PREDICTING SELF-HARM AND IDENTIFYING CAUSAL RISK FACTORS AMONG ADOLESCENTS WHO HAVE HAD CONTACT WITH U.S. CHILD PROTECTIVE SERVICES

by
Geoffrey David Kahn

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
May 2021

Abstract

Purpose: Examine self-harm among adolescents following an investigation by Child Protective Services (CPS) for maltreatment, validate a predictive model, and identify modifiable causal risk factors.

Methods: Data came from the second National Survey of Child and Adolescent Well-being cohort – a nationally representative, longitudinal survey. Following multiple imputation of missing data, descriptive statistics and multivariable logistic regression accounting for the complex survey design were used to examine the odds of self-harm. A hold-out random forest was used to predict self-harm based on a large (>1,500) set of variables encompassing individual, family, and environmental information. Among the significant predictors, three were identified as modifiable by CPS: feelings of worthlessness, presence of supportive adults, and parental psychological aggression. For each, propensity score weighting (PSW) was used to control for observed confounders and the average effect of exposure among the exposed (ATT) was estimated using weighted logistic regression.

Results: The prevalence among older adolescents (15-17 years) remained stable over time at ~10% while among younger adolescents (11-14 years) it declined from 13% to 6% to 3%; 5% of adolescents reported self-harm at multiple survey waves. Native American and Asian/Pacific Islander youth were more likely to report self-harm at multiple waves: odds ratio 6.88 (2.02-23.5) compared to White non-Hispanic. The final predictive model had an AUC of 0.72. Prior self-harm was the strongest predictor, with internalizing problems, suicidal ideation, depression, and psychiatric medication following. Other predictors included trauma symptoms, parental monitoring and maltreatment, running away from home, and having supportive adults. For parental psychological aggression, the PSW odds ratios comparing low and high aggression to none were 0.93 (0.35-2.45) and 1.25 (0.55-2.82), respectively. For feelings of worthlessness it was 1.73 (0.70-4.27), and for supportive adults 0.58 (0.28-1.19). Due to the weighting the effective sample size was substantially reduced, which may have affected statistical power.

Conclusions: Further research should explore why Native American and Asian adolescents experienced more persistent self-harm, and potentially design culturally appropriate interventions. Given the modest prediction accuracy, using a machine learning algorithm to estimate risk for individuals within CPS is not currently recommended. However, fostering supportive and encouraging relationships with adults may play an important part in preventing self-harm among adolescents with CPS contact.

Thesis Advisory Committee:

Holly C. Wilcox, PhD

Elizabeth A. Stuart, PhD

Jeannie-Marie Leoutsakos, PhD

Additional readers:

Michele R. Decker, ScD

Vadim Zipunnikov, PhD

Preface

This work was made possible by grants from the National Institute of Mental Health. Specifically: the T32 Mental Health Services and Systems Training Program (Grant # 1T32MH109436-01A1) and an individual NIMH F31 NRSA (Grant # 1F31MH120973-01).

I would also like to thank my TAC, Drs. Holly Wilcox, Elizabeth Stuart, and Jeannie-Marie Leoutsakos. The knowledge you shared with me about suicide prevention, statistical methods, and public mental health research has been incredible, but it pales in comparison to the kindness, patience, support, and respect you have shown me throughout my entire time at Johns Hopkins. Words cannot express how deeply grateful I am for your mentorship; one could not ask for better. Thanks also to Drs. Decker and Zipunnikov, who took the time to serve not only on my final defense committee but also my preliminary examination, in which they offered many insightful suggestions that shaped this work. Thanks to Patty Scott for keeping the academic trains running on time, and to Scott Hubbard for making the grant submission process smoother than I thought it could be. A tremendous thank-you to Holly Larrabee at the NDACAN at Cornell University, who went above and beyond on more than one occasion to help me wrap my head around the thousands of variables in the NSCAW. Thanks to Drs. Noah Greifer and Calliope Holingue for their assistance with propensity score methods and R coding, and to Dr. Jill Rabinowitz for her extremely helpful reviews of Chapters 3-5. Many thanks to Hannah Del Porto for proofreading this entire monstrosity start to finish. Thank you (thank you, a million times thank you) to the incredible people who kept me sane through this laborious process: my kung fu family at Goh's; Rachael Shapira, who, in addition to being a source of unflagging intellectual and emotional support, did her level best to keep me from getting expelled for being a smart-ass; and the best cohort to ever grace the halls of academia, Sherri-Chanelle Brighthaupt, Shoshanna Fine, Angela Lin, Kyle Moored, John Thorne, Kayla Tormohlen, and Lindsey Webb. Finally, thank you so very much to my family. I won the lottery on the day I was born, and the older I get the more I realize the incredible scope of my good fortune. So as inadequate as these few lines are, thank you. Thank you to my Uncle Rick for giving me my first taste of real science (mito' preps!) so many years ago, and to him and my Aunt Pat for later providing me with a true home away from home when I was taking the early steps that would eventually land me here. Thank you to my sister, Alexandra Grace Kahn, who embodies her middle name. And thank you to my parents, mom and dad. What compassion I have, inspiring me to use research to try and help others, I learned from you, counting coins into the Guest at Your Table box. What intelligence and diligence I have, enabling me to get to this point, grew from spending time after school being guided to look up unfamiliar words in the dictionary for myself, and spending nights listening to you read Tolkien aloud to me. Thank you.

To my family - blood and otherwise - and my mentors, I hope I have done you proud.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Suicide is the tenth leading cause of death in the United States, and the second leading cause of death among adolescents and young adults ages 10-24 years (CDC, 2003). Suicide rates have been increasing steadily among adolescents (and most other age groups) for over a decade (Curtin, 2020). Surveillance of high school students indicates that most youth risk behaviors, including illicit substance use, dating violence, and early sexual activity, have declined over the past ten years, but suicidal ideation and suicide attempts have increased (CDC, 2020). Youth who suffer abuse or neglect at home are at increased risk for suicidal behavior (Angelakis, Austin, & Gooding, 2020). In the U.S., Child Protective Services (CPS) is the state government agency responsible for responding to allegations of child maltreatment. CPS caseworkers investigate whether child abuse has occurred and, if it has, intervene to ensure a safe environment for the child(ren). In most cases this involves connecting families with appropriate support services but, in cases of extreme or persistent abuse, may involve relocating the child temporarily or permanently to a foster home. Although children and adolescents with CPS contact represent a high-risk population for suicidal behavior compared to the general population (Evans et al., 2017), they are also a boundaried population within an existing government structure that could, potentially, deliver preventive interventions to the most at-risk youth.

However, there are challenges to such an approach. One is identifying which children are in need/would benefit from preventive interventions. Predicting who is at risk for suicide in advance of an attempt has been extremely difficult, with very little improvement made over the past half century of research (Franklin et al., 2017). Poor predictive accuracy is likely due, in large part, to the fact that suicide risk is multifaceted – the result of a complex web of risk and protective factors with no single factor either necessary or sufficient to cause self-harm (Turecki & Brent, 2016). The authors of the largest meta-analysis of suicide risk factors ever conducted concluded that one of the limitations of prior predictive models was that they incorporated too few factors simultaneously and did not allow for complex interactive effects (Franklin et al., 2017). Of course, given the difficulty of prediction, one could argue for a universal approach to suicide prevention; just treat everyone who enters the CPS system. If maltreatment is a risk factor for suicide then why not? Apart from the issue of limited resources, there is another challenge that affects even universal approaches, and that is the fact that there are no gold standard interventions for adolescent suicide prevention. Effective preventive intervention requires an understanding of the etiology of suicide. Knowing what causes suicidal thoughts and behaviors informs care providers as to the changes needed to prevent these behaviors. Most current preventive interventions and treatments have grown out of either clinical experience or some (unproven) theoretical model, and not out of a robust body of empirical data on causal risk and protective factors (Franklin et al., 2017). There are some existing interventions that have shown efficacy in trials, but many have either not been rigorously evaluated, have shown unconvincing results, or promising results have not been replicated in the limited trials that have been done. So the challenges of how to identify youth who are at risk for suicide and what to do if they are identified, even when a system is in place to intervene, remain unresolved.

New statistical methods have been developed in recent years that may help address these challenges. Machine learning (ML) algorithms have shown improved predictive accuracy over traditional regression models in a number of medical settings. Researchers have started to use machine learning methods to predict suicidal behaviors, but to date these methods have been employed mostly in adult populations (Burke, Ammerman, & Jacobucci, 2019). ML algorithms allow for greater flexibility in modeling how factors interact with each other and can incorporate large numbers of predictors simultaneously, which make these methods potentially well-suited to studying and predicting suicidal behaviors. In fact, in the large meta-analysis cited above, the authors suggest that researchers move from thinking about individual risk factors to thinking in terms of risk algorithms, and posit ML methods as appropriate tools for advancing our understanding of suicide (Franklin et al., 2017). Not only can ML algorithms incorporate large numbers of variables into a predictive model, there are now methods for quantifying how strong each predictor is. Thus, ML methods may also be useful for identifying factors or sets of factors that might be useful targets for preventive intervention from among a large pool of candidate variables.

The work in this dissertation was undertaken to advance our understanding of suicidal behaviors in youth with CPS contact. The prevalence of self-harm over a three year period was examined in a cohort of adolescents who had been the subject of a CPS investigation. A machine learning algorithm, specifically a random forest, was used to build and validate a predictive model to see if it might be feasible to employ an ML algorithm to screen youth for suicide risk as they enter the CPS system. The random forest model was also used to identify significant predictors from a pool of over 1,500 variables. A set of predictors which were thought to be modifiable from within the CPS system were then extracted. Propensity score methods were used to assess whether these variables might be causal risk factors and to quantify their effect on self-harm in order to prioritize targets for preventive intervention. The remainder of this chapter will provide: additional information about the burden of suicide among adolescents in the U.S., overviews of the distinct epidemiological features of adolescents in the CPS system and the data source used for the current analyses, a brief introduction to the random forest ML method, and some illustrative examples of preventive interventions which could be used within CPS. For a full review of prior studies using ML methods to predict suicide in other populations and studies examining self-harm in the data source used for this work, see Chapter 2.

A note about terminology

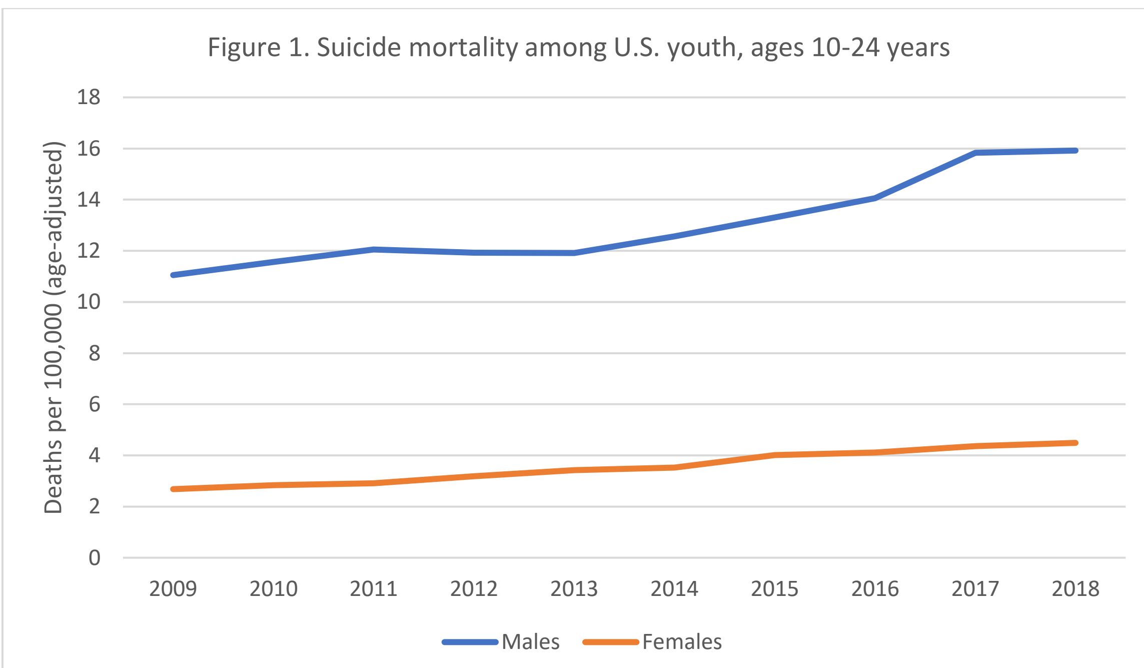
Throughout this manuscript, the reader may notice that most of the background information and discussion is related to suicide attempts, but the outcome from the dissertation analyses is referred to as “self-harm,” and not suicide attempt. As is described in Chapter 3, the survey questions used in the data source capture self-injurious behavior but do not confirm the presence of an intent to die, which is a necessary component for any action to be labelled a suicide attempt. It is likely that many, perhaps most, of the self-injurious behaviors in the NSCAW are suicide attempts but, because this cannot be confirmed, the decision was made to use the more accurate “self-harm” when describing the results. More detailed discussions of this limitation and the factors which justify the use of the current data source despite this limitation can be found in the proceeding manuscripts as well as the final concluding remarks in Chapter 6.

The burden of youth suicide and suicide attempts

As noted above, suicide is the second leading cause of death for person ages 10-24 years in the U.S. This information comes from the Centers for Disease Control and Prevention (CDC), which tracks the number of suicide deaths in the U.S. and makes the data publicly available through the Web-based Injury Statistics Query and Reporting System (WISQARS). While these data are frequently used in research, it is important to understand that there are some limitations when it comes to identifying deaths by suicide. Information from death certificates is reported to the CDC’s National Center for Health Statistics, which compiles, verifies, and then releases the data. Classifying a death as a suicide is usually the responsibility of a state medical examiner or coroner. While efforts have been made to provide standard criteria for making such a determination (CDC, 1988), suicides are still likely under-reported (Timmermans, 2005). Under-reporting is primarily due to the difficulty in establishing intent, but also to reticence to subject the surviving family to possible social stigma, guilt, and/or loss of insurance benefits.

With this fact in mind, the burden estimates provided by WISQARS should be considered at best a minimum estimate.

WISQARS shows that youth suicide rates have increased consistently over the past decade (Figure 1). The age-adjusted mortality rate for boys and young men (ages 10-24 years) increased from 11.05 per 100,000 in 2009 to 15.92 in 2018, an increase of 44%. For girls and young women, the mortality rate during the same period increased from 2.68 to 4.49, an increase of 68%. These rates correspond to a burden of 6,807 deaths by suicide in 2018 (CDC, 2003).

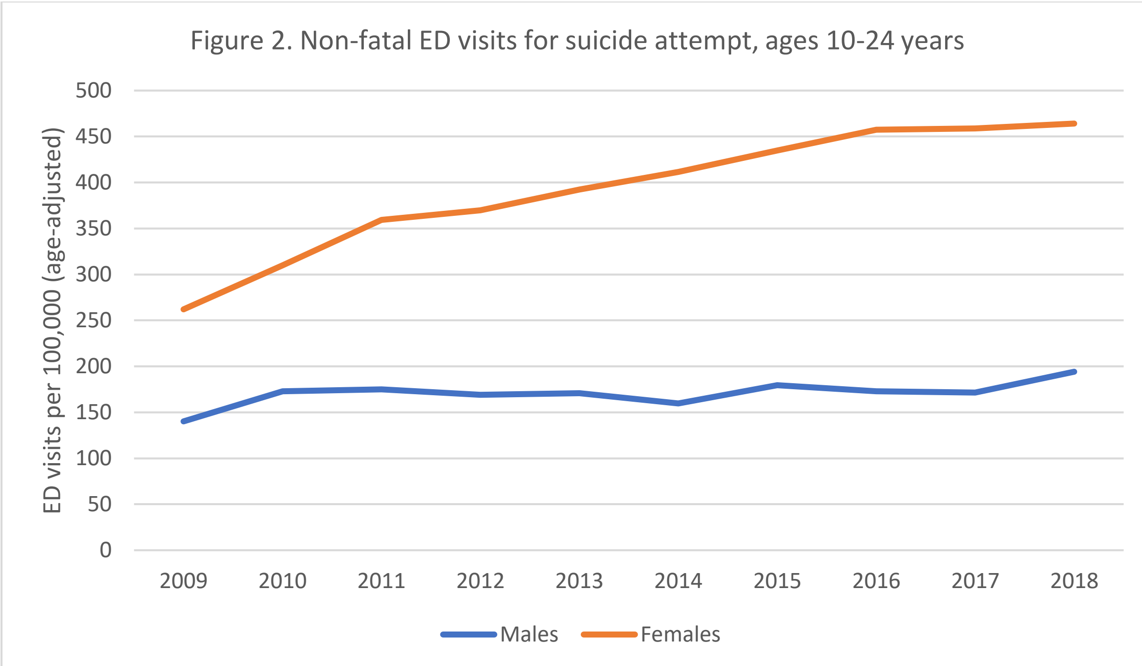


CDC Web-based Injury Statistics Query and Reporting System (WISQARS) www.cdc.gov/injury/wisqars/

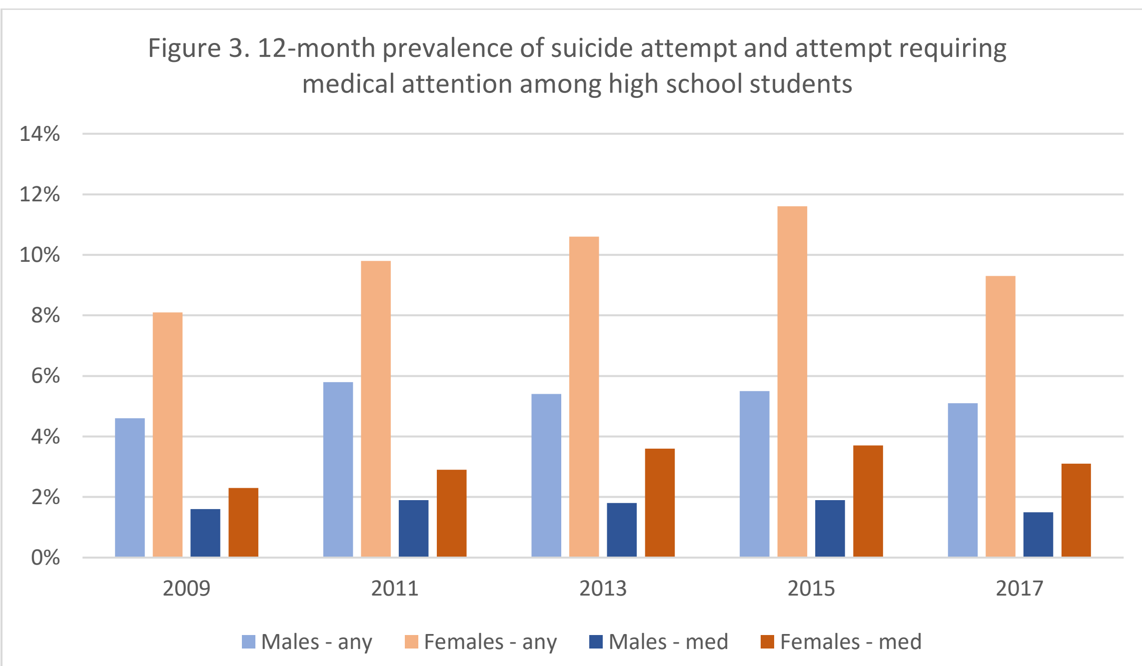
In additional to mortality data, the CDC collects data on non-fatal suicide attempts from two sources. The first is a hospital-based surveillance system operated by the Consumer

Product Safety Commission called the National Electronic Injury Surveillance System All Injury Program (NEISS-AIP). The NEISS-AIP collects data from emergency department (ED) admissions to a nationally representative sample of 66 U.S. hospitals. Reports from these 66 hospitals are then weighted to estimate the number of ED visits seen nationally; these data are also made publicly available through the WISQARS system. The second source of information is a nationally representative survey of high school students conducted by the CDC every other year, the Youth Risk Behavior Surveillance System (YRBSS). Respondents self-report whether they had any suicide attempt or an attempt requiring medical attention within the past 12 months. Data are publicly available through the CDC's YRBSS webpage. Many states also conduct a similar survey among middle school students, but there is currently no national dataset available for youth in middle school.

Non-fatal suicide attempts have, unsurprisingly, increased substantially over the past decade as well (Figure 2). The age-adjusted rate of emergency department visits for a non-fatal suicide attempt for boys and young men increased from 140.17 per 100,000 in 2009 to 194.18 in 2018, a 39% increase. For girls and young women, the rate of suicide attempts increased from 262.03 in 2009 to 464.11 in 2018, an increase of 77%. (CDC, 2003) Using WISQARS estimates, it appears that for every death by suicide there are 32.5 attempts treated in hospital EDs. And whereas surveillance of U.S. high school students indicates that most youth risk behaviors have declined over the past decade – including illicit substance use, dating violence, and early sexual activity – suicidal ideation and attempts have increased (CDC, 2020). Approximately 5% of boys and 9-10% of girls report at least one suicide attempt in the past year, and approximately 2% of boys and 3% of girls report at least one attempt that required medical attention (Figure 3).



CDC Web-based Injury Statistics Query and Reporting System (WISQARS) www.cdc.gov/injury/wisqars/



Youth Risk Behavior Surveillance System www.cdc.gov/healthyyouth/data/yrbs/

Children with CPS contact and the NSCAW data source

Children who suffer maltreatment at home are at an increased risk for suicidal behaviors above and beyond the rates described above; a recent meta-analysis showed a robust association between abuse/neglect before age 18 and suicidal thoughts and behaviors, with odds ratios for different forms of abuse ranging from 1.79 (95% CI: 1.27-2.53) for physical neglect to 3.41 (2.90-4.00) for sexual abuse (Angelakis et al., 2020). The authors concluded that, despite some heterogeneity in effect sizes, significant associations between maltreatment and suicidal behaviors were robust across 79 included studies of over 250,000 total participants. Unsurprisingly, research has shown that children under CPS supervision, either in out-of-home foster care or supervised home care, are at greater risk for suicidal behavior compared to the general population, with a recent meta-analysis estimating an odds ratio of 3.89 (3.14-4.83) (Evans et al., 2017). The prevalence of self-reported suicide attempts was also significantly higher among high school-aged children with CPS contact than for the general population, although the magnitude of the difference was moderate: 11.3% (95% CI: 6.5-19.0%) versus 7.8% (7.1-8.5%) (Heneghan et al., 2015).

