default/files/d7/priv/pep19-pl-guide-1.pdf>
- Tanner-Smith, E. E., & Lipsey, M. W. (2015). Brief alcohol interventions for adolescents and young adults: A systematic review and meta-analysis. *Journal of substance abuse treatment, 51*, 1–18.
- Volkow, N. D., Jones, E. B., Einstein, E. B., & Wargo, E. M. (2019). Prevention and treatment of opioid misuse and addiction: A review. *Journal of the American Medical Association Psychiatry, 76*, 208–216.
- Williams, N. J. (2016). Multilevel mechanisms of implementation strategies in mental health: Integrating theory, research, and practice. *Administration and Policy in Mental Health and Mental Health Services Research, 43*, 783–798.
- Zaninotto, P., & Falaschetti, E. (2011). Comparison of methods for modelling a count outcome with excess zeros: Application to Activities of Daily Living (ADLs). *Journal of Epidemiology & Community Health, 65*, 205–210.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or its NIH HEAL Initiative.