It remains unclear whether the increased risk of suicidal behaviors among CPS youth represents just the effect of the maltreatment to which CPS responds, or whether there are additional, independent risks associated with CPS contact (removal from the home, for instance) or other confounding factors (more parental psychopathology, for example). Some known risk factors for suicidal behavior are found at higher levels among youth in CPS care than the general population, but others are not and may even be less prevalent (Heneghan et al., 2015). Both internalizing behavior problems (e.g., anxiety and depression) and externalizing behavior problems (e.g., aggression and disruptiveness) are associated with suicidal behavior in adolescence (Verona, Sachs-Ericsson, & Joiner, 2004; Wanner, Vitaro, Tremblay, & Turecki, 2012), and nearly half of children when they enter the CPS system score above the cutoff for clinically significant problems on the Child Behavior Checklist, a common measure of both internalizing and externalizing problems (Burns et al., 2004). However, there is growing evidence that both internalizing and externalizing problems are the result of childhood abuse and, in fact, it may be these conditions which mediate the association between maltreatment and suicidal behavior (Duprey, Oshri, & Liu, 2020).

Among youth in general, a diagnosed psychiatric disorder is a strong risk factor for suicidal behavior, with multiple co-morbid disorders conferring even greater risk. Up to 80% of youth who attempt suicide and 90% who die by suicide had a diagnosable psychiatric disorder (Bridge, Goldstein, & Brent, 2006). Mental health treatment during adolescence is thus an important consideration. Data from the general population in the late 1990s suggested that approximately 20% of youth with mental health needs were receiving appropriate care (Kataoka, Zhang, & Wells, 2002). At roughly the same time, entry into the CPS system was found to significantly increase mental health service use; with 30% of youth with CPS contact who had mental health needs receiving care even when there were no additional welfare services provided by CPS, 35% of youth receiving care when at-home welfare services were provided by CPS, and 64% receiving care when youth were placed in foster care (Leslie et al., 2005). It should be noted that access to mental health specialty care has changed rather dramatically over the past 30 years, with the CDC now estimating that between 50-80% of youth who need it are receiving care (with variations by specific disorder) (CDC, 2021). Youth with CPS contact may represent a population with a high risk profile for suicidal behavior, but also a population with greater access to care and supports that could facilitate interventions.

In addition to being a vulnerable population with a unique epidemiological profile, youth with CPS contact also represent a large population in the U.S. The Administration for Children & Families (ACF) within the Department of Health & Human Services provides annual estimates of the numbers of children who have had contact with CPS and who are in foster care. In 2018, CPS received 4.3 million referrals involving 7.8 million children, 2.4 million (56%) of which required some response from CPS (ACF, 2020). The rate of CPS referrals requiring a response in 2018 was 32.5 per 1,000 children in the national population. That same year there were 437,283 children in foster care, with 262,956 having entered foster care that year, and 687,345 children served by the foster care system (ACF, 2019).

Recognizing these youth as an important sub-population, in 1999 the ACF contracted with RTI International to launch a nationally representative, longitudinal study of children and their families who had been the subject of a CPS investigation: the National Survey of Child and Adolescent Well-Being (now called the NSCAW I). This landmark study and its subsequent cohorts remain the most comprehensive source of information on the health trajectories of children involved with CPS in the U.S. The NSCAW included children and families regardless of the outcome of the CPS investigation, but over-sampled infants and children in foster care to allow for robust analyses of these sub-populations that were deemed to be uniquely vulnerable. Reports were collected from face-to-face interviews at regular intervals with children, their parents and other caregivers, teachers, and CPS caseworkers, as well as administrative records. Baseline interviews were conducted within approximately 4 months of the conclusion of the CPS investigation; baseline data for the NSCAW I were collected in 1999-2000, and the last wave of data collection occurred in 2007. In 2008, data collection began for a second cohort, which became the NSCAW II. Baseline data for the NSCAW II were collected in 2008-2009, with two follow-up waves at approximately 18- and 36-months post-baseline (Dowd et al., 2014). As of this writing, data collection for a third NSCAW cohort is currently ongoing, though no data have yet been made public.

Use of the NSCAW data to study suicide attempts and self-harm in youth has been limited to date. In particular, studies using the most recent cohort have used only the baseline survey and not taken full advantage of the longitudinal data. See Chapter 2 for a summary of prior research.

Introduction to prediction with a random forest classifier

As noted above, the random forest machine learning algorithm was selected for this analysis because it can incorporate large numbers of variables simultaneously and flexibly model multi-way interactions. Both of these characteristics are now thought to be important for modeling suicidal behavior, which is likely the result of a complex chain of multiple risk and protective factors. Prior studies have used random forests to predict suicide attempts in other populations with generally promising results; see Chapter 2. The method used in this study is described in detail in the corresponding manuscript (Chapter 4), but that description assumes familiarity with the random forest concept. This section is intended for readers who need a brief and non-technical introduction.

Imagine that one is interested in classifying individuals into one of two categories, those who attempt suicide and those who do not based, for example, on an arbitrary number of predictor variables. The random forest, initially developed by Breiman (Breiman, 2001), is an extension of an older method, the classification tree (Breiman, Friedman, Olshen, & Stone, 1983). In a classification tree, the full sample is split into two daughter samples (called nodes) that are each as homogenous as possible with regards to the outcome. The split occurs following a decision rule based on one of the predictors. For example, the sample may be split based on presence/absence of a psychiatric diagnosis, or based on age, with one node composed of individuals under 18 years, and the other individuals 18 years and older. Decision rules for each predictor are considered, and within each predictor splits on every possible cut-point for continuous variables and every possible set of categories for nominal variables are assessed, with the decision rule that results in the most homogeneous nodes being adopted. The process is then repeated for each subsequent node until all individuals are perfectly classified or, more likely, some stopping rule (like minimum node size) is hit. At the end of the process, one has a decision tree; the nodes at the end of each decision chain are called terminal nodes, and individuals are classified based on the mode outcome value of the terminal node into which they are placed.

Classification trees are flexible and make no parametric assumptions but are prone to over-fitting. Random forests help alleviate this problem without sacrificing predictive accuracy. A random forest is an assemblage of individual classification trees with two key differences. First, rather than fitting a tree to the actual sample, each tree is fit to a bootstrap sample, which is a dataset of the same size as the original dataset that was generated by sampling with replacement from the original data. Second, at each split rather than consider all possible predictors, only a random subset of a prespecified size is considered. The goal is to create trees that are each grown from the observed data but are as independent from each other as possible. Several hundred or thousand individual trees are grown, and the final classification for each individual is the mode of that individual's classifications across all trees.

Several slightly different methods based around this basic framework now exist, including the one described in Chapter 4. Readers interested in a more thorough introduction to random forests and other machine learning methods are encouraged to read An Introduction to Statistical Learning with Applications in R from Springer Texts in Statistics (James, Witten, Hastie, & Tibshirani, 2013).

Example preventive interventions

It should be noted that there is currently no gold standard for either treatment of youth with suicidal ideation or prior suicide attempts, or for primary prevention of suicidal thoughts and behaviors, and so the question of what to do if high-risk individuals or salient risk factors are identified is an apt one. Fortunately, suicide prevention is an area of active and on-going research, and there are a variety of family-focused interventions and therapies that have shown some effectiveness at reducing suicidal behavior and/or ideation in adolescents. These could be appropriate to implement within CPS, either through direct service provision by CPS staff or referral to care providers that have a relationship with CPS, depending on the structure of each state's CPS organization. For adolescents with active suicidal ideation or self-injurious behaviors, two treatment modalities have some evidence of efficacy. Integrated Cognitive Behavioral Therapy (I-CBT) combines individual and family CBT techniques with a parent training component. In a randomized trial 6 months of I-CBT therapy was shown to reduce suicide attempts compared to treatment as usual (TAU) controls over an 18-month study period in adolescents with alcohol or drug use disorder (Esposito-Smythers, Spirito, Kahler, Hunt, & Monti, 2011). Similarly, Attachment-Based Family Therapy (ABFT) seeks to improve parent-child emotional attachment bonds and also includes parent skills training. In a randomized trial, 3 months of ABFT was shown to reduce the severity of suicidal ideation over a 6-month study period compared to TAU among adolescents who had been treated in the ED or primary care for suicidal ideation (Diamond et al., 2010). Suicidal behaviors were not studied in the ABFT trial. While it is promising that both I-CBT and ABFT showed benefits that lasted beyond the treatment period, in both trials the control groups had significantly worse compliance with the treatment as usual regimen and so it is difficult to tell what the I-CBT and ABFT treatments were really compared to.

There are also primary prevention interventions that have shown promise at reducing suicide risk, although large-scale trials are lacking. The Family Check-Up (FCU) is an adaptive program targeting parenting skills and family functioning in which the specific intervention targets and treatment intensity are tailored to the needs of individual families. FCU employs mostly motivational interviewing techniques. When implemented as a school-based intervention, a randomized study showed that adolescents who received FCU in middle school had significantly lower “suicide risk” (a composite score incorporating both self-reported suicidal ideation and attempts, but not distinguishing between the two) at age 18-19 years compared to controls (Connell, McKillop, & Dishion, 2016). The Youth Aware of Mental Health (YAM) intervention is a program that is entirely youth-focused, as opposed to the family-focused interventions described to this point, and is intended to build resilience, emotional intelligence, and coping skills, as well as basic mental health literacy, in youth. YAM was tested in one of the largest randomized trials of a suicide prevention intervention to date: the Saving and Empowering Young Lives in Europe Study – a multicenter, cluster-randomized study enrolling over 11,000 children in 168 middle schools across 10 European countries. At 12 months follow-up, children in schools where YAM had been deployed had half the odds of suicidal ideation and suicide attempts compared to controls (Wasserman et al., 2015).

Promising interventions for suicide prevention exist that might be adapted for deployment in the CPS system, although currently there is nothing that could be considered a best practice for prevention. Programs targeting family dynamics, parenting skills, and youth resilience have demonstrated efficacy, and there are therapies for adolescents who are already experiencing suicidal thoughts and behaviors, as well as primary prevention strategies. The current research sought to inform which preventive interventions would be expected to have the greatest impact if deployed in a CPS setting by identifying risk factors that have the greatest causal association with adolescent self-harm among adolescents in the CPS system.

Chapter 2

Literature review

SYSTEMATIC REVIEWS

The most comprehensive review of suicide risk factors, and a seminal publication in the field of suicide research, was published in 2017 by Franklin and colleagues (Franklin et al., 2017). The authors conducted an unprecedented review and meta-analysis of papers published from December 1965 through January 1, 2015. Papers which examined risk factors for suicide mortality, non-fatal suicide attempts, or suicidal ideation – and in which the factors were known to precede the outcome – were included. In all, 365 studies (3,428 risk factor effect sizes) were assessed using random effects meta-analysis models to pool the results. The authors found that longitudinal prediction was only slightly better than chance for all outcomes; that no broad category or subcategory of risk factors stood out as improving prediction significantly; and that predictive accuracy has not improved over time, despite all of the research that has been done. The average study was nearly 10 years long, but longer studies did not produce better prediction. The authors further note that studies rarely examined the combined effect of multiple risk factors, and that five broad categories of risk factor (internalizing psychopathology, demographics, externalizing psychopathology, prior suicidality, and social factors) have accounted for nearly 80% of all risk factor tests. The authors conclude that models which account for the complex interaction of multiple risk factors are needed, and recommend the use of new machine learning models for improving prediction.

The field seems to have taken Franklin's recommendation to heart. Numerous studies using a variety of machine learning methods to predict self-harm have been published in the few years following the Franklin review. So many that two additional systematic reviews have been published which provide important updates to the Franklin review. The first, published by Belsher and colleagues in 2019, was a systematic review of studies, conducted through August 2018, that characterized longitudinal predictive accuracy of suicide mortality or attempts

(Belsher et al., 2019). This review only included studies of adult populations or populations in which >80% of the participants were over 18 years of age. Seventeen cohort studies met the inclusion criteria, evaluating 64 different statistical models in over 14 million participants. Studies came from five countries, although 7 of the 17 studies were conducted solely with U.S. military participants (either active duty Army or military veterans). The authors note that studies were generally of high quality and with low risk of bias. Overall global classification accuracy was good, with most models predicting suicide mortality and nonfatal attempts having an AUC >0.8. However, only 5 of 11 mortality studies reported a PPV >1%, and those that did were generally in unrepresentative population samples. ML methods generally outperformed traditional regression. The authors conclude that machine learning represents an improvement in the field, but that PPV is still too low for algorithms to be recommended for practical use. The other review was conducted by Burke and colleagues, also published in 2019 (Burke et al., 2019). It was a systematic review of studies, conducted through February 2018, which specifically used a machine learning method to predict self-harm. However, studies did not have to be longitudinal, and of the 35 studies which met inclusion criteria, only 10 studies were longitudinal; 5 predicting suicidal behavior and 5 ideation. The authors agree with Belsher at al. that machine learning methods improve prediction compared to traditional statistical methods, but that there remains substantial room for further improvement. ML studies have often confirmed the importance of known risk factors for self-harm, and have occasionally uncovered new risk factors or, more commonly, distinct sub-groups of high-risk individuals. This despite that fact that the most effective prediction models are often “black box” methods where interpreting the effect of individual predictors is less straightforward.

The following section will briefly describe those individual studies which are most pertinent to the current work: studies which used a machine learning method to longitudinally predict self-harm.

INDIVIDUAL STUDIES

United States

All studies conducted in the U.S. to date have used a retrospective cohort design, drawing from electronic medical records (EMR) and sometimes supplemented with other data sets, like the National Death Index (NDI). Half of the studies published in the U.S. have concerned either military veterans or active duty Army soldiers. After noticing a rise in suicide deaths among soldiers – who historically had had suicide rates below the national average – the U.S. Army in partnership with the NIH launched a large-scale research project in 2008, the Study to Assess Risk and Resilience in Servicemembers (Army STARRS, https://starrs-ls.org/#/). Its aim was to develop actionable data to help reduce suicide and self-harm among service members. Owing to its focus on applied research, many studies using STARRS data have focused on algorithms for identifying high risk groups and evaluating predictive accuracy, but also potential clinical utility specifically within these groups. This is in contrast to other studies in the U.S. and elsewhere that report measures of predictive accuracy, like area under the ROC curve (AUC), for the whole study sample.

The first U.S. study was conducted among veterans but actually predates the Army STARRS study. Ilgen and colleagues conducted a retrospective cohort study of patients in the Veterans Administration (VA) healthcare system who were treated for severe depression (defined as either a diagnosis plus a prescription for anti-depressant medication, or two or more separate depression diagnoses, and excluding patients with co-morbid bipolar or schizophrenia diagnosis) from 1999-2004 (N = 887,859) (Ilgen et al., 2009). The outcome of interest was suicide death, which was identified through a search of the NDI. This study was focused on identifying high-risk sub-groups, but not prediction *per se*. A Bayesian Dirichlet Equivalent decision tree was built with a very limited predictor set: demographics plus five variables related to other conditions, including substance use disorder (SUD) and post-traumatic stress disorder (PTSD). One third of the sample was set aside for model validation. The model did identify groups of patients at different risk for suicide, although it is difficult to compare this model's performance with others. AUC and other prediction measures were not reported. Instead, chi-squared tests were used to assess whether suicide was distributed non-randomly within each terminal node of the decision tree; all tests were significant in the validation set. However, the prevalence of suicide within each node varied only from 0.08% to 0.77%, and so the tree is not useful for classifying individuals. The first split in the tree (a qualitative measure of variable importance) was on SUD; most splits were then on race or gender with one for inpatient psychiatric treatment in the previous year. Single decision trees are now seldom used for prediction, as tree ensembles are known to produce more accurate results and, most importantly, are less likely to overfit, something decision trees are prone to. It is also unclear why the authors excluded patients with bipolar disorder and schizophrenia but included other psychiatric conditions as predictors.

Two analyses emerging from the Army STARRS study have used machine learning to predict suicide mortality among slightly different populations. One study was of U.S. Army soldiers who had been hospitalized for a psychiatric illness between 2004 and 2009 (N = 40,820) (Kessler et al., 2015). The outcome was suicide within 12 months of discharge from the hospital. Data from several administrative and medical databases were integrated, and 421 predictors were evaluated. The authors were interested in using machine learning not as the primary analysis method, but to support variable selection for a conventional discrete-time survival model; it is not entirely clear why, but one would guess for interpretability. Following tests of bivariate associations, 100 single classification trees were built from bootstrapped samples to identify possible interactions. Interactions appearing in >10% of trees would have been included, but no such interactions were identified. Candidate predictors were included in an elastic net, a type of penalized regression which can aid in variable selection by allowing users to drop the most heavily penalized coefficients (i.e. those with the least effect on the outcome). The penalty coefficient was chosen through internal 10-fold cross validation. Seventythree predictors were initially retained, but the authors reported multicollinearity and high variance inflation factors for some variables, so forward stepwise selection was used to reduce the predictor set further to a final set of 20 variables. There was no independent model validation. The final model had an AUC of 0.84. The elastic net model with 73 predictors had an AUC of 0.85, although again this number was not generated from independent validation data. The authors were interested in stratifying people by risk, and noted that 53% of all suicides occurred among the top 5% of risk scores. Significant predictors included male sex, Army career characteristics (high age at enlistment and high AFQT score), access to firearms, crime perpetration, prior suicidality, prior psychiatric treatment, and hospitalization in a civilian hospital.

The other STARRS study was of soldiers who had a visit to an outpatient specialty mental health clinic between 2004 and 2009 (Kessler, Stein, et al., 2017). The outcomes were suicide within 5 and 26 weeks of the last clinic visit. Owing to the large and potentially cumbersome number of people included in the study population, a probability sample of ~100 controls was selected for each suicide case (N = 569) and then weighted to equal full sample. Approximately 1,000 variables were assessed, falling into six broad categories: socio-demographics, Army career, characteristics of the index visit, prior clinical factors, crime involvement, and other contextual factors (e.g. weapon ownership). Variables that were significant in univariate discrete-time survival analysis were further considered. Four analytic methods were compared: naïve Bayes, random forest, support vector regression, and elastic net. The “best” model was defined as the one with the highest cross-validated sensitivity in the subgroup of the 5% of visits with the highest risk score. Variable selection occurred in the full sample, but coefficients were re-estimated using the 2004-2007 data, and validated with 2008-2009 data and 2010-2012 data. The elastic net classifier was selected as the optimal model. It produced an AUC of 0.72 for suicide within 5 weeks and 0.67 for 26 weeks. Sensitivity in top 1-3 risk ventiles was 29.8-47.4% for 5 weeks and 26.7-41.3% for 26 weeks. Sensitivity was much lower for 2010-2012 data, 18.1-27.4% and 13.3-36.1% for 5 and 26 weeks, respectively; the AUC was not reported. Fewer than 20 variables were included in final models. Among cases with a prior psychiatric hospitalization, predictors were mostly indicators of prior suicidality and depression treatment; among cases without a prior hospitalization, predictors included various prior treatment indicators and crime perpetration, but no demographic variables.

The final study involving the U.S. military population was a retrospective cohort of 6,359 suicide deaths among veterans from 2009-2011, identified through an NDI search, who used VA healthcare services up to 2 years prior to death, and a 1% probability sample time-matched to deaths (N = ~2.1 million) then weighted to equal the full control population (Kessler, Hwang, et al., 2017). 2009-2010 decedent data (and their controls) were used for model fitting (split-half) and then validated against 2011 decedents. 381 measures of VA service use up to 2 years prior to death were evaluated as predictors. An elastic net discrete-time logistic regression model predicting suicide in the next 30 days was compared to eight other machine learning methods. The final net model included 61 predictors and had a sensitivity for the top 5% of risk scores of 26.3%. A Bayesian additive regression tree slightly outperformed the elastic net with sensitivity 28.1%. AUCs were not reported, nor were any measures of predictor importance. The authors mention that PPV would max out at 0.3% in this population, so PPV was also not reported.
Most studies of suicide and self-harm in the U.S. civilian population, like the studies above, have used the retrospective cohort design. Data on predictors and non-fatal outcomes generally come from EMR collected at individual hospitals or large healthcare systems, with data on suicide mortality coming from the NDI although most civilian studies have looked at non-fatal suicide attempts rather than mortality. Two studies, described below, used data from nationally representative longitudinal surveys. Barak-Corren and colleagues conducted a retrospective cohort using EMR data from patients at two large academic medical centers in Boston who had at least three inpatient and/or outpatient visits between 1998 and 2012 (Barak-Corren et al., 2017). The outcome of interest was suicide attempt treated in hospital or suicide death, with

both outcomes pooled. Suicide attempts were captured in the EMR and deaths identified through the NDI. After excluding patients with missing data (18.6%) 16,588 cases (852 deaths) and 1,708,197 controls were enrolled. There was an average of 5.3 years follow-up per patient, and only 10% of the sample was <25 years old. The authors used a naïve Bayesian classifier, and split the data randomly in half for model training and validation. Separate models for men and women were constructed. Predictors consisted of demographic data and diagnostic codes, lab results, and medications; codes and medications were not grouped, so thousands of variables were assessed. Variable importance was quantified as odds ratios. Model performance was similar between men and women, with AUC of 0.76 for men and 0.77 for women. With specificity set at 90% sensitivity and PPV in men were 44% and 5% (compared to a base rate of 1.55%), and 46% and 3% (compared to a base rate of 0.9%) in women. The strongest predictors included opioid abuse, and bipolar and personality disorders; in general, mental health disorder codes topped the list. Medium predictors included miscellaneous injuries like open wounds, superficial injury, crushing injury, and other external injuries which could all be unrecognized suicide attempts, although the authors do not mention this possibility. Important medications tend, unsurprisingly, to be psychiatric medications.

Walsh and colleagues conducted a retrospective cohort study of 5,543 adult patients from a single hospital (Walsh, Ribeiro, & Franklin, 2017). Cases were defined as patients with a nonfatal suicide attempt (N = 3,250 after dropping 376 suicide deaths) and were compared to two sets of controls: patients who had another self-injury diagnosis, either accidental injury or non-suicidal self-injury (NSSI) (N = 1,917), and a random sample of the general hospital population with no history of suicidality (N = 12,695). The authors used a random forest analysis, with decrease in Gini as the measure of variable importance. Bootstrap optimism adjustment was used instead of independent model validation. Predictors included socio-demographics, diagnosis, prior healthcare use, medication, and prior suicidality measured at eight different time windows ranging from 1 week to 2 years prior to the attempt. Multiple imputation was used to fill in missing data. The optimism-adjusted AUC for NSSI controls varied 0.80-0.84 depending on the time window, with predictors measured closer to the attempt tending to perform better, although not by much. The authors also ran separate analyses for cases with a single attempt versus multiple attempts, but results did not differ substantially. The model performed better with general hospital controls, with an adjusted AUC ranging 0.86-0.92. The random forests substantially out-performed traditional logistic regression. Demographic variables and diagnoses of recurrent depression with psychosis or schizophrenia/schizo-affective disorder were important predictors at most time windows. Substance use disorder was important at shorter time windows. Hospital utilization history and visit tallies were important at longer windows, as was medication use (except melatonin, which was important at short windows). Prior suicidality codes as well as codes that could indicate intentional self-harm (e.g. poisoning) were also consistently important.

Walsh and colleagues also used the same data source and methods to examine nonfatal suicide attempts among adolescents (Walsh, Ribeiro, & Franklin, 2018). This is the only study to date which has examined self-harm among young people, although one other study has examined suicidal ideation among adolescents. This retrospective cohort of EHR from 1998-2015 included 974 cases and this time three different control groups: 476 adolescents with NSSI, 7,059 with a history of depression but no self-harm, and 25,081 with no history of self-harm randomly drawn from the general hospital population. Otherwise, the analysis methods were identical to the previous paper. Optimism-adjusted AUC across time windows for NSSI controls were 0.82-0.85, for depressed controls 0.87-0.90, and for general controls 0.94-0.97. For the 90 day window and using general controls, 91% of attempters in were in the top 5% of risk scores. Important variables (roughly in order of importance) included BMI, age, use of anti-inflammatories and SSRIs, history of MDD, history of episodic mood disorders, number of ER visits, history of self-poisoning, use of antipsychotic medication, history of recurrent depression and/or PTSD, use of narcotics, diagnosis of ADD/ADHD, use of benzodiazepines, and gender.

Only one other study has been done in a youth population, and it predicted suicidal ideation but not behavior. Hill and colleagues used data from the National Longitudinal Study of Adolescent to Adult Health (https://addhealth.cpc.unc.edu/), and included youth who had completed waves 1 (data collected 1994-1995) & 2 (data collected 1 year later) (Hill, Oosterhoff, & Kaplow, 2017). 4,799 youth were included, with age ranging 11 to 21 years and a mean of 15 at baseline. The authors used CART to build a single classification tree with variables selected based on a priori theory. Survey variables were collapsed into summary scores where possible, and the final set included 16 predictors plus 7 indicators related to wave 1 suicidal behaviors in self or family/friends. 10-fold cross-validation used to train the tree, and there was no independent validation. Three trees were considered. Sensitivity ranged 44.6-77.6%, specificity 68.2-91.1%, and PPV, while not reported, can be calculated as ranging 23-38%; AUC was not reported. Variable importance can be qualitatively assessed based on order of split in the tree: suicidal ideation at wave 1 was the first split, then depressive symptoms, then family/friend suicidal behavior. Some additional variables appeared in the largest tree (e.g. social support, school absences) but they were deep in the tree so are of questionable validity given the lack of validation and the possibility of overfitting. Additionally, the survey that the authors used collected data on self-harm at both waves, so it is unclear why the authors chose to model only suicidal ideation and not behavior.

The largest prediction study to date was published by Simon and colleagues (Simon et al., 2018). This retrospective cohort study combined EMR data from seven health systems with insurance claims, and state census and mortality data. Two samples were analyzed: patients with a mental health specialty clinic visit, and patients with a primary care visit for a mental health diagnosis between 2009 and 2015. Patients aged 13 years and older were included; only 10% of the mental health specialty sample and 4% of the primary care sample were under age 18. The study included 2,960,929 patients total. The outcomes were suicide attempt and suicide mortality within 90 days of the index visit. (This study is the only one to date to examine suicide mortality in the U.S. civilian population.) Predictors included 313 demographic and clinical characteristics: 149 dichotomous variables coded for three overlapping time points: 90 days, 1 year, 5 years prior to the outcome. LASSO models were built for the two samples, and 10-fold cross-validation was used to set the tuning parameter. A random 35% of the data was used for validation. The AUC for suicide attempt in the specialty sample was 0.85, and in the primary sample 0.85. The AUC for suicide death was 0.86 in the specialty sample, and 0.83 in the primary. In the specialty sample, 48% of deaths and 43% of attempts occurred in the top 5% of risk scores; 43% of death and 48% of attempts in the primary sample. In the specialty sample, PPV for attempts based on cut-points at the 50^{th}-99^{th} percentile of risk scores ranged 1.1-10.4% and for deaths 0.05-0.62%. PPV in the primary sample for attempts ranged 0.5-6.1% and for deaths 0.03-0.31%. The strongest predictors based on effect size included depression and SUD diagnosis, medication for same, inpatient stays, and prior suicide attempt. Most diagnosis variables were more predictive in the 5 years timeframe, suggesting an elevated risk associated with chronic conditions, while prior suicide attempt was more predictive in earlier time windows.

The most recent study as of this writing was also the only one to use data from a nationally representative survey rather than EMR data to examine suicidal behavior. De la Garza and colleagues used data from the first two waves of the National Epidemiological Survey on Alcohol and Related Conditions (NESARC), a nationally representative survey of non-institutionalized, civilian adults 18 years and older (Garcia de la Garza, Blanco, Olfson, & Wall, 2021). Wave 1 data were collected from face-to-face interviews in 2001-2002, and wave 2 in 2004-2005. The sample was 34,653 adults who completed the wave 2 interview (70.2% of the baseline sample). The outcome was self-reported, non-fatal suicide attempt at any point in the 3 years between survey waves 1 and 2. Predictors included past year and lifetime substance use, and mental health diagnoses. A balanced random forest model (bootstrap samples contain equal numbers of each class, attempters and non-attempters, which may improve model performance when class prevalence is highly imbalanced) was constructed, with 10-fold cross-validation used to tune model parameters. AUC and other prediction metrics were based on the aggregated out-of-fold classifications for the final model, not on a separate validation dataset, and were weighted based on the NESARC design and non-response weights so that the final results would be representative of the U.S. population based on the 2000 Census. Relative variable importance measures were calculated (scaled to be from 0 to 100), and risk severity groups were defined. The AUC was 0.857 with sensitivity 85.3% and specificity 73.3%. Based on the model, 73.1% of the U.S. population is estimated to be at low risk for suicide attempt, 17.5% at medium risk, 7.6% at high risk, and 1.8% at very high risk. Depending on the threshold used, PPV ranged from 2.0% to 10.4%, NPV from 99.9% to 99.6%, alarms per 100 evaluation from 27 to 2, and number needed to evaluate to find one new attempter from 51 to 10. The most important predictors included prior suicide attempts and ideation, followed by several symptoms of anhedonia and low mood, and then demographic factors (age, income, education, marriage status) and recent financial crises. Supplemental analyses revealed that model accuracy decreased with increasing time to suicide attempt, was less accurate for younger people, and was less accurate for non-White people. Using only the top 5 and 10 most important variables, models were constructed with AUCs of 0.818 and 0.845, respectively.

Other Countries

Studies predicting suicide and self-harm have to date mostly come from the U.S., though one study each have come from the United Kingdom (Wales), Denmark, Australia (two studies but coming from the same sample), and South Korea. Studies from developing and non-Western countries are lacking. One study using machine learning to predict suicide in Iran has been published, but it is unclear from the methods whether the data are longitudinal or all come from a single cross-sectional survey, and so this paper is not described in detail (Amini, Ahmadinia, Poorolajal, & Moqaddasi Amiri, 2016).

DelPozo-Banos and colleagues conducted a case-control study in Wales using EMR data from a centralized anonymous database plus national death statistics (DelPozo-Banos et al., 2018). Cases were 2,604 suicide deaths occurring from 2001 to 2015, and 20 matched controls per case. The age distribution of the sample was not described, but the mean age was 48 years. The authors used an artificial neural network with 10-fold cross-validation but no independent validation data. Variables (healthcare visits) were defined based on primary complaint, medication prescription, and hospital admission at four non-overlapping time frames: 1 month, 6 months, 1 year, and 5 years, for a total of 60 predictors (15 factors * 4 times). The best performing model had an AUC of 0.80, sensitivity 64.5% and specificity 81.8%; PPV was not reported. Variable importance was expressed semi-quantitatively as the distribution of risk scores within a factor. Self-harm and drug/alcohol abuse had the highest proportion of high/very high risk scores, and results did not vary substantially over time frames.

Gradus and colleagues conducted a case-cohort study using Danish registry data from 1995-2015 (Gradus et al., 2020). The sample included all suicide deaths and a random 5% sample of living individuals at start of cohort. The age distribution of the sample was not described, but the mean age was 38 years (SD 18.8). Variables from numerous domains were dummy coded to be time-varying (0-6, 0-12, 0-24, and 0-48 months prior to event). In all, 2554 variables culled to 1339 by dropping rare predictors (<11 observations), predictors with unadjusted odds ratios 0.9-1.1, and diagnoses occurring in the ED. The authors used two methods. CART with 10-fold cross-validation to train the tree, and random forest built with all cases and a random sample of equal N controls for each tree, split-half cross-validation to calculate the error rate, and mean decrease in accuracy after permutation for measuring variable importance. Models were stratified by sex, and there was no independent validation. The AUC for the single classification tree was 0.77 for men and 0.87 for women. The highest risk groups for both men and women are those with a prior suicide attempt or prior poisoning and who were not receiving psychiatric pharmacotherapy. The random forest AUC was 0.80 for men and 0.88 for women. 90-91% of the variables had a positive variable importance value. (Note that there are now methods for testing whether importance measures are statistically significantly different from zero.) The top predictors across both folds were often consistent, but sometimes there were wide discrepancies, possibly a result of using split-half cross-validation not on the full sample. Top predictors included various psychiatric medications, SUD/BP/MDD diagnoses at 48 months, prior self-harm/poisoning, various physical injuries (which again may be unrecognized suicide attempts), early retirement for men but not women. Factors at 48 and 24 months tended to be more predictive, similar to results observed in the Simon study in the U.S., possibly indicating the importance of chronicity since timeframes do overlap in this study.

Two studies were published in Australia using the same EMR data from one tertiary care hospital. The first was a retrospective cohort study by Tran and colleagues, enrolling 7,399 inpatient and ED patients who had at least one suicide risk assessment from 2009 to 2012 (Tran et al., 2014). For patients with multiple risk assessments, one visit was randomly selected for inclusion. Patients were aged over 10 years, and only 16% of the sample was under 21 years. The outcome was suicide attempt 30, 60, 90, or 180 days following assessment, and this was treated as 3-level ordinal variable: high lethality attempt, moderate lethality, and low lethality/no attempt. 202 demographic and clinical predictors from five non-overlapping time points up to 48 months prior were included. The authors used an L1-penalized continuation-ratio model, which is essentially LASSO for an ordinal outcome, and used a random one third of the data validation. Variable importance was measured by using the bootstrap to calculate empirical distributions for the model coefficients (effect sizes). The AUC ranged 0.71-0.79 across time frames, with no apparent trend between times, and no difference in predicting high lethality vs rest, and moderate/high lethality vs rest. The model consistently performed better than clinician risk assessment. The best predictors of high lethality attempt were prior high lethality attempts, prior other injuries (which, again, may be misclassified attempts), and male sex. Predictors of moderate lethality attempt were prior ED visit, number of psychiatric diagnoses, emotional distress, and depression or bipolar diagnosis.

The second study, published by Karmakar and colleagues, used the exact same data set but a different analysis method (Karmakar, Luo, Tran, Berk, & Venkatesh, 2016). The authors had a specific interest in using physical illness as a predictor of suicide attempt. The authors constructed probability tables based on the number of ICD-10 codes a patient had recorded over five time windows which ranged from 3 to 48 months. ICD codes were counted by chapter of the ICD-10, excluding chapters 5 (mental and behavioral disorders), 16 and 17 (congenital and perinatal disorders), and merging chapters 7 and 8 (diseases of the eye and ear). A probability of suicide attempt was then empirically calculated for every possible combination of physical illness counts over time, using 10-fold cross-validation to estimate the final probabilities. There was no independent validation. The AUC of the table was 0.71, and it outperformed the hospital's clinical risk score.

The final study, by Ryu and colleagues, used data from a nationally representative annual survey of noninstitutionalized South Koreas over age 19 in 2007-2012 (Ryu, Lee, Lee, & Park, 2018). All 5,814 individuals with suicidal ideation and an equal number without, chosen randomly from the full survey of approximately 35,000, were included. The authors excluded those who were missing the ideation question (8% of the total sample), though used MICE to multiply impute missing predictor data. Forty-seven predictors were chosen based on theory, and backwards selection was used to cull predictors down to 15. The authors then used random forest analysis, although specifics of the model were not described except to note that 10-fold cross-validation was used to train the model, and a random 10% of the data used for validation. The model had an AUC of 0.85, and reported sensitivity of 77% and specificity 79%. The model also had a PPV of 46% when applied to the full survey population, although this is likely inflated since only some of the non-suicide ideators were independent of model training. Top variables included depressed mood, measures of stress, anxiety, general health status, sex, and participation in employment and social activities.

PRIOR STUDIES USING NSCAW DATA TO EXAMINE SELF-HARM

A handful of studies have been published examining self-harm in the NSCAW I and II cohorts. The first was a descriptive analysis of the NSCAW I baseline data published by Leslie and colleagues in 2010 (Leslie et al., 2010). The study included children aged 11 to 15 years (unlike the NSCAW II, which recruited children <18 years, the NSCAW I only recruited children through age 15). Self-harm, what the authors called "suicidality," was coded the same as in the current study (a dichotomous indicator of responses greater than Never on *either* the YSR or CBCL item 18, relating to self-harm) and was one of the nine outcomes under the umbrella of "health-risk behaviors" that the authors were assessing. In addition to reporting population prevalence estimates of self-harm, the authors also reported associations with a variety of theoretically presumed risk and protective factors. The prevalence of self-harm was 7.9%. Prevalence did not vary significantly by child demographics, alleged abuse type, out-of-home placement; or measures of youth behavioral or cognitive functioning, religiosity, or school engagement. There were qualitative differences in prevalence across some variables, but the authors did not provide confidence intervals or specific p-values, so it is difficult to assess whether or not there is an issue of low power in this sample, which included only 64 cases of self-harm. Self-harm was significantly less prevalent when caregivers reported higher than a high school education (4.8% vs 8.6-9.8%), and children who reported self-harm also had significantly lower scores on measures of caregiver monitoring (3.5 vs 4.1) and expectations for the future (3.5 vs 4.0). Multivariate analyses of self-harm were not conducted in this study.

Heneghan and colleagues conducted a similar descriptive analysis of NSCAW II baseline data in 2013 (Heneghan et al., 2013). They assessed prevalence of "suicidality" using the same case definition as Leslie et al. in children aged 12 to 17 years, along with four other mental health-related outcomes. The prevalence of self-harm was 13.9% overall. In univariate analyses, self-harm was significantly more prevalent among females than males (19.0% vs 6.4%), children 12-14 years old than those 15-17 (15.9% vs 11.3%), and significantly less prevalent among Black children than other races (5.0% vs 12.9-23.3%). Differences by alleged abuse type, out-of-home placement, prior maltreatment, chronic health conditions, and caregiver depression were not significant. In multivariate analysis, the variables associated with self-harm were female sex (odds ratio, 95% CI: 2.38, 0.99-5.56), Black race (0.27, 0.09-0.76), and alleged physical or sexual maltreatment compared to neglect or other reasons for CPS involvement (2.27, 0.96-5.26; and 3.23, 1.56-6.67, respectively). Heneghan and colleagues also published a study comparing the prevalence of "suicidality" and other health-risk behaviors between the NSCAW I, II, and the 2011 Youth Risk Behavior Surveillance (YRBS) survey, a nationally representative survey of U.S. high school students (Heneghan et al., 2015). The prevalence estimates and risk factor associations were reported in the prior two papers, but this analysis does add one additional risk factor that was not examined previously. Self-harm was more prevalent among children with scores on the CBCL ≥64 (29.0% vs 7.3%). Self-harm was more prevalent in the NSCAW II than the YRBS, 11.3% (95% CI: 6.5-19.0%) versus 7.8% (7.1-8.5%).

Few studies of self-harm have been conducted using the full longitudinal NSCAW data, and all studies to date have looked at suicidal ideation (SI) but not behavior. Both NSCAW I and II measured suicidal ideation using a single item on the Child Depression Inventory that asks youth about thoughts of suicide in the past 2 weeks. Anderson used data from the NSCAW I to assess the association between out-of-home (OOH) placement and SI reported at the last wave of data collection, and examine the possible mediating role of depressive symptoms (Anderson, 2011). Using a series of multivariate logistic regression models, Anderson showed that OOH placement (measured as a count of the number of survey waves in which a child reported currently being OOH) was associated with SI at the last survey wave while controlling for prior

SI (SI at any wave before the last) (odds ratio 1.68, 1.08-2.62). OOH placement was also associated with depressive symptoms, measured as a standardized CDI score >65 at any of the first four survey waves (OR 1.62, 95% CI not reported but $p < 0.05$). When depressive symptoms was added to the model of OOH placement on subsequent SI, the effect of OOH placement was attenuated (OR 1.48, 0.86-2.56). The author concludes that there is evidence that time spent in OOH placement increases the risk of subsequent suicidal ideation by increasing depressive symptoms.

Two studies have been published examining suicidal ideation across multiple waves of the NSCAW II. Sellers and colleagues examined the associations between SI and substance use, deviant peer association, and caregiver health (Sellers, McRoy, & O'Brien, 2019). The authors used a mixed effects model to analyze all three waves of data simultaneously, but the associations examined were cross-sectional. That is, the authors modeled SI as a function of the covariates at the same wave, they simply included all three waves and controlled for intra-individual correlation with random intercepts, so temporality cannot be gauged in this study. The authors found that SI was associated with alcohol use (quite strongly) and with greater deviant peer associations, but not with marijuana use nor any measure of caregiver health, including caregiver substance use. However, in their analysis the authors did not properly account for the complex survey design of the NSCAW. In fact, they did not even mention that the NSCAW was a complex survey. This serious limitation undermines the validity of the standard errors in the models, and thus the conclusions that can be drawn. For this reason, the specific results are not reported here.

The last study, by Fulginiti and colleagues, examined suicidal ideation and measures of social connectedness in the baseline and first follow-up wave of the NSCAW II (Fulginiti, He, & Negriff, 2018). Three measures of connectedness, treated as continuous variables, were examined: connectedness with caregivers, peers, and school. The authors used a survey-weighted, cross-lagged path model to model the reciprocal effects of connectedness and SI at baseline with each other at follow-up (i.e. the full path model was fit). Controlling for demographics, OOH placement, and a few other constructs theoretically related to SI, the model showed that SI at baseline reduced caregiver and peer – but not school – connectedness at follow-up, but no measures of connectedness significantly impacted SI at follow-up. Effect sizes were relatively small: standardized coefficients for peer connectedness -0.17 and caregiver connectedness -0.12. Model results did not vary significantly by gender. The authors conclude that interventions to bolster social connectedness for youth may be needed in the *aftermath* of a suicidal event not just, as is commonly stated in the literature, as a primary preventive measure.

Chapter 3

Self-harm over time among adolescents with Child Protective Services (CPS) contact

ABSTRACT

Purpose: To examine the prevalence of self-harm over a three-year period among adolescents following an investigation by Child Protective Services (CPS) into alleged maltreatment, and to assess whether self-harm prevalence differed by demographic factors and characteristics of the event that precipitated the CPS investigation.

Methods: Data came from the second National Survey of Child and Adolescent Well-being cohort (NSCAW II), a nationally representative longitudinal survey. Descriptive statistics at baseline, 0-4 months following the CPS investigation, and follow-up waves at 18 and 36 months were generated. Multivariable logistic regression accounting for the complex survey design was used to assess differences in the odds of self-harm at follow-up by demographic characteristics, baseline maltreatment type, and any out-of-home placement. Multiple imputation of missing data was used to minimize possible selective non-reporting in the follow-up waves.

Results: The overall prevalence of self-harm at any point in the study was 16%, and 5% of adolescents reported self-harm at multiple survey waves. The wave-specific prevalence among youth aged 15-17 years remained stable over time at ~10% while among youth 11-14 years it declined over successive waves from 13% to 6% to 3.5%; among young adults 18+ the prevalence at wave 3 was 6%. At baseline, females and children suffering physical abuse had greater odds of reporting self-harm (odds ratio and 95% confidence interval: 2.95, 1.46-5.97 for females vs males; 3.50, 1.48-8.30 for physical abuse vs neglect) and Black non-Hispanic youth had lower odds (0.39, 0.16-0.95 vs White non-Hispanic), but these differences did not persist over follow-up. Native American and Asian/Pacific Islander youth were not more likely to report

self-harm at any particular wave, but were significantly more likely to report self-harm at multiple waves (5.25, 1.43-19.2 compared to White non-Hispanic).

Conclusions: Prior research reported that younger adolescents with CPS contact had greater odds of self-harm than older adolescents, but this was true only at baseline. There was no evidence that out-of-home placement was associated with an increased risk for self-harm, controlling for maltreatment. Further research is warranted to determine the reasons why Native American and Asian adolescents experienced more persistent self-harm, and potentially to design culturally appropriate interventions to interrupt these behaviors.

INTRODUCTION

Suicide is the tenth leading cause of death in the U.S., and is the second leading cause of death among adolescents and young adults aged 10-24 years of age (CDC, 2003). Surveillance of U.S. high school students indicates that suicidal ideation and suicide attempts have increased over the past decade, while most other youth risk behaviors have declined; including illicit substance use, dating violence, and early sexual activity (CDC, 2020). The causes of suicidal behavior in youth are still not fully understood, but it is recognized that youth who suffer abuse and/or neglect at home are at significantly increased risk. A systematic review and meta-analysis of the association between abuse/neglect before age 18, and suicidal thoughts and behavior through age 24 found that all types of maltreatment were significantly associated with suicide attempt, with pooled odds ratios for different types ranging from 1.8 (95% CI 1.3-2.5) for physical neglect to 3.4 (2.9-4.0) for sexual abuse (Angelakis et al., 2020). The authors concluded that, despite some heterogeneity in effect sizes, significant associations between abuse/neglect and suicidality were robust across 79 included studies of over 250,000 total participants. However, they also noted that most of the included studies used a cross-sectional design, and that longitudinal studies were needed to examine the development of suicidal behaviors over time following maltreatment. The current study aimed to address this gap in the literature.

In the U.S., Child Protective Services (CPS) is the state government agency responsible for responding to allegations of child maltreatment. Recognizing that children with CPS contact represent an especially vulnerable population, the Administration for Children & Families commissioned the National Survey of Child and Adolescent Wellbeing (NSCAW), a nationally representative, longitudinal study of children and their families who had been the subject of a CPS investigation. The two cohorts enrolled (NSCAW I and II) remain the most comprehensive source of data on the health, welfare, and development of children in the CPS system. Self-harm[1] in the NSCAW cohorts has not been extensively studied, although the prevalence of self-harm at baseline in both NSCAW I and II has been assessed. In the NSCAW I, which enrolled children from 1999-2000, the prevalence of self-harm at baseline was 8% and did not vary by demographic factors, type of maltreatment, or whether the child was removed from the home and placed with other relatives or in foster care (out-of-home placement, OOH) (Leslie et al., 2010). Results were different in the NSCAW II, which enrolled children from 2008-2009 and had a higher baseline prevalence, 14%, which differed across some key variables (Heneghan et al., 2013). Self-harm was significantly higher among females than males, younger adolescents 12-14 years than those 15-17, and significantly lower among African American youth than youth of other races. In multivariate analyses the sex and race differences persisted, plus children with alleged physical/sexual maltreatment had significantly higher odds of self-harm than other abuse types or neglect. However, no study to date has examined the prevalence and correlates of self-harm longitudinally in the NSCAW cohorts, and so the current study was undertaken to assess how the prevalence of self-harm changed over time in the most recent NSCAW cohort,

[1] We use the term “self-harm” rather than “suicidality” although the latter has been used in other NSCAW studies (Heneghan et al., 2013) because the survey questions capture self-injurious behavior but do not confirm an intent to die (see Methods). Therefore, we feel that “self-harm” is the more accurate term and use it throughout.

and to examine demographic and maltreatment-related correlates of self-harm up to 3 years following an initial CPS investigation.

METHODS

Data came from the National Survey of Child and Adolescent Wellbeing II (NSCAW II). The NSCAW II was a 3-year longitudinal study of youth <18 years of age and their families who had been the subject of an investigation by Child Protective Services (CPS) for alleged maltreatment between February 2008 and April 2009 (Dowd et al., 2014). Baseline interviews with children and their primary caregiver were conducted within approximately four months of the conclusion of the CPS investigation, and families were enrolled regardless of the investigation outcome. Follow-up interviews were conducted at approximately 18- and 36-months post-baseline. Children remained in the study even if they were older than 18 years at follow-up; however, caregivers were only interviewed if the child was under 18. The current study includes all three waves of data.

The NSCAW II followed a two-stage stratified sampling design. In the first stage, the U.S. was divided into nine strata corresponding to the eight states with the largest CPS caseloads and the remaining 42 states and the District of Columbia. Within these strata, primary sampling units (PSUs) were defined and selected. The PSU was defined as the geographic area containing a population served by a single child welfare agency; typically this was a single county. Eighty-one PSUs were selected (71 of the 92 PSUs that were included in NSCAW I, plus an additional 10 PSUs to replace those that declined to participate in the NSCAW II). The sample was weighted to be representative of the overall U.S. child welfare population.

The outcome of interest was child- or caregiver-reported self-harm, measured at each wave by a single item on the Youth Self-Report and the Child Behavior Checklist for youth and caregivers, respectively. Both instruments are part of the ASEBA, a suite of psychometric measurement instruments that have been well validated and widely used in child research

(Achenbach, 2009). The text of the question is similar between the YSR and CBCL, asks about behavior in the previous 6 months, and reads:

"I deliberately try to hurt or kill myself" -YSR

"My child deliberately harms his/herself or attempts suicide" -CBCL

Both items are scored on a 3-point Likert scale, with possible responses *Never True, Somewhat or Sometimes True, or Very or Often True*. For the analysis, responses were dichotomized with Somewhat and Very collapsed into a single, positive indicator of self-harm. For the multivariate analysis (described below), children were considered positive for self-harm if **either** the youth or caregiver response was positive. For the descriptive analysis child and caregiver responses were also examined separately to assess agreement between them.

The weighted prevalence of self-harm and Wald 95% confidence intervals were estimated for each of the three survey waves and for the study period overall (i.e. self-harm at any or several of the three waves). The analytic sample included all observations from children 11 and older, so the sample size for each successive wave increased as children aged into the analytic sample. Descriptive results were stratified by age: 11-14 years, 15-17 years, and 18+ years. Suicidal behavior is known to vary by sex, age, and race (Bridge et al., 2006), and a recent study found that youth with CPS contact in Ohio who died by suicide were significantly more likely to have had out-of-home placement (Ruch et al., 2021), although an association between out-of-home placement and self-harm was not seen using NSCAW II baseline data (Heneghan et al., 2013). Therefore, survey-weighted, multivariate logistic regression models were run to evaluate whether the odds of self-harm varied by demographic factors or out-of-home placement, controlling for abuse severity. Three analyses were conducted. In the first, cross-sectional models were fit separately for each wave, including the following variables: child sex (male/female), race/ethnicity (collapsed into White non-Hispanic, Black non-Hispanic, Hispanic, or Other), age (three categories described above), primary alleged maltreatment type (collapsed into physical abuse, sexual abuse, other abuse, neglect, or request for service(s) access), and whether the child was currently in out-of-home placement, e.g. foster care, group home. Next, a model was fit for the odds of reported self-harm at either follow-up wave and included the following covariates: sex, race/ethnicity, age at baseline, most severe maltreatment type prior to self-harm or wave 3 (whichever came first), any out-of-home placement prior to self-harm or wave 3 (whichever came first), and self-harm at baseline. A final model was fit modeling odds of reported self-harm at multiple waves, including all covariates previously described, except self-harm at baseline, and limited to youth with at least two waves of data.

Prior to data analysis, missing data were imputed. Child's age, sex, and race/ethnicity were imputed deterministically, with sex and race/ethnicity set as the mode across waves and age equal to the age at the prior wave plus two years (age was not missing from any child at baseline). Self-harm, primary maltreatment type at baseline, and out-of-home placement were multiply imputed using chained equations, with 15 imputed datasets created. Maltreatment at follow-up was missing for >80% of cases and so was not imputed; for all multiply imputed variables the percent missing varied from 0.9% to 18.1%. Chi-squared tests were used to examine whether the prevalence of self-harm varied between observed and imputed observations for each survey wave. All analyses were conducted using R version 3.6.2 (R Core Team, 2021); the 'mice' package was used for imputation, the 'survey' package was used to generate descriptive statistics and regression models accounting for the survey design for each imputed dataset, and the 'mitools' package was used to combine estimates from the different imputations.

All procedures for the NSCAW II were approved by the Research Triangle Institute's Institutional Review Board and all analytic work with the NSCAW II Restricted Release dataset was approved by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board.

RESULTS

Prevalence of self-harm

A total of 1,513 youth contributed at least one wave of data to the analysis. Demographic characteristics and alleged maltreatment for the sample are shown in Table 1. Overall, 15% of youth reported self-harm at at least one wave and, of those with two or three waves of data (n = 1,278), 5% self-harmed at multiple waves. Wave-specific prevalence of self-harm for the baseline and follow-up waves were 12%, 7%, and 6%, respectively. A substantial minority of youth with self-harm at follow-up had prior reported self-harm; 38% of youth who self-harmed at wave 2 had also self-harmed at baseline, and 44% of youth who self-harmed at wave 3 had self-harmed at a prior wave. The prevalence of self-harm over time differed by age (Table 2). Among youth 15-17 years at each wave, the prevalence was nearly constant at ~10% but among youth 11-14 years, the prevalence declined over successive waves from 13% to 6% to 3%. Among youth 18 years and older, the prevalence at wave 2 was <1% (possibly influenced by small sample size, n = 118) and at wave 3 was 6%. Concordance between child and caregiver reports of self-harm was low, averaging 11% (percent positive agreement). Agreement varied by child age and survey wave, but with no discernable trends (Table 2).

Associations with covariates

When stratifying by survey wave (Table 3), females and children suffering physical abuse had significantly higher odds of self-harm at baseline, with odds ratios (OR) and 95% confidence intervals 3.0, 1.5-6.1 and 4.0, 1.7-9.9, respectively; Black children had significantly lower odds of self-harm at baseline (OR 0.4, 0.2-0.9). However, there were no significant differences by sex, race/ethnicity, or abuse type at waves 2 and 3. Youth 15-17 had increasingly higher odds compared to youth 11-14 at waves 2 and 3, significant at wave 3 (OR 3.1, 1.3-7.4), which matches the trend noted above that the prevalence of self-harm among older youth remained constant while among younger children it declined with each successive wave. In the longitudinal model, self-harm at either follow-up wave did not differ significantly by demographic characteristics, type of prior maltreatment, or home placement (Table 4). Self-harm at baseline was strongly associated with self-harm at follow-up (OR 6.5, 3.2-13.3). Youth suffering physical abuse and Other race youth had significantly greater odds of self-harm at multiple waves (3.4, 1.2-9.7, and 6.9, 2.0-23.5, respectively; Table 5).

Table 1. Sample characteristics (weighted proportions and unweighted N)

Total	1573
Child sex	
Male	46.7% (762)
Female	53.3% (811)
Child race	
White, non-Hispanic	45.0% (630)
Black, non-Hispanic	21.5% (458)
Hispanic	28.2% (386)
Other	5.3% (98)
Child age at baseline	
<11 years	34.0% (519)
11-14 years	42.9% (657)
15-17 years	23.1% (397)
Primary maltreatment type	
Physical maltreatment	21.7% (314)
Sexual maltreatment	9.2% (170)
Other maltreatment	6.0% (100)
Neglect	47.5% (731)
Service access	1.5% (25)
Unknown	14.0% (233)
Primary abuse substantiated	
Yes	22.3% (788)
No	73.7% (744)
Unknown	4.0% (41)
Additional investigation(s) at follow-up	
Yes	7.4% (158)
No	13.0% (446)
Unknown	79.6% (969)
Out-of-home placement at any wave	
Yes	20.9% (615)
No	79.1% (958)
Self-harm reported at any wave*	
Yes	14.7% (138.5)
No	85.3% (1,434.5)
Self-harm reported at multiple waves**	4.5% (51.5)

*Average N across imputed datasets

**Average across imputations and limited to those with at least 2 waves of data, N = 1,278

Table 2. Prevalence of self-harm by wave and age, survey weighted proportion and 95% confidence interval

		Wave 1 (N = 1,054)	Wave 2 (N = 1,290)	Wave 3 (N = 1,513)
Child reported	11-14 years	7.4% (3.2-11.5%)	2.8% (0.6-5.0%)	1.0% (0-2.5%)
	15-17 years	3.0% (1.0-5.1%)	4.8% (1.8-7.7%)	5.3% (0.9-9.6%)
	18+ years	N/A*	0.1% (0-0.2%)	5.8% (1.3-10.2%)
Caregiver reported	11-14 years	7.7% (4.5-10.8%)	4.4% (2.2-6.5%)	2.9% (0.8-5.0%)
	15-17 years	9.7% (4.1-15.3%)	5.2% (1.7-8.8%)	7.4% (2.8-12.0%)
Child or caregiver reported	11-14 years	13.2% (8.7-17.7%)	6.0% (3.4-8.6%)	3.4% (1.2-5.6%)
	15-17 years	11.3% (5.5-17.1%)	9.8% (5.3-14.3%)	9.5% (4.9-14.2%)
Percent positive agreement	11-14 years	11.7% (0.9-22.4%)	15.4% (0-31.6%)	6.1% (0-14.8%)
	15-17 years	10.4% (0.8-19.9%)	1.7% (0-5.2%)	26.9% (0-59.3%)

*All youth were <18 years at the time of enrollment into the NSCAW

Table 3. Survey-weighted odds of self-harm, cross-sectional by survey wave

	Wave 1 (N = 1,053)		Wave 2 (N = 1,147)		Wave 3 (N = 1,308)	
	Odds ratio (95% conf int)	p value	Odds ratio (95% conf int)	p value	Odds ratio (95% conf int)	p value
Sex						
Male	ref	-	ref	-	ref	-
Female	2.97 (1.45-6.08)	0.003	1.84 (0.86-3.96)	0.118	0.63 (0.25-1.57)	0.318
Age						
11-14 years	ref	-	ref	-	ref	-
15-17 years	0.96 (0.50-1.84)	0.893	1.66 (0.84-3.32)	0.147	3.13 (1.33-7.39)	0.009
18+ years	N/A*	-	N/A*	-	1.83 (0.80-4.15)	0.151
Race/ethnicity						
White non-Hispanic	ref	-	ref	-	ref	-
Black non-Hispanic	0.38 (0.16-0.94)	0.036	0.85 (0.31-2.39)	0.764	0.74 (0.23-2.35)	0.610
Hispanic	1.27 (0.61-2.66)	0.524	1.81 (0.76-4.29)	0.178	1.27 (0.49-3.33)	0.623
Other	2.22 (0.76-6.53)	0.147	2.16 (0.51-9.12)	0.295	3.97 (0.91-17.4)	0.068
Primary maltreatment type						
Neglect	ref	-	ref	-	ref	-
Physical maltreatment	4.04 (1.65-9.88)	0.002	1.75 (0.71-4.32)	0.226	1.16 (0.41-3.31)	0.778
Sexual maltreatment	1.37 (0.40-4.63)	0.613	1.48 (0.46-4.72)	0.507	0.67 (0.19-2.33)	0.526
Other maltreatment	1.19 (0.33-4.26)	0.789	0.49 (0.09-2.80)	0.422	0.69 (0.14-3.33)	0.649
Service access	1.84 (0.25-13.5)	0.551	1.20 (0.14-10.6)	0.871	0.36 (0.03-4.24)	0.416
Child living situation						
At home	ref	-	ref	-	ref	-
Out of home placement	1.40 (0.62-3.12)	0.417	1.24 (0.47-3.27)	0.662	0.65 (0.22-1.85)	0.415

*All youth were <18 years at the time of enrollment into the NSCAW

Table 4. Survey-weighted odds of self-harm at any follow-up wave

	Odds ratio (95% confidence interval)	p value
Sex		
Male	ref	-
Female	0.75 (0.32-1.74)	0.506
Age at baseline		
<11 years	ref	-
11-14 years	1.28 (0.51-3.19)	0.604
15-17 years	1.15 (0.48-2.77)	0.752
Race/ethnicity		
White non-Hispanic	ref	-
Black non-Hispanic	0.95 (0.41-2.18)	0.903
Hispanic	1.54 (0.67-3.55)	0.307
Other	3.45 (1.12-10.6)	0.031
Primary maltreatment type		
Neglect	ref	-
Physical maltreatment	0.99 (0.45-2.17)	0.971
Sexual maltreatment	1.15 (0.36-3.66)	0.811
Other maltreatment	0.42 (0.11-1.60)	0.202
Service access	0.75 (0.11-4.98)	0.768
Child living situation		
At home	ref	-
Out of home placement	1.09 (0.53-2.20)	0.821
Self-harm at baseline		
No	ref	-
Yes	6.52 (3.18-13.3)	<0.001

Table 5. Survey-weighted odds of repeated self-harm

	Odds ratio (95% confidence interval)	p value
Sex		
Male	ref	-
Female	1.45 (0.62-3.37)	0.391
Age at baseline		
<11 years	ref	-
11-14 years	5.17 (1.07-25.1)	0.042
15-17 years	8.81 (1.59-48.7)	0.013
Race/ethnicity		
White non-Hispanic	ref	-
Black non-Hispanic	0.55 (0.12-2.45)	0.430
Hispanic	1.72 (0.56-5.30)	0.342
Other	6.88 (2.02-23.5)	0.002
Primary maltreatment type		
Neglect	ref	-
Physical maltreatment	3.44 (1.22-9.73)	0.020
Sexual maltreatment	1.68 (0.46-6.10)	0.432
Other maltreatment	0.90 (0.15-5.56)	0.912
Service access	0.001 (<0.001-14.2)	0.121
Child living situation		
At home	ref	-
Out of home placement	1.49 (0.50-4.41)	0.469

DISCUSSION

This study is the first to examine self-harm using the full longitudinal NSCAW II data, a cohort of children who have had contact with Child Protective Services (CPS). Several findings emerge that would not have been apparent from examining the baseline data – or any cross-sectional data – alone. For example, prior research using the NSCAW II concluded, “Of note, suicidality was higher among younger teens than among older teens,” (Heneghan et al., 2013). While this was true at baseline, the prevalence of self-harm (what Heneghan termed suicidality) among younger teens declined over subsequent survey waves while the prevalence among older teens remained constant until, 3 years after baseline, older youth had significantly higher prevalence of self-harm. Self-harm and suicide attempts are more common among older adolescents in community samples (Bridge et al., 2006), and so the decline in self-harm among younger adolescents in the NSCAW may represent a return to what is normal following the CPS investigation and precipitating abuse. This would suggest that these events convey a short-term increase in risk for self-harm among younger adolescents. It is worth noting that the use of imputation in this study minimizes the chance that selective non-reporting could have accounted for the observed decrease in prevalence among younger teens. In longitudinal studies, individuals who are "sicker" may be more likely to drop out; in this case, children in more chaotic circumstances and at greater risk for self-harm could have been less likely to respond to the follow-up interviews, which would have resulted in a decreasing prevalence of observed self-harm among the sample of children who did respond. When missingness depends only on observed variables, imputation can alleviate this problem. That being said, there was limited evidence of selective non-reporting in the NSCAW. The prevalence of self-harm between the observed and imputed cases was not significantly different at any wave; the largest difference was at wave 3 where the observed (unweighted) prevalence of self-harm was 5.2% and the imputed prevalence averaged across imputations was 10.3% (chi-squared statistic 2.372, $p = 0.13$).

Use of longitudinal data also allows for the assessment of repeated/persistent self-harm. Approximately 15% of children with CPS contact reported self-harm at least once in the 3 years following initial CPS investigation, and 5% of children (37% of those reporting self-harm) reported self-harm at multiple survey waves. These results are similar to what was observed in another high-risk group: youth diagnosed with bipolar disorder. A study that followed youth for 5 years after their diagnosis of bipolar disorder found that 18% of the sample ($n = 413$) made at least 1 suicide attempt during follow-up and 8% of the sample (41% of attempters) made multiple attempts (Goldstein et al., 2012). In the current study, most self-harm reports at follow-up were from children with no prior reports and thus would not have been recognized as self-harming from analysis of the baseline data. Nevertheless, a substantial minority of self-harm reports at follow-up were from adolescents with prior reported self-harm, and the persistence of self-harm in over one-third of children who reported any self-harm underscores the need to identify and intervene with these high-risk youth. Children who suffered physical maltreatment were at significantly greater risk for persistent self-harm compared to other types of maltreatment, though the effect size was similar to the association between physical abuse and self-harm at baseline. More interesting is that children of other races were at moderately but not statistically significantly increased risk of self-harm when the survey waves were analyzed cross-sectionally and longitudinally, but were at a significantly higher risk when the outcome was persistent self-harm. The effect of Other race was even stronger than physical maltreatment. In the NSCAW, two racial groups comprise the Other category: American Indian and Asian/Hawaiian/Pacific Islander. A post-hoc analysis was done parsing Other into these separate groups, with the hypothesis that Native American children might be driving the increased risk of persistent self-harm since it is well known the Native Americans in general suffer the highest rates of suicide mortality of any racial/ethnic group in the U.S. (Leavitt et al., 2018). Surprisingly, the results were similar between Native American and Asian children, with both showing significantly higher odds of persistent self-harm compared to White children (OR 6.52, 1.45-29.3 for Native American and 7.85, 1.38-44.7 for Asian), and elevated but non-significant odds of self-harm at any individual survey wave. These results suggest that both Native American and Asian youth with CPS contact may be especially likely to continue self-harming. Additional research should seek to elucidate the cause(s) of this particular risk.

In addition to its longitudinal nature, this study has several other strengths. The data are nationally representative, and so these results can be generalized to the entire population of children served by CPS, which in the U.S. is substantial: in 2018, CPS received 4.3 million referrals involving 7.8 million children (ACF, 2020). Self-harm was measured by two different respondents, the child themselves and their primary caregiver, which should reduce under-reporting. The low concordance between child and caregiver reports of self-harm has been seen in other studies (Klaus, Mobilio, & King, 2009). In community samples, youth often report higher prevalence of self-harm than their caregivers (Klaus et al., 2009). However, a study of youth in foster care, a population more comparable to the NSCAW, found results similar to the current study: similar overall rates of endorsing self-harm between children and caregivers but low concordance, and with caregivers more likely to endorse when the child did not than the reverse (Gabrielli et al., 2015).

There are also important limitations to consider. The wording of the questions measuring self-harm did not distinguish between suicide attempts and non-suicidal self-injurious behavior. Research has shown these to be distinct phenomena (Huang, Ribeiro, & Franklin, 2020) and so they would ideally be examined separately. Additionally, most cases lacked information on whether there were additional incidents of abuse after the precipitating incident. A post-hoc analysis limited to those cases where the presence or absence of additional CPS reports at follow-up was known (n = 493 for wave 2 and n = 319 for wave 3) found no evidence that additional CPS reports of any kind were associated with greater odds of self-harm at either wave 2 or 3. The degree to which repeated self-harm during follow-up is associated with ongoing maltreatment is an important question that remains to be answered.

This study expands our understanding of self-harm behavior in children with CPS contact, and highlights high-risk sub-populations and periods within this already vulnerable population. Younger adolescents experience a spike in the risk of self-harm shortly following alleged maltreatment and CPS investigation which decreases over time, whereas older adolescents maintain a similarly high level of risk for at least 3 years following the incident, including into young adulthood. Social workers, clinicians, and others interacting with youth in the CPS system may need to tailor services to children of different ages, with older children requiring more sustained emotional support and/or monitoring. Over a third of children who self-harm will do so multiple times, with Native American and Asian children appearing to be at particularly high risk for repeated/persistent self-harm. Research into both the mechanism(s) underlying this risk and culturally appropriate interventions to reduce it are warranted.

REFERENCES

Administration for Children and Families. (2020). Child Maltreatment 2018. Retrieved from https://www.acf.hhs.gov/sites/default/files/cb/cm2018.pdf

Achenbach, T. M. (2009). The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Angelakis, I., Austin, J. L., & Gooding, P. (2020). Association of Childhood Maltreatment With Suicide Behaviors Among Young People: A Systematic Review and Meta-analysis. JAMA Netw Open, 3(8), e2012563. doi:10.1001/jamanetworkopen.2020.12563

Bridge, J. A., Goldstein, T. R., & Brent, D. A. (2006). Adolescent suicide and suicidal behavior. J Child Psychol Psychiatry, 47(3-4), 372-394. doi:10.1111/j.1469-7610.2006.01615.x

Centers for Disease Control and Prevention. (2018). Youth Risk Behavior Survey Data Summary & Trends Report 2007-2017. Retrieved from https://www.cdc.gov/healthyyouth/data/yrbs/pdf/trendsreport.pdf

Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS) [Online]. (2003). National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (producer). Available from: www.cdc.gov/ncipc/wisqars. Accessed July 2020.

Dowd, K., Dolan, M., Smith, K., Orin, D., Keeney, J., Wheeless, S., & Biemer, P. (2014). NSCAW II Combined Waves 1-3 Restricted Release Version Data File User's Manual. Ithaca, NY: National Data Archive on Child Abuse and Neglect, Cornell University.

Gabrielli, J., Hambrick, E. P., Tunno, A. M., Jackson, Y., Spangler, A., & Kanine, R. M. (2015). Longitudinal Assessment of Self-Harm Statements of Youth in Foster Care: Rates, Reporters, and Related Factors. Child Psychiatry Hum Dev, 46(6), 893-902.

Goldstein, T. R., Ha, W., Axelson, D. A., Goldstein, B. I., Liao, F., Gill, M. K., . . . Birmaher, B. (2012). Predictors of prospectively examined suicide attempts among youth with bipolar disorder. Arch Gen Psychiatry, 69(11), 1113-1122. doi:10.1001/archgenpsychiatry.2012.650

Heneghan, A., Stein, R. E., Hurlburt, M. S., Zhang, J., Rolls-Reutz, J., Fisher, E., . . . Horwitz, S. M. (2013). Mental health problems in teens investigated by U.S. child welfare agencies. J Adolesc Health, 52(5), 634-640. doi:10.1016/j.jadohealth.2012.10.269

Huang, X., Ribeiro, J. D., & Franklin, J. C. (2020). The Differences Between Individuals Engaging in Nonsuicidal Self-Injury and Suicide Attempt Are Complex (vs. Complicated or Simple). Front Psychiatry, 11, 239. doi:10.3389/fpsyt.2020.00239

Klaus, N. M., Mobilio, A., & King, C. A. (2009). Parent-adolescent agreement concerning adolescents' suicidal thoughts and behaviors. J Clin Child Adolesc Psychol, 38(2), 245-255. doi:10.1080/15374410802698412

Leavitt, R. A., Ertl, A., Sheats, K., Petrosky, E., Ivey-Stephenson, A., & Fowler, K. A. (2018). Suicides Among American Indian/Alaska Natives - National Violent Death Reporting System, 18 States, 2003-2014. MMWR Morb Mortal Wkly Rep, 67(8), 237-242.

Leslie, L. K., James, S., Monn, A., Kauten, M. C., Zhang, J., & Aarons, G. (2010). Health-risk behaviors in young adolescents in the child welfare system. J Adolesc Health, 47(1), 26-34.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ruch, D. A., Steelesmith, D. L., Warner, L. A., Bridge, J. A., Campo, J. V., & Fontanella, C. A. (2021). Health Services Use by Children in the Welfare System Who Died by Suicide. Pediatrics, 147(4), e2020011585.

Chapter 4

Using a random forest algorithm to predict self-harm among adolescents with Child Protective Services (CPS) contact

ABSTRACT

Purpose: To use a machine learning algorithm to build and validate a predictive model for self-harm in a sample of adolescents who have had contact with Child Protective Services (CPS), and identify variables that are significantly predictive of self-harm.

Methods: Data came from the second National Survey of Child and Adolescent Well-being cohort (NSCAW II). A hold-out random forest as described by Janitza and colleagues (2018) was used to predict future self-harm based on a large (>1,500) set of variables encompassing demographic, family, health, developmental, and environmental information. Area under the ROC curve was used to measure prediction accuracy on an independent validation sub-sample, and p-values for the measures of variable importance generated using the method described by Janitza, et al. were used to distinguish significant predictors.

Results: The final model displayed low-moderate predictive accuracy, with an AUC of 0.72. Prior self-harm was the single strongest predictor, with measures of internalizing problems, suicidal ideation, depression, and taking medication for an emotional or behavioral problem the next strongest predictors. Three-quarters (66/86) of the significant predictors came from just three survey instruments: the Youth Self Report and Child Behavior Checklist, similar instruments administered to the adolescent and their caregiver, respectively, measuring behavior problems, and the Children's Depression Inventory, a child-administered measure of depression symptoms. Other significant predictors included trauma symptoms, parental monitoring and maltreatment at home, running away from home, the presence of supportive adults, substance

abuse, and mental health diagnoses. When youth and caregivers reported on the same construct, the youth's report was almost always a better predictor.

Conclusions: Given the modest prediction accuracy, using a machine learning algorithm such as the one described here to estimate risk of self-harm for individuals within CPS is not recommended at this time. Variables that significantly predicted self-harm in this sample came from a smaller number of domains than expected, with the majority corresponding to measures/symptoms of internalizing behavior problems and depression. These results are generally consistent with the interpersonal-psychological model for suicidal behavior, and point especially to the importance of feelings of burdensomeness as significant risk factors for adolescent self-harm.

INTRODUCTION

Suicide is an important public health problem, and the U.S. has seen rising suicide rates for over a decade. Suicide is the tenth leading cause of death in the U.S., and the second leading cause of death among youth ages 10-24 years of age (CDC, 2003). After a period of stability from 2000 to 2007, the suicide rate among youth 10-24 in the United States increased 57.4% – from 6.8 per 100,000 in 2007 to 10.7 in 2018 (Curtin, 2020). In 2018, there were 6,807 deaths among youth ages 10-24 and an estimated 208,218 non-fatal suicide attempts treated in hospital emergency departments (CDC, 2003). Surveillance of U.S. high school students indicates that suicidal ideation and attempts have increased over time, with the most recent data (2019) showing that 19% of all students have seriously considered suicide and 9% have attempted suicide at least once in the past 12 months (CDC, 2020).

Children who suffer abuse or maltreatment at home are at an increased risk for suicidal behaviors; a recent meta-analysis showed a robust association between suicidal behaviors and childhood maltreatment in children and young adults, with odds ratios ranging from 1.79 (95%

CI: 1.27-2.53) for physical neglect to 3.41 (2.90-4.00) for sexual abuse (Angelakis et al., 2020). In the U.S., it is the responsibility of each state's office of Child Protective Services (CPS) to respond to alleged child maltreatment. Unsurprisingly, research has shown that children under CPS supervision, either in out-of-home foster care or supervised home care, are at greater risk for suicidal behavior compared to the general population, with a recent meta-analysis estimating an odds ratio of 3.89 (3.14-4.83) (Evans et al., 2017). The number of children under CPS supervision in the U.S. is substantial: in 2018, CPS received 4.3 million referrals involving 7.8 million children, 2.4 million (56%) of which required some response from CPS (ACF, 2020). Children with CPS contact, therefore, represent a sizeable, high-risk population for suicidal behaviors and an important target for prevention efforts.

In this study, we used nationally representative data on children and their families who have had contact with CPS to build a predictive model for youth self-harm and to identify key risk factors from a dataset of over 1500 possible predictors. To do this, we used a random forest machine learning (ML) algorithm. Such methods have shown promise to improve prediction over traditional regression, but to date they remain under-utilized to study suicidal behavior, especially in youth (see Current Status of the Field below). Children and families under CPS supervision represent a monitored population, which could facilitate the delivery of preventive interventions to these families. If it is possible to accurately predict youth self-harm, then indicated preventive interventions could be deployed within the CPS system; if not, then efforts should focus rather on universal interventions. In either case, identifying key risk factors can help to develop and tailor interventions effectively. To that end, this is the first study to use new methods for hypothesis testing the importance of predictors within an ML framework (Degenhardt, Seifert, & Szymczak, 2019) which allowed us to quantify not just the relative importance of predictors but their statistical significance as well. This study aimed to assess the predictive accuracy of a machine learning model for self-harm and to identify statistically significant predictors from among a large pool of variables in an understudied but important population: adolescents with CPS contact.

Current Status of the Field

Suicide is also a complex phenomenon to study. In a seminal paper published in 2017, Franklin and colleagues reviewed the extensive literature on risk and protective factors for suicide and concluded that, while hundreds of factors have been observed to be associated with suicidal behavior, our ability to identify who is at highest risk for suicide before they make an attempt – to longitudinally predict suicidal behavior – is scarcely better than chance, and has not improved over the past 50 years of research (Franklin et al., 2017). Prior self-injurious behavior is widely recognized as the strongest single predictor of a future suicide attempt, but Franklin et al. demonstrated that no broad class of predictors (e.g. socio-demographics, internalizing psychopathology, social isolation) was substantially associated with improved predictive accuracy. Part of the challenge is obviously the low base rate of suicide in the general population. But the authors also conclude that most studies have only examined predictors individually and have not accounted for the complex interactions that may occur to lead to suicidal ideation/behavior. Contemporary theoretical models for suicide conceptualize suicidal behavior as the result of a large web of risk and protective factors. The model proposed by Turecki & Brent for suicide risk in the general population posits the existence of distal or predisposing risk factors that can be present from (or even before) birth, such as family history and genetics. These distal factors then lead to developmental or mediating risk factors, like certain personality traits and maladaptive cognitive processing, that can act directly on suicide risk but can also modify an individual's susceptibility to proximal or precipitating risks, like stressful life events or depressive episodes or acute substance use. Proximal risk factors have short-term but potentially dramatic impacts on suicide risk. All of these individual risk factors are posited to act under the umbrella of broader social cohesion and environmental risk factors, like access to lethal means and medical care (Turecki & Brent, 2016). Given the complexity of modeling suicide, Franklin and colleagues recommend the use of newer machine learning (ML) methods which can effectively model complex, non-linear interactions between large numbers of possible predictors, and have been shown to improve predictive accuracy over traditional regression models in other fields.

The research community appears to have taken Franklin's recommendation to heart. In the past few years, numerous studies have been published using a variety of machine learning methods to attempt to improve the prediction of suicide and self-harm. Two new reviews provide important updates to the state of the literature. Belsher and colleagues reviewed all longitudinal, predictive self-harm models regardless of statistical methodology (Belsher et al., 2019). Seventeen cohort studies met the inclusion criteria, evaluating 64 different statistical models in over 14 million participants. Studies came from five countries, although 7 of the 17 studies were conducted solely with U.S. military participants (either active-duty Army or military veterans). The authors note that studies were generally of high quality and with low risk of bias. Overall global classification accuracy was good, with most models predicting suicide mortality and nonfatal attempts having an AUC >0.8. However, only 5 of 11 mortality studies reported a positive predictive value (PPV) >1%, and those that did were generally in unrepresentative population samples. ML methods generally outperformed traditional regression. The authors conclude that machine learning represents an improvement in the field, but that PPV is still too low for algorithms to be recommended for practical use. The other review was conducted by Burke and colleagues, also published in 2019 (Burke et al., 2019). It was a systematic review of studies which specifically used ML to predict self-harm. However, studies did not have to be longitudinal and, in fact, of the 35 studies that met inclusion criteria, only 10 studies were longitudinal, 5 predicting suicidal behavior and 5 ideation. The authors agree with Belsher at al. that ML improved prediction compared to traditional statistical methods, but that there remains substantial room for further improvement. ML studies have often confirmed the importance of known risk factors for self-harm, and have occasionally uncovered new risk factors or, more commonly, distinct sub-groups of high-risk individuals. This despite that fact that the most effective prediction models are often “black box” methods where interpreting the effect of individual predictors is less straightforward.

Of note is the paucity of studies predicting self-harm among youth. Most studies have been limited to adults and, in those conducted in the general population, youth have generally comprised less than 10% of the total sample. Only one study we are aware of has used ML to predict self-harm longitudinally in youth. Walsh and colleagues used electronic medical records (EMR) from a single hospital to conduct a retrospective cohort study of 974 adolescents with a nonfatal suicide attempt from 1998-2015 (Walsh et al., 2018), with three different control groups: 476 adolescents with non-suicidal self-injury (NSSI), 7,059 with a history of depression but no self-harm, and 25,081 with no history of self-harm randomly drawn from the general hospital population. The authors built random forest prediction models but did not use an independent sample to validate them, instead using a statistical method called optimism adjustment to estimate the validated predictive accuracy. The models were able to distinguish suicide attempts from hospital controls with a non-suicidal self-injury with an optimism-adjusted AUC of 0.82-0.85 (depending on the time window); an AUC for depressed controls 0.87-0.90, and for general hospital controls 0.94-0.97. Important variables (roughly in order of importance) included BMI, age, use of anti-inflammatories and SSRIs, history of MDD, history of episodic mood disorders, number of ER visits, history of poisoning self, use of antipsychotic medication, history of recurrent depression and/or PTSD, use of narcotics, diagnosis of ADD/ADHD, use of benzodiazepines, and gender. However, owing to the data source, EMR, the authors were not able to evaluate other constructs known to be associated with self-harm in adolescents, like social isolation, impulsivity, and substance use. And while the authors reported an optimism-adjusted AUC, they did not validate their model against an independent sample.

The current study seeks to fill a gap in the literature by using machine learning methods and a dataset with information on a wide variety of health, social, and behavioral domains to longitudinally predict self-harm in an adolescent population.

METHODS

Data Source

Data came from the National Survey of Child and Adolescent Wellbeing II (NSCAW II). The NSCAW II was a 3-year longitudinal study of youth <18 years of age and their families who had been the subject of an investigation by Child Protective Services (CPS) for alleged maltreatment between February 2008 and April 2009 (Dowd et al., 2014). Baseline interviews with children and their primary caregiver were conducted within approximately 4 months of the conclusion of the CPS investigation, and families were enrolled regardless of the investigation outcome. Follow-up interviews were conducted at approximately 18- and 36-months post-baseline. A single interview was also conducted with the CPS caseworker who led the investigation, and with the child's schoolteacher. However, due to substantial missingness in the data, the teacher survey was not used in this analysis.

The NSCAW II followed a two-stage stratified sampling design. In the first stage, the U.S. was divided into nine strata corresponding to the eight states with largest CPS caseloads and the remaining 42 states and the District of Colombia. Within these strata primary sampling units (PSUs) were defined and selected. The PSU was defined as the geographic area containing a population served by a single child welfare agency; typically this was a single county. Eighty-one PSUs were selected (71 of the 92 PSUs that were included in the first NSCAW study, plus 10 additional PSUs to replace those that declined to participate in the NSCAW II). The sample was weighted to be representative of the overall U.S. child welfare population. The current study limited the analytic sample to children ages 11-17 years (N =

1,296), as data on self-harm were not collected from younger children, and older youth (legal adults) were asked a very different set of survey questions.

All procedures for the NSCAW II were approved by the Research Triangle Institute's Institutional Review Board and all analytic work with the NSCAW II Restricted Release dataset was approved by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board.

Variables

The outcome of interest was child- or caregiver-reported self-harm at survey waves 2 and 3. Self-harm was measured by a single item on the Youth Self-Report and the Child Behavior Checklist for youth and caregivers (asking about the child's behavior), respectively, dichotomized into *Any* or *No* self-harm in the prior 6 months. Both instruments are part of the ASEBA, a suite of psychometric measurement instruments which have been well-validated and widely used in child research (Achenbach, 2009). Children were considered positive for self-harm if either the youth or caregiver response was positive. Variables (described below) measured at waves 1 and 2 were used to predict self-harm at waves 2 and 3, respectively. Children who were 11 years or older at baseline could therefore contribute 2 observations to the analysis, while children who were younger than 11 at baseline but 11 or older at wave 2 contributed 1 observation.

The NSCAW contains over 10,000 variables collected at each wave; details about all NSCAW variables are available in the NSCAW Data File User's Manual (Dowd et al., 2014). Briefly, information was collected from the following domains: demographic, family structure, and socio-economic characteristics of the child and caregiver; child's and caregiver's physical and mental health history; relationship between child and caregiver; child's cognitive and scholastic performance; child's mental, emotional, and social development; child's peer relationships; child's engagement in problem behaviors, and both risky and prosocial activities; and details of the CPS investigation, including case outcome, service referrals from CPS, and characteristics of the investigating CPS agency. Most variables were measured at all 3 survey waves, with a few exceptions, e.g., CPS agency characteristics, certain cognitive tests.

Effort was made to be as inclusive as possible when defining the set of predictors. Variables were dropped only if they: were not asked of children 11-17; were asked only of a subset of children (e.g. those in foster care); were redundant with other variables (e.g. raw and standardized scores; however, summary scores and their component individual items were both included); were captured more efficiently in existing derived variables (e.g. specific court findings related to the alleged abuse were dropped in favor of a summary variable indicating whether the abuse was substantiated or not); captured a level of detail that was considered too granular for the current analysis (e.g. each specific date that a child's living arrangement changed); or were missing from >60% of cases. This led to 1569 predictors ultimately being included in the analysis; see Appendix A for a list of included variables.

Imputation of missing data

When child's age was missing, it was imputed as 2 years from the age at prior wave (age was not missing at baseline for any child). Sex and race/ethnicity were copied forward from prior waves. For sex and race/ethnicity, there were occasional discrepancies in some waves and, in these cases, the mode value for the individual was imputed to all waves. For all other variables, missing data were multiply imputed using chained equations (MICE) (van Buuren & Groothuis-Oudshoorn, 2011). Multiple imputation has been shown to improve the validity of statistical inferences when data are missing at random and has been shown to work well in analyses using tree-based machine learning methods (Valdiviezo & Aelst, 2015). MICE is a common imputation method particularly valued for its capacity to impute multiple variables simultaneously, and it has been shown to function well for imputing longitudinal data. When the dataset is balanced and the interval between observations is the same, it is recommended to treat observations at each time point as independent variables and to use observations of the same variable across time points to predict each other (Huque, Carlin, Simpson, & Lee, 2018).

MICE was implemented using the 'mice' package in R version 3.6.2 (R Core team, 2021). For all variables except summary scores and other derived variables, the default prediction methods were used: predictive mean matching for continuous variables, logistic regression for dichotomous variables, proportional odds regression for ordinal variables (including all Likert scale variables), and polytomous regression for unordered categorical variables. Imputation models did **not** account for complex design effects nor did they include survey weighting. In general, variables were modeled as a function of the same variable at both other waves, child- or caregiver-reported self-harm, and a set of demographic variables. Predictors of child self-reported and caregiver-reported child self-harm were chosen through forward variable selection of all child and caregiver variables, respectively. Passive imputation was used to impute summary scores and other derived variables based on the imputed values of the component variables. Models were iterated eight times, and 20 datasets were generated.

Two survey instruments were either not administered to all children in the analytic sample (Independent Living, administered only to ages 14 years and older) or different versions were administered within the sample (Vineland Adaptive Behavior Scale, different versions administered to children 11-12 years vs 13+ years old). For these instruments, an out-of-age dummy value was added, and all variables were treated as categorical in the predictive models.

Random forest for prediction and variable selection

The goal of the analysis was two-fold: to generate and evaluate the accuracy of a statistical model for predicting future self-harm, and to identify which variables were most predictive of self-harm. Prediction models were based on the random forest classifier initially described by Breiman (Breiman, 2001). Random forests are well-suited to this analysis; they can easily incorporate large numbers of predictors, and the decision trees that make up the forest intrinsically model complex multi-way interactions between predictors. As noted above, suicide and self-harm are likely the result of a complex web of risk and protective factors, making the flexibility of random forests an appropriate method for modeling this outcome from the large number of predictors in the NSCAW. The specific random forest method used was an adaptation of the Breiman forest, a so-called "hold-out" forest as described by Janitza et al. (Janitza, Celik, & Boulesteix, 2018). In a hold-out forest, rather than bootstrapping the training sample for each tree and calculating an out-of-bag error rate and variable importance measure, the dataset is split in half with the tree grown on one half and validated against the second half (also known as 2-fold cross-validation). The cross-validation approach avoids potential bias in the variable importance measure that can be induced by bootstrapping (Strobl, Boulesteix, Zeileis, & Hothorn, 2007). The training data was split randomly by *person*, not by observation, following the method described by Karpievitch, et al. for applying random forests to cluster-correlated data (Karpievitch, Hill, Leclerc, Dabney, & Almeida, 2009). Splitting the sample (or bootstrapping, in the case of traditional RF) at the cluster level helps to preserve the independence of trees within the forest, by ensuring that (correlated) observations from the same person are never present in both the training and validation data sets. 2000 trees were grown using the Gini index for splitting nodes; number of candidate predictors considered at each split was the square root of the total (40). Prior to growing the forests, one-third of the data was set aside for model validation, selected randomly but preserving the overall prevalence of the outcome. Results reported are based on the validation data. Sensitivity, specificity, positive and negative predictive values were calculated for a range of cut-offs; cut-off being the minimum proportion of trees in the forest in which an observation was classified as positive in order for the final classification to be positive; the area under the ROC curve (AUC) was also calculated. Random forests were grown using the 'ranger' package in R.

Variable importance is the measure of how predictive a given variable is, and is quantified in a random forest by taking all trees that include a given variable and randomly permuting[1] the variable, then taking the difference in prediction accuracy between the true and permuted trees. Variables that are important predictors will show a significant decrease in accuracy when those variables are permuted, while variables which are unimportant will show little to no change. Recently, several methods were developed to calculate p-values for the variable importance scores. In this analysis, we used a method called the Vita method for calculating variable importance and associated p-values. The Vita method for variable selection was preferred because it has been shown to be one of the most powerful, stable variable selection methods currently available, and is the least computationally intensive among comparable methods (Degenhardt et al., 2019). It works by using the negative importance values generated by the forest to construct an empirical null distribution, from which p-values for the importance measures can then be calculated (Janitza et al., 2018). In order to implement the Vita method for calculating variable importance, random forests were constructed that differed from those used for prediction in two ways. First, the full dataset was used with nothing held out for validation. Second, the bias-corrected Gini index described by Sandri & Zuccolotto was used to split the nodes, rather than a standard Gini index (Sandri & Zuccolotto, 2008). The bias-corrected Gini index is required to implement the Vita method, but is not recommended for prediction.

The procedures described above were repeated separately for each of the imputed datasets. Final variable importance measures and prediction statistics were obtained by averaging the individual results; variables which were significant ($p<.05$) in 19 or more of the 20 datasets, **or** were significant in 18 or more datasets and also had an average importance measure in the 95th percentile were considered significant.

[1] In random permutation, all the values for a given variable are assigned to individuals/observations at random. Thus, the distribution of values for the variable is identical between the original and permuted datasets but, in the new dataset, the permuted variable is not associated with any other variables except by random chance.

RESULTS

The predictive models, averaged across all 20 imputed datasets, produced an AUC of 0.697. Eighty-six of the 1521 variables appeared to be significant predictors of future self-harm. The models were re-run including only these significant variables, which resulted in a small improvement in predictive accuracy. The average AUC for the final models was 0.723; the ROC curve is shown in Figure 1. Prediction statistics are shown in Table 1.

The variables that were significantly predictive are shown in Table 2. Three quarters (66/86) of the significant predictors came from just three survey instruments: the Youth Self Report (YSR) and Child Behavior Checklist (CBCL) – child- and caregiver-administered instruments, respectively – measuring behavior problems, and the Children's Depression Inventory (CDI), a child-administered measure of depression symptoms. The 10 strongest predictors also all came from these three instruments. Prior self-harm was the strongest predictor, followed by the YSR subscale for internalizing problems, and a single YSR item measuring suicidal ideation. Three more YSR subscales (self-destructive, anxious/depressed, and somatic complaints) and the YSR total score were in the top 10. The negative self-esteem subscale of the CDI was the strongest predictor from that instrument (7th overall), and the CBCL total score was the strongest from that instrument (9th overall). In general, if an item or construct was a significant predictor then both the child- and caregiver-reported measure of the construct was significant; however, child-reported measures were generally stronger predictors than the corresponding caregiver-reported measures. The most dramatic example was the YSR item measuring suicidal ideation, the 3rd strongest predictor. The CBCL item measuring child's suicidal ideation was ranked 24th overall, not even in the top 20%.

Other significant predictors included items from instruments measuring trauma symptoms, parental monitoring and maltreatment at home, running away from home, the presence of supportive adults, substance abuse, and mental health diagnoses and medication. From these other scales, only one variable was in the top 20% of predictors: the number of prescription medications the child was currently taking for emotional/behavioral problems. Only one variable from the CPS caseworker interview was a significant predictor: the t-score for a measure of the CPS agency's resistance to change. No variable related to the CPS investigation, the index maltreatment allegation, or the case outcome and CPS response was significantly predictive.

Figure 1. ROC curve for final prediction model (AUC 0.723)

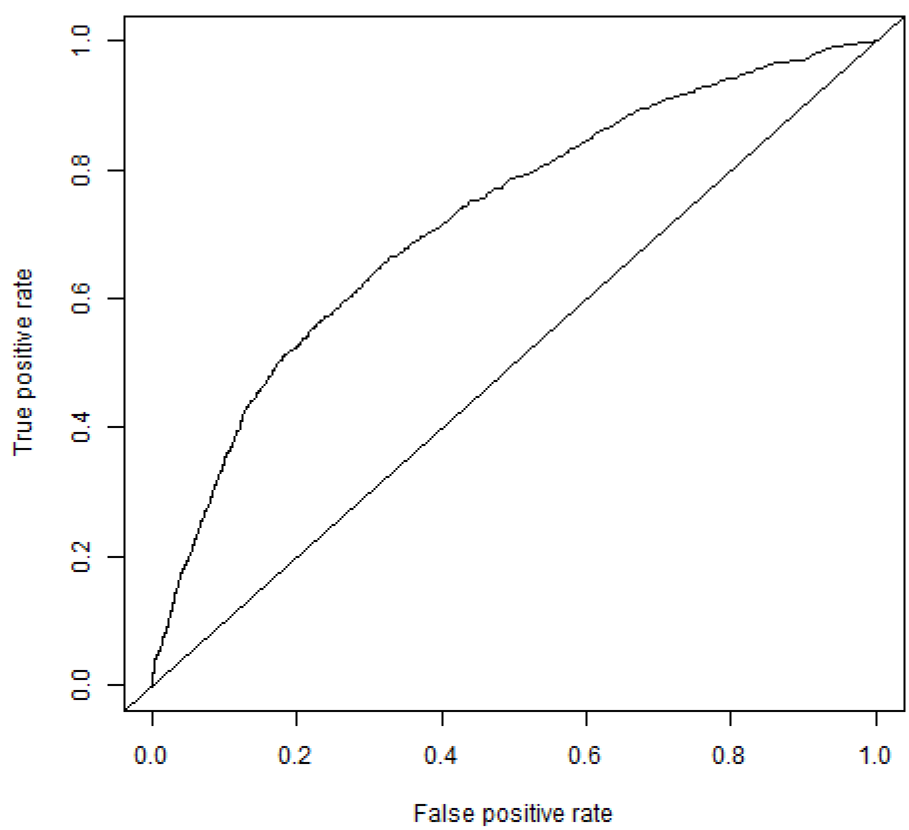


Table 1. Prediction statistics averaged across the 20 random forests

Cut-off	**Sensitivity**	**Specificity**	**PPV**	**NPV**
0.5	0.01	>0.99	N/A	0.91
0.4	0.05	0.99	0.39	0.92
0.3	0.15	0.97	0.29	0.92
0.2	0.33	0.91	0.25	0.94
0.1	0.63	0.70	0.17	0.95

Table 3. Significant predictors of future self-harm

Instrument (Construct)	Items identified as significant predictors
Child variables	
N/A	**Self-harm (child- or caregiver-reported) in past 6 months**
Youth Self Report (behavior problems)	**Internalizing problems subscale score** **Suicidal ideation: "I think about killing myself" (q91)** **Self-destructive subscale score[2]** **Anxious/depressed subscale score** **Total behavior problems score** **Somatic complaints subscale score** **Thought problems subscale score** **Low self-worth: "I feel worthless or inferior" (q35)** Social problems subscale score Aggressive behavior subscale score Attention problems subscale score Depression: "I am unhappy, sad, or depressed" (q103) Externalizing problems subscale score Feeling unloved: "I feel that no one loves me" (q33) Anxiety: "I am too fearful or anxious" (q50) Screaming: "I scream a lot" (q68) Dizziness: "I feel dizzy" (q51) Loneliness: "I feel lonely" (q12) Nightmares: "I have nightmares" (q47) Withdrawn subscale score Pyromania: "I set fires" (q72) Social problems: "I don't get along with other kids" (q25) Disobedience: "I disobey my parents" (q22) Fighting: "I get in many fights" (q37) Trouble sleeping: "I have trouble sleeping" (q100) Guilt: "I feel too guilty" (q52) Nausea: "I have nausea or feel sick without a known medical cause" (q56c) Runaway: "I run away from home" (q67) Paranoia: "I feel that others are out to get me" (q34) Violence: "I physically attack people" (q57)

Instrument (Construct)	Items identified as significant predictors
Children's Depression Inventory (mental health)	**Negative self-esteem subscale score** **Total depression score** **Negative mood subscale score** Relative worth: "I am just as good as other kids; I can be as good as other kids if I want; I can never be as good as other kids" (q24) Self-blame: "I do most things OK, I do many things wrong; I do everything wrong" (q3) Interpersonal problems subscale score Low appetite: "I eat pretty well; Many days I do not feel like eating; Most days I do not feel like eating" (q18) Anhedonia subscale score Feeling unloved: "I am sure that somebody loves me; I am not sure if anybody loves me; Nobody really loves me" (q25) Self-loathing: "I like myself; I do not like myself; I hate myself" (q7) Ineffectiveness subscale score Suicidal ideation: "I do not think about killing myself; I think about killing myself but would not do it; I want to kill myself" (q9) Loneliness: "I do not feel alone; I feel alone many times; I feel alone all the time" (q20) Self-image: "I look OK; There are some bad things about my looks; I look ugly" (q14) Violence: "I get along with people; I get into fights many times; I get into fights all the time" (q27) Crying: "I feel like crying once in a while; I feel like crying many days; I feel like crying every day" (q10) Self-blame: "Bad things are not usually my fault; Many bad things are my fault; All bad things are my fault" (q8)
Denver Youth Survey (delinquency)	Runaway to where: "Where did you run to: relative's; friend's; other" (q2d) # times child ran away from home in past 6 months (q2) Alone when ran away (vs with others or never ran away) (q2f)
Parent-Child Conflict Tactics Scale, adapted (maltreatment)	Parent psychological aggression score Parent nonviolent discipline score # times parents threatened to kick child out of the house (q12)
Trauma Symptom Checklist for Children – PTSD section, adapted (mental health)	Trauma total score Unwanted mental imagery: "Scary ideas or pictures just pop into my head" (q2) Fear of men: "Feeling scared of men. how often does this happen?" (q6)
Resiliency Scale – LongSCAN (protective factors)	Protective factors total score Parental support: "Do you feel you can go to a parent or someone who is like a parent with a serious problem?" (q2)

Instrument (Construct)	Items identified as significant predictors
CRAFFT (substance abuse)	# times sniffed glue/paint/aerosols to get high, lifetime (q26)
Supervision-Child Scale from Fast Track Project (behavioral monitoring)	Child stays somewhere other than home and is unsupervised in the evening (q21)
Caregiver variables	
Child Behavior Checklist (child behavior problems)	**Total behavior problems score** **Thought problems subscale score** **Internalizing problems subscale score** **Somatic complaints subscale score** Suicidal ideation: “Child talks about killing him/herself” (q91) Other problems subscale score Anxious/depressed subscale score Nightmares: “Child has nightmares” (q47) Externalizing problems subscale score Auditory hallucinations: “Child hears sounds or voices that aren’t there” (q40) Nausea: “Child has nausea or feels sick without a known medical cause” (q56c) Aggressive behavior subscale score Destructive: “Child destroys his/her own things” (q20) Attention problems subscale score Dizziness: “Child feels dizzy” (q51) Pains: “Child has aches or pains (not stomach or headaches) without a known medical cause” (q56a) Trouble sleeping: “Child has trouble sleeping” (q100) Stomachaches: “Child has stomachaches or cramps without a known medical cause” (q56f) Injuries: “Child gets hurt a lot or is accident-prone” (q36)
PHS (health & disabilities services received by child)	**# prescription medications child currently taking for emotional or behavioral problems (q3a42)** # times child been to ER/Urgent Care for an illness or injury Child currently has depression, anxiety, an eating disorder, or other emotional problem (q3a6) Doctor ever recommended child take medication for an emotional/behavioral problem (q3a34)
Parent-Child Conflict Tactics Scale	Nonviolent discipline score

Instrument (Construct)	Items identified as significant predictors
Caseworker variables	
Organizational Social Context (climate of CPS agency)	OSC Resistance [to change] t-score

1) Variables in top 20% of importance scores with p <0.001 are in **bold**
2) This sub-scale was not part of the original YSR instrument; it is a re-combination of existing YSR items used in the NSCAW, see NSCAW User's Manual Appendix 3 for details

DISCUSSION

In this study, a random forest model was able to predict self-harm 12-18 months in advance with low to moderate accuracy: an AUC of 0.72 for the final model. Predictive accuracy in this adolescent population is lower than what has been reported using machine learning methods in adults. Only one other study to date has used ML to examine self-harm in youth, a retrospective cohort study using EMR to predict admission to a hospital for a suicide attempt (Walsh et al., 2018). Walsh and colleagues attempted to distinguish suicide attempters from three control groups: patients admitted for non-suicidal self-injury or accidental injury, patients with depression, and the general hospital population. Predictive accuracy in the current study was lower than the results observed by Walsh et al., who found AUCs ranging 0.82-0.85 for other injury control adolescents, 0.87-0.90 for depressed control adolescents, and 0.94-0.97 for general control adolescents. There are several possible explanations for this discrepancy.

Walsh et al. estimated the AUC using an adjustment for over-fitting, the so-called optimism adjustment proposed by Harrell (Harrell, Lee, & Mark, 1996), while the current study used an independent validation sub-sample of the dataset to calculate the AUC. It is possible that optimism-adjusted metrics may still be inflated; the adjustment was originally proposed for use in small datasets, where withholding data for validation might severely compromise efficiency, not for widespread use. Additionally, Walsh et al. had two suicide experts and a third adjudicator review the medical records to confirm suicidal intent in patients with a self-harm ICD code, and thus the outcome in their study is likely highly specific. The current study was limited to the use of two survey questions that did not distinguish between suicide attempts and non-suicidal self-injury (see Methods). This non-specific outcome measure is a general limitation of the current study, and likely contributes to the poorer model performance.

A final explanation relates to the population that the current study is drawn from. The NSCAW includes only children who have been the subject of a CPS investigation for maltreatment or neglect. While it does include children regardless of the outcome of the investigation, the whole sample may have a more similar risk profile than would a random sample from the general U.S. youth population. The more similar individuals are, the more difficult splitting them into different classes becomes, and the Walsh study shows decreasing predictive accuracy as the controls become more similar to the suicide attempters, from general hospital population to depressed patients to patients with non-suicidal self-injury.

This is the first study to assess not just the relative importance of predictors of self-harm, but to use tests to quantify statistical significance. We found 88 variables that were significant predictors. Despite assessing variables from a wide range of conceptual domains, three-quarters of all the significant predictors came from just two domains: behavior problems – reported by both child and caregiver – and child-reported depressive symptoms. Internalizing problems were more predictive than externalizing ones, and most constructs that were predictive were significant when reported by both children and caregivers, which supports the validity of these results and the reliability of their measurement. However, in most cases, youth responses were more predictive than caregiver responses for the same items/constructs, suggesting that children and youth may be better reporters of their own emotional status than their caregivers.

It is important to remember that, even using longitudinal data where the temporal precedence between the risk factors and the outcome is known, models developed for prediction are not the same as models developed for making causal inferences. Variables that are significant predictors of self-harm may or may not be causally related, because prediction models make no effort to control for confounding. Therefore, it should not be automatically assumed that variables that are significant predictors of self-harm are necessarily important to the etiology of self-harm or would be viable targets for preventive interventions. Nevertheless, the degree to which results from prediction models are compatible with existing etiologic theories can be illuminating, and prediction results can suggest potentially promising avenues for future research that does use methods designed for causal inference. In particular, machine learning prediction models that can be used with high-dimensional data, like random forests, can be an efficient tool for assessing large numbers of possible risk factors prior to conducting more targeted causal analyses. Keeping the limitations of prediction models in mind, the results of the current study have several theoretical and practical implications.

One of the most popular theoretical models of suicidal behavior is the interpersonal-psychological theory proposed by Thomas Joiner (Joiner, 2005). In the Joiner model, the desire to die (one of two prerequisite conditions for suicide, the other being capability to kill oneself) develops from the simultaneous presence of two psychological/emotional states within an individual: perceived burdensomeness and thwarted belongingness. The significant predictors identified in this study are generally consistent with this model, although there are some noteworthy discrepancies. No question in the NSCAW measures burdensomeness directly, e.g. by asking something like, “My family would be better off without me.” However, related constructs like self-worth and self-blame were among the strongest single-item predictors.

After suicidal ideation and some composite scores for internalizing problems, the next strongest predictors were children reporting that, “I feel worthless or inferior most of the time,” and “I do everything wrong,” and several similar sentiments. Items suggesting thwarted belongingness were also significant predictors, including “I feel that no one loves me”/”Nobody really loves me” (measured from different instruments) and “I don’t get along with other kids,” and “I feel lonely.” However, instruments that measured relationships with peers and social support, constructs that seem closely related to belongingness, were not significant predictors. Additionally, the significant items related to self-worth were generally stronger predictors than those related to belongingness. This may be related to the age of this population. Adolescence is a time of evolving identity and social transition for most youth, and thwarted belongingness may, at least temporarily, be much more prevalent than at older ages, which would make it less useful for stratifying people into different self-harm risk groups even if it remains an important construct.

The relatively poor performance of the predictive model suggests that, at this time, using indicated suicide prevention interventions may be challenging in the CPS context, at least using the variables currently available in the NSCAW. Universal interventions that seek to bolster children's self-worth, and teach coping strategies for dealing with internalizing problems and depressive symptoms, may hold more promise than attempting to target high-risk individuals. Many of the significant predictors are related to psychological conditions that may not be easily malleable without professional medical help; diagnosed mental disorders and prescription medications were strong predictors. However, one potential target is the use of psychological aggression by parents. This was a significant predictor and, equally important, parental behavior is an area into which CPS might be able to intervene. However, as noted above, analyses that control for bias should be conducted before strong conclusions are drawn regarding intervention targets, as intervening on something implies a causal relationship with self-harm.

This study is only the second to use machine learning methods to predict self-harm among adolescents, and the contrast between our results and prior results that showed substantially greater predictive accuracy, but no independent sample validation suggest that more work is needed to determine in what populations prediction might be feasible and what information is necessary for a good model. This study is the first to use methods to evaluate the statistical significance of predictors. The results are generally concordant with the interpersonal-psychological theory of suicide. Several targets for preventive intervention within the CPS system emerge, including child feelings of self-worth and parental discipline practices, but more work is needed with models that account for confounding before we can say with confidence that these likely have a causal association with self-harm and would, therefore, be viable intervention targets.

REFERENCES

Administration for Children and Families. (2020). Child Maltreatment 2018. Retrieved from https://www.acf.hhs.gov/sites/default/files/cb/cm2018.pdf

Achenbach, T. M. (2009). The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Angelakis, I., Austin, J. L., & Gooding, P. (2020). Association of Childhood Maltreatment With Suicide Behaviors Among Young People: A Systematic Review and Meta-analysis. JAMA Netw Open, 3(8), e2012563. doi:10.1001/jamanetworkopen.2020.12563

Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., . . . Skopp, N. A. (2019). Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. JAMA Psychiatry, 76(6), 642-651. doi:10.1001/jamapsychiatry.2019.0174

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.

Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. J Affect Disord, 245, 869-884. doi:10.1016/j.jad.2018.11.073

Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS) [Online]. (2003). National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (producer). Available from: www.cdc.gov/ncipc/wisqars. Accessed July 2020.

Curtin, S. C. (2020). State Suicide Rates Among Adolescents and Young Adults Aged 10-24: United States, 2000-2018. Natl Vital Stat Rep, 69(11), 1-10. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33054915

Degenhardt, F., Seifert, S., & Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. Brief Bioinform, 20(2), 492-503. doi:10.1093/bib/bbx124

Dowd, K., Dolan, M., Smith, K., Orin, D., Keeney, J., Wheeless, S., & Biemer, P. (2014). NSCAW II Combined Waves 1-3 Restricted Release Version Data File User's Manual. Ithaca, NY: National Data Archive on Child Abuse and Neglect, Cornell University.

Evans, R., White, J., Turley, R., Slater, T., Morgan, H., Strange, H., & Scourfield, J. (2017). Comparison of suicidal ideation, suicide attempt and suicide in children and young people in care and non-care populations: Systematic review and meta-analysis of prevalence. Child Youth Serv Rev, 82, 122-129.

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., . . . Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychol Bull, 143(2), 187-232. doi:10.1037/bul0000084

Harrell, F. E., Jr., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med, 15(4), 361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Huque, M. H., Carlin, J. B., Simpson, J. A., & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. BMC Med Res Methodol, 18(1), 168. doi:10.1186/s12874-018-0615-6

Janitza, S., Celik, E., & Boulesteix, A.-L. (2018). A computationally fast variable importance test for random forests for high-dimensional data. Advances in Data Analysis and Classification, 12(4), 885-915. doi:10.1007/s11634-016-0276-4

Joiner, T. E. (2005). Why people die by suicide. Cambridge, MA: Harvard University Press.

Karpievitch, Y. V., Hill, E. G., Leclerc, A. P., Dabney, A. R., & Almeida, J. S. (2009). An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. PLoS One, 4(9), e7087. doi:10.1371/journal.pone.0007087
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Sandri, M., & Zuccolotto, P. (2008). A bias correction algorithm for the Gini variable importance measure in classification trees. J Computational Graphical Stats, 17(3), 611-628.
Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics, 8, 25. doi:10.1186/1471-2105-8-25
Turecki, G., & Brent, D. A. (2016). Suicide and suicidal behaviour. Lancet, 387(10024), 1227-1239. doi:10.1016/S0140-6736(15)00234-2
Valdiviezo, H. C., & Aelst, S. V. (2015). Tree-based prediction on incomplete data using imputation or surrogate decisions. Information Sciences, 311, 163-181.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. J Statistical Software, 45(3).
Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. J Child Psychol Psychiatry, 59(12), 1261-1270. doi:10.1111/jcpp.12916

Chapter 5

Identifying causal risk factors for self-harm among adolescents with Child Protective Services (CPS) contact

ABSTRACT

Purpose: To prioritize potential targets for self-harm preventive interventions in adolescents who have had contact with Child Protective Services by estimating causal effects of several potentially modifiable predictors.

Methods: Data came from the second National Survey of Child and Adolescent Well-being cohort (NSCAW II), a nationally representative longitudinal survey. Three variables were identified as significant predictors of self-harm in a prior NSCAW II study and considered theoretically modifiable within the CPS system. They were: child's feelings of worthlessness, the presence of supportive adults in the child's life, and parental psychological aggression. Propensity score weighting was used to control for observed confounders measured at baseline, and then the association between each factor and future self-harm was estimated using weighted logistic regression. The average effect of exposure among the exposed (ATT) was estimated.

Results: No factor was significantly associated with self-harm in the analyses, although the presence of supportive adults had a substantial non-significant effect. For parental psychological aggression, the odds ratios (OR) comparing low and high aggression to none were 0.93 (0.35-2.45) and 1.25 (0.55-2.82), respectively. For feelings of worthlessness, the OR was 1.73 (0.70-4.27), and for supportive adults 0.58 (0.28-1.19). Due to the combination of NSCAW survey weights and propensity score weights the effective sample size was

substantially reduced – from 881 to ~140-305 depending on the analysis – which may have affected statistical power.

Conclusions: The null findings when factors were examined separately support theoretical models describing suicidal behavior as the result of multiple chains of risk and protective factors interacting together. Preventing suicide and self-harm in adolescents likely requires a multifaceted approach, and fostering supportive and encouraging relationships with adults may play an important part in preventive intervention packages. Despite the lack of statistical significance, youth in the NSCAW with high levels of adult support had 42% lower odds of self-harm than if they had had low/no adult support.

INTRODUCTION

Suicide is the tenth leading cause of death in the U.S. and the second leading cause of death among adolescents and young adults aged 10-24 years of age (CDC, 2003). Suicide rates among adolescents and young adults have been increasing steadily for the past decade (Curtin, 2020). There is general agreement that suicide and self-harm are the end results of a complex web of risk and protective factors, with no single factor either necessary or sufficient to cause self-harm (Franklin et al., 2017). Most contemporary theoretical models are based on a diathesis-stress paradigm, where predisposing vulnerabilities interact with negative life events in a particular environmental context that facilitates self-harm (Hawton, Saunders, & O'Connor, 2012; Turecki & Brent, 2016). There has been a wealth of research identifying factors associated with suicide attempts and self-harm, but very few studies have moved beyond association to rigorously evaluate possible causal relationships. A recent systematic review and meta-analysis noted that, “there is very little existing research on the causal risk factors for [suicidal thoughts and behaviors]” (Franklin et al., 2017). In the current study, we began to address this deficit by using longitudinal data and statistical methods that enable the evaluation of causal relationships.

This work builds upon prior work that used a random forest machine learning algorithm to predict self-harm and identify which variables contributed significantly to predictive accuracy in a sample of high-risk adolescents (Kahn, unpublished). Both the prediction study and the current analysis were conducted with adolescents who have had contact with the U.S. Child Protective Services (CPS) system. Prior research has documented increased prevalence of suicidal behavior among youth with CPS contact compared to the general population (Heneghan et al., 2015) as well as increased odds of suicide attempt among youth in foster care or CPS-supervised home care compared to the general population (Evans et al., 2017). These youth also represent a substantial population (>2 million) in a boundaried setting, the CPS system, that may be positioned to facilitate or deliver preventive interventions to children and their families.

The aim of the current study was to identify possible targets for intervention, and so candidate variables should be both strong predictors of self-harm and feasibly modifiable within CPS. In the random forest analysis, the strongest predictors were summary measures of various internalizing behavior problems and depression, with most significant predictors being component variables of these scales. Because these summary measures represent multiple – and potentially distinct – pathologies, interpretation of any results was deemed more challenging. Therefore, the decision was made to focus instead on a more narrowly defined construct. A child-reported indicator of feelings of worthlessness was the only single survey item (besides suicidal ideation) that had an effect on predictive accuracy that was comparable to the depression/internalizing summary scores. Additionally, assessing worthlessness had the potential to inform theoretical models of suicide, many of which posit worthlessness or related mental states (thwarted belongingness, perceived burdensomeness (Joiner, 2005)) as pre-conditions for suicide attempts. Based on the strength of its association with self-harm and the potential to inform understanding of the etiology of suicide, feelings of worthlessness was selected in lieu of a summary measure of internalizing problems as one of the candidates for the current analysis.

Two variables were chosen that were also significant predictors in the random forest model and seemed to be targets for interventions that would naturally fall within the purview of CPS. The first was a summary score for child-reported frequency and intensity of parental psychological aggression; things like threatening to kick the child out of the house. The second was another summary score measuring the degree to which children reported that they had a supportive adult presence. Called “protective factors” in the NSCAW survey, it encompassed items such as whether the child had an adult they could turn to if they needed help, or whether an adult had made a difference in the child’s life. All three variables – feelings of worthlessness, parental psychological aggression, and adult support – have been shown to be significantly predictive of self-harm in a population of adolescents with CPS contact. In the present analysis, we aimed to evaluate whether any of these predictors were likely to be causal risk factors for self-harm, and thus would make viable targets for preventive interventions that could be delivered within CPS.

METHODS

Data

Data came from the National Survey of Child and Adolescent Wellbeing II (NSCAW II). The NSCAW II was a 3-year longitudinal study of youth <18 years of age and their families who had been the subject of an investigation by Child Protective Services (CPS) for alleged maltreatment between February 2008 and April 2009 (Dowd et al., 2014). The NSCAW II followed a two-stage stratified sampling design. In the first stage, the U.S. was divided into nine strata corresponding to the eight states with largest CPS caseloads and the remaining 42 states and the District of Columbia. Within these strata primary sampling units (PSUs) were defined and selected. The PSU was defined as the geographic area containing a population served by a single child welfare agency; typically this was a single county. Eighty-one PSUs were selected, and all families with active cases in the selected PSUs were invited to participate in the study. Baseline interviews with children and their primary caregiver were conducted within approximately 4 months of the conclusion of the CPS investigation, and families were enrolled regardless of the investigation outcome. Follow-up interviews were conducted at approximately 18- and 36-months post-baseline. The sample was weighted for nonresponse and to be representative of the overall population with CPS contact. The analytic sample included children aged 11-17 years at baseline.

Variables

We used the same definition of self-harm as has been reported in detail in other studies. Briefly, the Youth Self-Report and the Child Behavior Checklist for youth and caregivers, respectively, each contain a single item capturing frequency of self-harm and/or suicide attempts in the previous 6 months. These items were dichotomized, and youth who either self-reported or had a caregiver report any instance of the child's self-harm were considered positive for self-harm.

Candidate risk factors were chosen based on prior work using a random forest model to identify significant predictors of self-harm in the NSCAW. The three candidates were: (1) child reported feelings of worthlessness; (2) a summary measure of the degree to which children reported they had an adult or adults who supported and believed in them, and whom they could turn to for help (termed Protective Factors in the NSCAW), and (3) a summary measure of the frequency of verbal abuse the child reported experiencing from their primary caregiver (termed Child Maltreatment: Psychological Aggression in the NSCAW). Details on how each of these variables were measured are available in the NSCAW Data File User's Manual.

Prior to data analysis, missing data were multiply imputed. Details of the imputation procedure have been reported previously (Kahn, unpublished). We used the same 20 imputed datasets as the random forest analysis that informed the choice of candidate risk factors (see Chapter 4).

Statistical analyses

To adjust for the observed confounders, we used propensity score weighting (PSW) as described elsewhere (Austin & Stuart, 2015). Briefly, when the gold standard for causal inference – a randomized study – is not possible, PSW allows researchers to mimic the unbiasedness of a randomized trial, at least with respect to the observed covariates, using observational data and considering the risk factor of interest in lieu of a randomly assigned treatment. Propensity scores are the estimated probability that an individual was "treated," i.e., positive for the risk factor (hereafter referred to as "exposed"), based on a set of covariates. Weighting by the inverse of this probability creates (ideally) a dataset where the distribution of observed covariates is the same for exposed and unexposed individuals; essentially, exposure can be considered randomly assigned with respect to the observed covariates. Assuming no unmeasured confounding and with known temporal ordering of the risk factor and outcome, effect estimates from a PSW model can be interpreted as causal effects. In the current study, to maintain temporal ordering, covariates measured at baseline were used to estimate the probability of exposure at first follow-up, and exposure was used to model self-harm at the second follow-up.

With regards to unmeasured confounding, the NSCAW is a rich dataset which permitted us to control for many important possible confounders (see the NSCAW Data File User's Manual or Kahn et al. Appendix A for detailed descriptions of the range of variables included in the NSCAW). Covariates in the PS models included representative variables from each domain that was identified as significantly predictive in the random forest model, and additional variables shown to be associated with self-harm in the literature. Covariates included sex, race/ethnicity, and age at baseline. Alleged maltreatment type was collapsed into a 4-level categorical variable based on type of abuse and whether the allegation was substantiated or not. Physical and sexual abuse were combined, and all other abuse/neglect types were combined with a binary indicator for substantiated or not. Other covariates included self-harm at baseline; and summary scores for internalizing and externalizing behavior problems, depression, trauma symptoms, loneliness/social dissatisfaction, delinquency, substance use, and parental physical violence. The exposure supportive adults was also used as a covariate for the other two exposures. The item measuring feelings of worthlessness was a component of the internalizing problems scale, and a very similar question was part of the depression scale. To avoid adjusting for the exposure itself, two related subscales which did not include the item were used instead in the PS model for worthlessness: the somatic complaints subscale of the internalizing problems scale (the most strongly predictive of the subscales that did not include worthlessness), and the anhedonia subscale of depression. (Table 1) Given the scope of included covariates, the assumption of unmeasured confounding seems reasonable.

There are a number of ways to implement PSW depending on the estimand of interest. Again, propensity scores (PS) estimate each individual's probability of being exposed. To estimate the effect of the exposure on the full sample, the so-called average treatment effect (ATE), cases are weighted by the inverse of the probability of their exposure status, i.e., 1/PS for the exposed and 1/(1-PS) for the unexposed. To estimate the effect of the exposure on just those individuals who were actually exposed, the average treatment effect among the treated (ATT), exposed cases are given a weight of 1 and unexposed cases are weighted by their odds of exposure, i.e., PS/(1-PS) (J. Lee & Little, 2017). For this analysis, we present both ATE and ATT estimands, but focus on the ATT under the assumption that future preventive interventions would be targeted to individuals who actually experienced the given risk factor.

To simplify both the estimation and interpretation of the propensity scores, prior to estimation the candidate risk factors were transformed into binary or categorical variables (Table 1). Feelings of worthlessness was dichotomized into none/any, supportive adults was dichotomized into those with a score in the bottom quartile versus in the top three quartiles, and parental psychological aggression was split into three categories, those reporting zero aggression, those with a positive aggression score in the bottom three quartiles, and those with a score in the top quartile. Multivariable logistic regression was used to estimate the propensity scores, using the covariates described above. Generalized boosted modeling was also examined to estimate PS, and although some studies have found that GBM can perform better than logistic regression (B. K. Lee, Lessler, & Stuart, 2011), in this analysis better covariate balance was obtained using logistic regression. Covariate balance was assessed by examining the standardized bias for the propensity score and all covariates; a standardized bias of less than 0.25 was considered adequate for balance (Ho, Imai, King, & Stuart, 2007). Variables with a bias of between 0.25 and 0.5 were included in the final effect estimation model to adjust for the remaining imbalance.

For complex survey data, sampling weights need to be incorporated into the PS estimation (Dugoff, Schuler, & Stuart, 2014). This was done, and additionally the top 1% of weights were trimmed (set to be equal to the 99th percentile), as this can improve the accuracy and precision of the final effect estimates (B. K. Lee et al., 2011). For psychological aggression, which had three categories, the highest level of exposure was used as the "treatment" group when calculating the PS weights for the ATT analysis. PS estimation and weighting was done separately for each imputed dataset. Weighted logistic regression was used to estimate the association between each risk factor and self-harm, covariates with remaining imbalance were included as needed, and the results across imputations were combined following Rubin's rules.

All procedures for the NSCAW II were approved by the Research Triangle Institute's Institutional Review Board and all analytic work with the NSCAW II Restricted Release dataset was approved by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board. All analyses were conducted in R version 4.0.4 (R Core Team, 2021). The 'MatchThem' and 'cobalt' packages were used to estimate the propensity scores and visualize the covariate balance, and the 'survey' package was used for the weighted outcome models. Multiple imputation was done previously with the 'mice' package.

Table 1. Propensity score modeling decisions for candidate risk factors

	Feelings of worthlessness	Supportive adults	Parental psychological aggression
Raw variable	3-level ordinal, never/sometimes/ often	continuous, range 0-5	continuous, range 0-125
Transformed variable	dichotomous, never/ever	dichotomous, 0-3/4-5	3-level ordinal, 0, 1-16, 17-125
Variables included in PS model	sex race/ethnicity age alleged maltreatment self-harm at baseline somatic complaints externalizing behavior problems depression - anhedonia trauma symptoms supportive adults in child's life loneliness/social dissatisfaction delinquency substance abuse parental physical violence	sex race/ethnicity age alleged maltreatment self-harm at baseline internalizing behavior problems externalizing behavior problems depression trauma symptoms loneliness/social dissatisfaction delinquency peers substance abuse parental physical violence	sex race/ethnicity age alleged maltreatment self-harm at baseline internalizing behavior problems externalizing behavior problems depression trauma symptoms supportive adults in child's life loneliness/social dissatisfaction delinquency peers substance abuse parental physical violence

RESULTS

The analytic sample included 881 youth; characteristics of the sample are shown in Table 2. The average sample characteristics before and after ATT weighting, including the mean and range of covariate values between imputed datasets, are shown in Figures 1-3. Adequate covariate balance was achieved across exposure groups defined by having supportive adults, but there were lingering imbalances between groups defined by feelings of worthlessness and parental psychological aggression. Therefore, the final model for worthlessness also adjusted for somatic complaints and loneliness/social dissatisfaction, and the model for psychological aggression adjusted for externalizing problems and delinquency. Of note is the fact that the combination of survey design and propensity score weighting resulted in substantially reduced effective sample sizes for all three analyses (Figures 1-3).

In the PSW analyses, no candidate risk factor was statistically significantly associated with self-harm (Table 3). The effect estimate for protective factors was in the expected direction and approached significance. Among youth with high levels of adult encouragement and support, the odds of self-harm were 42% lower than if they had had less support (odds ratio 0.58, 95% CI 0.28-1.19, $p = 0.14$). The effect of worthlessness was also in the expected direction, but also not statistically significant. Among adolescents who reported any feelings of worthlessness, their odds of self-harm were 73% higher than if they had not had those feelings (OR 1.73, 0.70-4.27, $p = 0.24$). The effect of parental psychological violence was essentially null, with no clear dose response and odds ratios close to 1. Only the category for the highest violence scores showed increased odds of self-harm compared to no violence, and the confidence intervals were wide for all categories (OR 0.93, 0.35-2.45, $p = 0.88$ for youth with scores 1-15; 1.25, 0.55-2.82, $p = 0.59$ for scores 16-125).

Table 2. Sample characteristics[1] (N = 881)

	Unweighted N (weighted %)[2] -or- Weighted median (quartiles)[2]
Female	497 (61.0%)
Age (years)	13 (12, 15)
White, non-Hispanic	355 (44.0%)
Black, non-Hispanic	243 (19.1%)
Hispanic	222 (30.4%)
Other race	61 (6.5%)
Major maltreatment, substantiated	143.6 (8.5%)
Major maltreatment, not substantiated	157.5 (26.6%)
Minor maltreatment, substantiated	320 .0 (17.0%)
Minor maltreatment, not substantiated	260.0 (47.9%)
Self-harm	106.8 (13.8%)
Self-harm at follow-up 2	72.4 (8.1%)
Feelings of worthlessness (sometimes)	113.3 (14.6%)
Feelings of worthlessness (frequently)	29.1 (3.4%)
Protective factors score	5 (4, 5)
Parental psychological aggression score	3.0 (0, 15.6)

[1] Measured at baseline unless otherwise noted

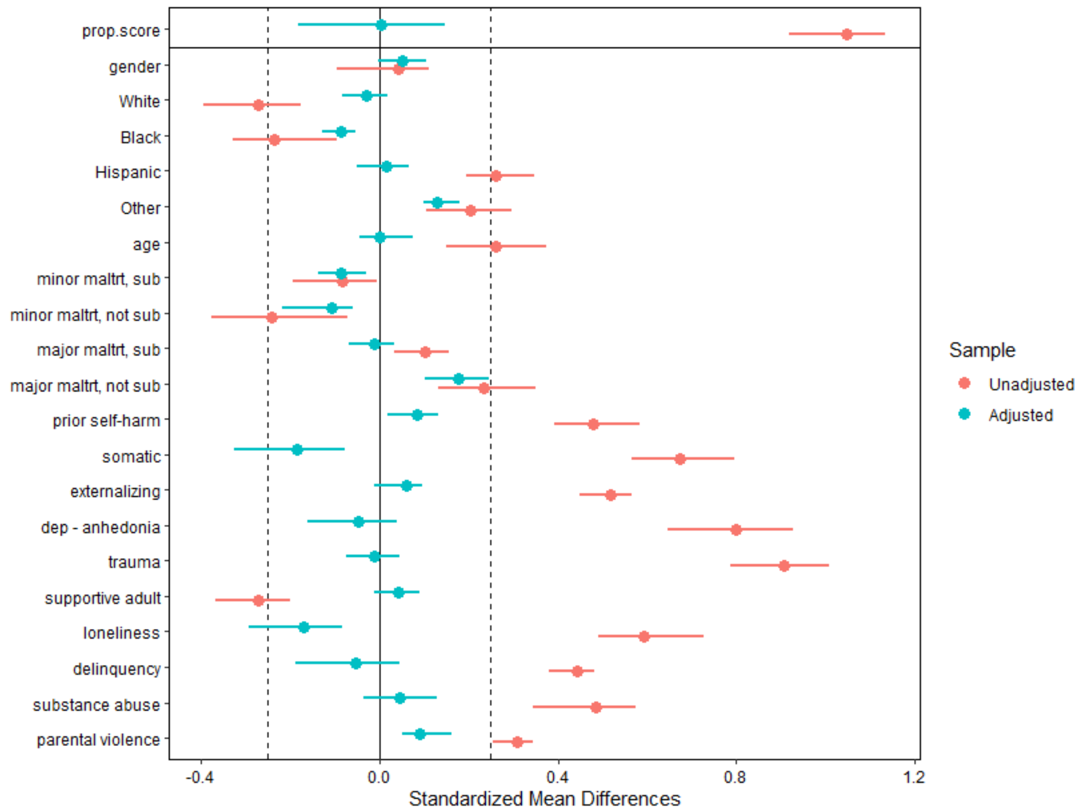
[2] Mean value across imputed datasets

Table 3. Propensity score weighted logistic regression

	ATT		**ATE**	
	OR (95% CI)	p-value	OR (95% CI)	p-value
Feeling worthless (dichotomous)	1.73 (0.70-4.27)	0.236	1.75 (0.65-4.69)	0.268
Protective factors (dichotomous)	0.58 (0.28-1.19)	0.139	0.59 (0.30-1.17)	0.135
Parental psychological aggression (score = 0)	ref	-	ref	-
Parental psychological aggression (1 - 16)	0.93 (0.35-2.45)	0.882	0.79 (0.36-1.71)	0.546
Parental psychological aggression (17 - 125)	1.25 (0.55-2.82)	0.591	1.21 (0.55-2.67)	0645

Figure 1. Covariate Balance, Feeling worthless (ATT)

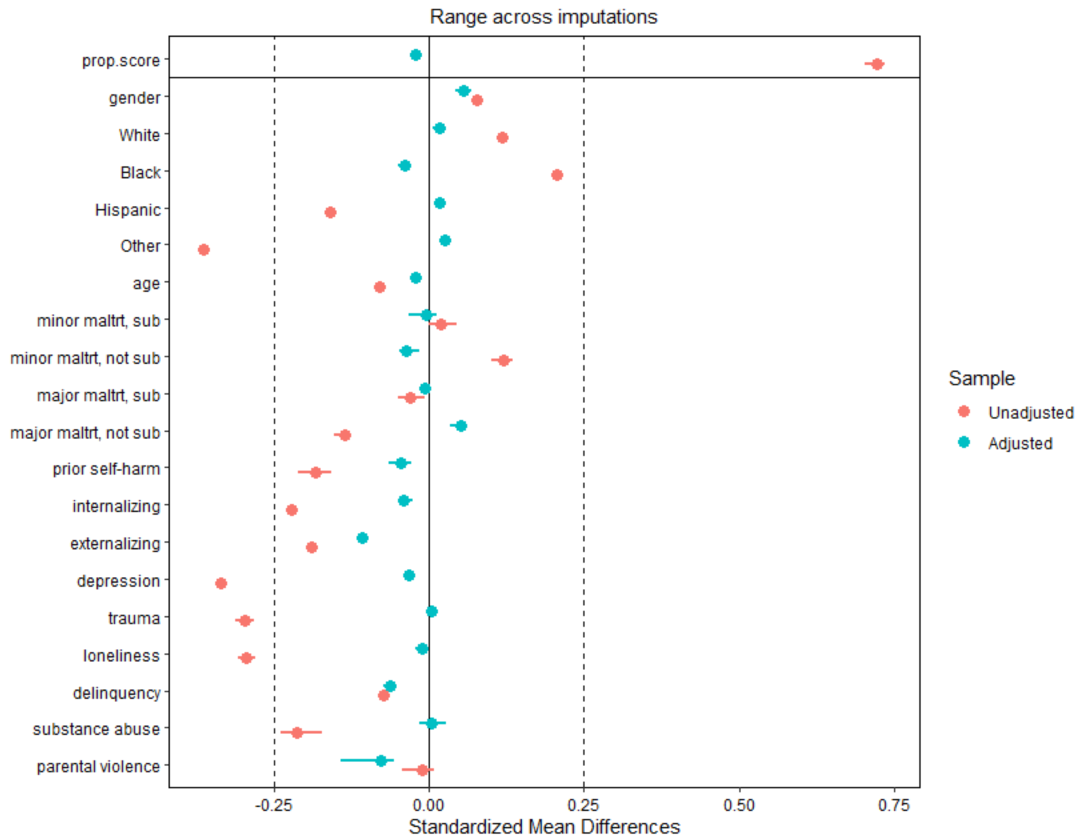
Range across imputations



<u>Average effective sample sizes across imputations</u>

	Control	Treated
Unadjusted	272.34	41.18
Adjusted	93.08	41.18

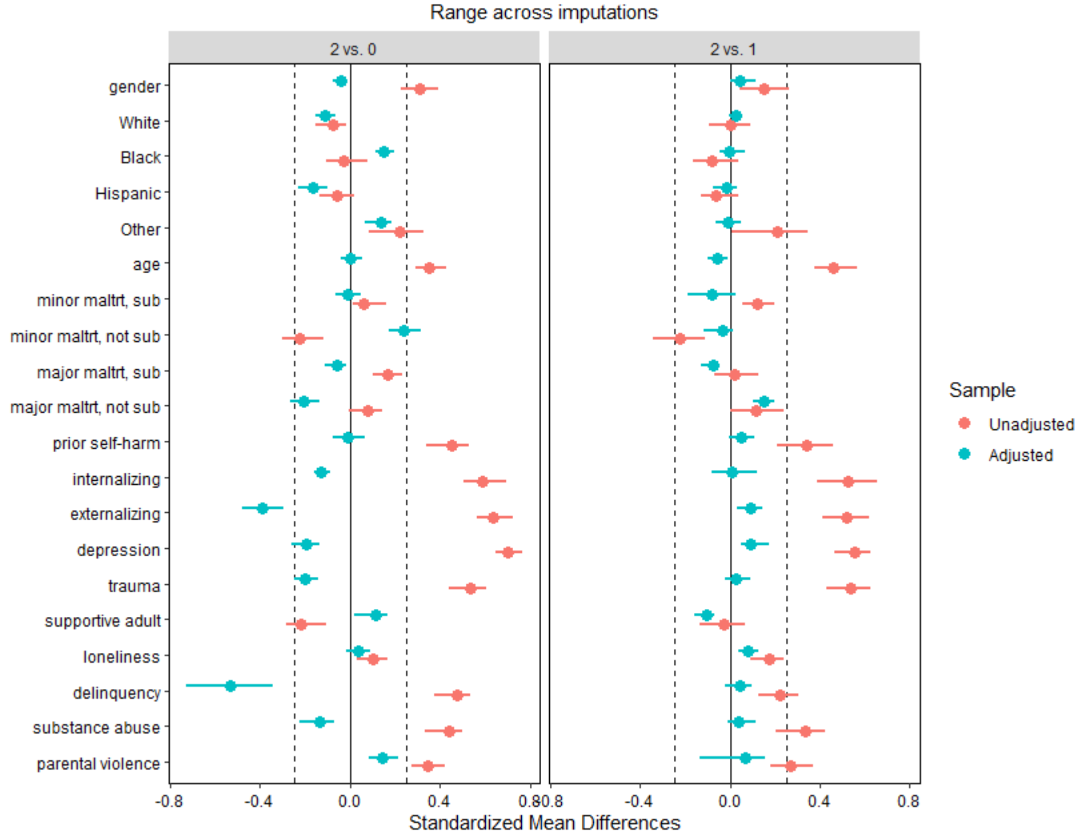
Figure 2. Covariate Balance, Supportive adults (ATT)



<u>Average effective sample sizes across imputations</u>

	Control	Treated
Unadjusted	61.65	253.31
Adjusted	49.24	253.31

Figure 3. Covariate Balance, Parental psychological aggression (ATT, highest aggro)



<u>Average effective sample sizes across imputations</u>

	0	1	2
Unadjusted	133.37	115.04	68.23
Adjusted	34.78	86.65	68.23

DISCUSSION

This study is one of the first to look at causal risk factors for self-harm in adolescents, and the first we are aware of to examine adolescents in the CPS system. As such, there are not many other studies against which to directly compare the results. The null results for parental violence seem to be at odds with the robust literature showing an association, albeit non-causal, between childhood physical maltreatment and suicidal ideation and attempts in adolescents, but should be interpreted within the context of the different study populations. Prior results are almost universally derived from comparisons of children in community settings with any abuse history compared to children with none, or between children in the child welfare system and those in the community (Miller, Esposito-Smythers, Weismoore, & Renshaw, 2013). In contrast, all children in the current study have been investigated by CPS for some form of maltreatment. The null results suggest that efforts to reduce parental psychological violence towards children, while a crucial goal in its own right, may not have measurable protective effects on adolescent self-harm once children have had CPS contact and/or enter the child welfare system. It may also be the case that, although we were able to control for the index maltreatment incident, parental violence prior to and following the index event may vary in ways that were not captured in the analysis and confound the results. Nevertheless, additional efforts to prevent self-harm in this population beyond reducing parental violence are likely warranted.

It is tempting to look towards fostering supportive adult/child relationships as a possible intervention target, and this approach merits additional study but also caution. The results of the current study – a reduction in self-harm risk by approximately 40% – are not statistically significant. The effect size is large and there is a reasonable theoretical basis for believing that supportive relationships with adults may reduce self-harm risk. There is a robust literature showing non-causal, negative associations between parental support and suicide attempts in adolescents (Bridge et al., 2006). In particular, one large analysis using Add Health data showed a significant protective association between a measure of parent-child connectedness and future suicide attempt among >12,000 U.S. adolescents (Kidd et al., 2006). Additionally, lower levels of criticism and parent-child conflict, and higher levels of parental warmth are associated with lower internalizing problems in youth (Yap & Jorm, 2015). The interpersonal-psychological theory of suicide posits thwarted belongingness as a necessary precursor to suicide attempt, and children who report that they can go to a parent or other adult with a problem and have an adult who made in difference in their lives are likely to feel like, at minimum, they belong within their family unit. Self-harm and even suicide attempts can be considered maladaptive coping strategies, and children who have an adult to turn to with a serious problem are, theoretically, less likely to turn to other extreme measures like self-harm. There are also elements of the analysis that may legitimately have biased the statistical tests towards the null. First, there is the fact that a continuous measure was dichotomized. Preliminary regression analyses (not reported) did suggest a small loss of information due to this transformation. More importantly, the low effective sample sizes imply limited statistical power to detect differences. Additional research is warranted to confirm the findings in the current study and explore possible effect modifiers, but fostering supportive adult relationships may be a viable target for preventing adolescent self-harm.

There are several limitations of this study to consider. One important limitation is that the outcome measure conflated suicide attempts and non-suicidal self-injury. Although these distinct phenomena do share many risk factors, the lack of specificity may have biased the results of the current analysis in ways that cannot be predicted. Another is that the analysis was limited to one baseline measure of the risk factors and one outcome measure at a fixed time point ~18 months post-baseline. Adolescence is a period of comparatively rapid growth and development for any youth, and youth in the NSCAW are likely in particularly volatile circumstances owing to the intervention of CPS into their family life. Correlations between the values for risk factors at baseline and follow-up were low for all variables (data not shown). It is possible that any or all effect estimates were biased due to unmeasured changes in the risk factors over time. It is also important to remember that no significant effect was observed on self-harm at 12-18 months specifically. These factors may be important determinants of self-harm within different timeframes. Future studies on self-harm should be thoughtful about the timing of data collection.

Additionally, identifying the appropriate set of covariates to balance in each PS model in an observational study is challenging but important. In order to interpret the effect estimates as causal effects, one must assume that there is no unmeasured confounding. We used a combination of theory and empirical assessment to select covariates for each PS model, but there nevertheless remains an untestable assumption of no unmeasured confounding. Finally, there is the fact that we attempted to measure the causal effect of one variable at a time, but most researchers and clinicians now agree that suicide attempts and self-harm are the result of numerous factors interacting together. The fact that feelings of worthlessness was not significantly associated with self-harm in this study, for example, does not mean that this cognitive state is likely unrelated to suicidal ideation and behavior. Rather, it suggests that attempting to intervene *solely* on feelings of worthlessness is not a promising approach to preventing self-harm in adolescence.

Despite these limitations, the current study makes several important contributions to the literature of adolescent suicide prevention. It is one of the first to use methods from causal inference and longitudinal data to evaluate causal risk factors. The data come from a nationally representative study, so results are generalizable to the whole U.S. population of youth with CPS contact. The results suggest that merely reducing the occurrence of parental psychological violence may be insufficient to prevent future self-harm once maltreatment has occurred. The null results of this study underscore the need for multi-faceted approaches to suicide prevention. That said, fostering supportive relationships with adults in the child's life may protect against self-harm and be an important component to a preventive intervention package, although future research is needed to bolster this finding and investigate possible moderating variables.

REFERENCES

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med, 34(28), 3661-3679. doi:10.1002/sim.6607

Bridge, J. A., Goldstein, T. R., & Brent, D. A. (2006). Adolescent suicide and suicidal behavior. J Child Psychol Psychiatry, 47(3-4), 372-394. doi:10.1111/j.1469-7610.2006.01615.x

Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS) [Online]. (2003). National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (producer). Available from: www.cdc.gov/ncipc/wisqars. Accessed July 2020.

Curtin, S. C. (2020). State Suicide Rates Among Adolescents and Young Adults Aged 10-24: United States, 2000-2018. Natl Vital Stat Rep, 69(11), 1-10. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33054915

Dowd, K., Dolan, M., Smith, K., Orin, D., Keeney, J., Wheeless, S., & Biemer, P. (2014). NSCAW II Combined Waves 1-3 Restricted Release Version Data File User's Manual. Ithaca, NY: National Data Archive on Child Abuse and Neglect, Cornell University.

Dugoff, E. H., Schuler, M., & Stuart, E. A. (2014). Generalizing observational study results: applying propensity score methods to complex surveys. Health Serv Res, 49(1), 284-303. doi:10.1111/1475-6773.12090

Evans, R., White, J., Turley, R., Slater, T., Morgan, H., Strange, H., & Scourfield, J. (2017). Comparison of suicidal ideation, suicide attempt and suicide in children and young people in care and non-care populations: Systematic review and meta-analysis of prevalence. Child Youth Serv Rev, 82, 122-129.

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., . . . Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychol Bull, 143(2), 187-232. doi:10.1037/bul0000084

Hawton, K., Saunders, K. E., & O'Connor, R. C. (2012). Self-harm and suicide in adolescents. Lancet, 379(9834), 2373-2382. doi:10.1016/S0140-6736(12)60322-5

Heneghan, A., Stein, R. E., Hurlburt, M. S., Zhang, J., Rolls-Reutz, J., Kerker, B. D., . . . Horwitz, S. M. (2015). Health-risk behaviors in teens investigated by U.S. Child Welfare Agencies. J Adolesc Health, 56(5), 508-514. doi:10.1016/j.jadohealth.2015.01.007

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis, 15, 199-236.

Joiner, T. E. (2005). Why people die by suicide. Cambridge, MA: Harvard University Press.

Kidd, S., Henrich, C. C., Brookmeyer, K. A., Davidson, L., King, R. A., & Shahar, G. (2006). The social context of adolescent suicide attempts: interactive effects of parent, peer, and school social relations. Suicide Life Threat Behav, 36(4), 386-395. doi:10.1521/suli.2006.36.4.386

Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. PLoS One, 6(3), e18174. doi:10.1371/journal.pone.0018174

Lee, J., & Little, T. D. (2017). A practical guide to propensity score analysis for applied clinical research. Behav Res Ther, 98, 76-90. doi:10.1016/j.brat.2017.01.005

Miller, A. B., Esposito-Smythers, C., Weismoore, J. T., & Renshaw, K. D. (2013). The relation between child maltreatment and adolescent suicidal behavior: a systematic review and critical examination of the literature. Clin Child Fam Psychol Rev, 16(2), 146-172. doi:10.1007/s10567-013-0131-5

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Turecki, G., & Brent, D. A. (2016). Suicide and suicidal behaviour. Lancet, 387(10024), 1227-1239. doi:10.1016/S0140-6736(15)00234-2
Yap, M. B., & Jorm, A. F. (2015). Parental factors associated with childhood anxiety, depression, and internalizing problems: a systematic review and meta-analysis. J Affect Disord, 175, 424-440. doi:10.1016/j.jad.2015.01.050

Chapter 6

Conclusions

The research comprising this dissertation has yielded new insights into the epidemiology of self-harm in an adolescent population that has had contact with Child Protective Services (CPS). This was the first study to look at adolescent self-harm in the NSCAW II follow-up waves, and some results were different from what was observed in prior studies using just the baseline data. Notably, the prevalence of self-harm remained stable among older adolescents (15-17 years) at 10-11% while among younger adolescents (11-14 years) it was elevated in the time period immediately following the CPS investigation, but declined over subsequent follow-up waves, from 13% at baseline to 6% to 3.5%. In the general population, suicide attempts are rare among younger adolescents (Bridge et al., 2006), so parents, CPS caseworkers, and clinicians should be aware that the period immediately following an investigation may represent a time of especially heightened risk for self-harm among young adolescents specifically, but that the risk wanes among younger adolescents while it does not among older adolescents, who may require longer-term monitoring and potentially preventive intervention. There is surprisingly little research on the duration and/or recurrence of suicidal ideation and behaviors among adolescents in the general population. In the NSCAW II, a substantial minority of adolescents reported self-harm at multiple survey waves, and Native American and Asian/Pacific Islander youth were significantly more likely to report persistent self-harm. Further research is warranted to examine the reasons for this discrepancy and to see whether these racial differences are present among the general population as well, or whether they are, for whatever reason, specific to youth with CPS contact.

This study is only the second that the author is aware of to use machine learning (ML) algorithms to predict self-harm in an adolescent population. The first study by Walsh and colleagues of adolescents admitted to a hospital ED for self-inflicted injury used EMR data and a random forest classifier, and yielded very promising results (Walsh et al., 2018). The area under the ROC curve (AUC) varied from 0.82 to 0.94 in the Walsh study, depending on the control group used (see Chapters 2 and 4 for details). In the current study, the final prediction model achieved an AUC of 0.72, well below the results from Walsh, et al. These differences in predictive accuracy may be due to differences in the methods used in each study or differences in the populations under assessment. The ability to accurately predict an individual's risk of future suicide attempt would be a tremendous help to both clinicians and policy-makers in intervening to prevent self-harm. The discrepant results from the only two studies to examine longitudinal prediction among adolescents suggests the need for further research to clarify whether and for what populations individual-level prediction is possible given currently available methods and data. As it stands, the author would not recommend the use of a ML risk algorithm to guide decision-making in the CPS system. However, the current analysis did provide information about the types of variables that best predicted self-harm in the NSCAW II cohort, which does have implications for the provision of preventive interventions within CPS and potentially to adolescents in general.

The NSCAW contains over 1,500 variables applicable to all adolescents, covering a wider range of domains than has been examined in prior ML studies, which have traditionally drawn from electronic medical records and thus lacked such characteristics as developmental and social functioning, home environment, and sub-clinical symptomatology and trauma history. Despite the variety of information available, significant predictors of self-harm in the NSCAW II came from a limited number of domains, fewer than was expected. The primacy of scales measuring internalizing behavior problems and depression, and to a lesser extent externalizing problems, as well as quantity of prescription medication taken for emotional and behavioral problems points to the central role that mental disorders play in adolescent self-harm. Such results are in line with the prior observation that a diagnosable mental disorder was present in approximately 90% of youth who died by suicide and up to 80% of youth who attempted suicide

(Bridge et al., 2006). Unfortunately, "effectively treat adolescent mental disorders" is a broad and ambitious strategy. Access to mental health services remains a challenge nationwide (Lipari, Sarra, Blau, & Rubenstein, 2016), and nearly all serious mental disorders have a proportion of patients who remain persistently non-responsive to treatment (Maalouf, Atwi, & Brent, 2011; Masi & Liboni, 2011; Masi et al., 2010), although progress is being made on both of these fronts. Luckily, the results of this study do point to an additional risk factor that may be more feasible to intervene on within the CPS system and would have an impact on the incidence of self-harm.

The results of the propensity score (PS) weighted analysis showed that having a supportive and encouraging adult presence in the child's life was strongly protective against future self-harm even after controlling for other known risk factors, cutting the risk by approximately 40%. The richness of the NSCAW allowed the author to control for a host of potential confounders when supportive adults was examined as a possible causal risk factor. A key assumption of such analyses is the absence of unmeasured confounding, and the ability to control for as many known risk factors as was done makes this assumption more plausible. Although the estimated odds ratio was not statistically significant ($p = 0.14$), there were several aspects of the data and analysis that may have influenced statistical power, including dichotomizing a continuous measure of supportiveness and low effective sample sizes following survey and PS weighting. It is also relevant that supportiveness, self-harm, and covariates were all measured at 18-month intervals in a sample of adolescents who were often changing rapidly due to the growth and development inherent in adolescence compounded by, in many cases, the intervention of CPS into their lives and homes following an investigation. Additional measurements at shorter follow-up periods might have improved our ability to accurately estimate the association between supportive adults and self-harm by measuring supportiveness with greater precision prior to the outcome. Finally, it is important to remember that suicidal behaviors are usually the result of a complex chain of factors, and so trying to estimate the effect of any single factor in isolation may be difficult at best and missing the point at worst.

In considering all of these factors, it is the author's opinion that interventions to promote supportive and encouraging behavior from parents in cases where adolescents report lacking an adult they can turn to for help may be effective at preventing self-harm within the U.S. CPS system, and that further studies in this area are warranted. It would be valuable to replicate these results in a study that addresses some of the limitations of the current work, namely by using a more precise definition of self-harm that distinguishes between suicide attempt and non-suicidal self-injury, collecting data at intervals sufficient to capture changes in important variables in a timely manner, and ensuring adequate effective sample size. As mentioned in Chapter 1, there are a variety of programs that aim to improve parenting skills, parent-child attachment, and family functioning. A review should be conducted to determine which existing programs, if any, have considered as an outcome the construct "adult supportiveness" as measured in the NSCAW, since this is the specific construct that appears to be protective. Additional research to examine in what specific circumstances adult supportiveness has the biggest impact, i.e. look for moderating variables, would help to tailor intervention delivery as well as to inform theoretical models. It is the author's opinion that work on moderating variables could be done prior to or as a part of an intervention trial.

In a broader sense, there remains a need for more research into causal risk/protective factors for suicidal behavior. The field has started to embrace machine learning methods in recent years, following the influential review by Franklin and colleagues (Franklin et al., 2017), but research into causal factors remains sorely lacking. This study is one of the first to use methods from the field of causal inference to estimate the effect of individual risk factors, and without additional studies against which to compare, it is difficult to draw strong conclusions. Moreover, there is a need to examine whether it makes sense to estimate the effect of individual factors on suicidal behavior, or whether it would be better to examine sets of factors; the effect of the co-occurrence of depressive symptoms and binge drinking in adolescents with versus without emotionally supportive parents, for example. Such research would undoubtedly be more difficult, but would also be more consistent with existing theoretical models for suicidal behavior as the result of a multitude of factors (Turecki & Brent, 2016).

Finally, as rich as the NSCAW is, one of the important elements it lacks is any information on short-term stressors and changes in mental state that may play a crucial role in the etiology of self-harm in adolescents; factors like a romantic break-up or drinking alcohol or exposure to suicidal behavior in media or online. Ecological momentary assessment (EMA), which involves gathering information from participants in real-time, often utilizing smart phones or other wearable technology, is a method which holds promise for collecting data on just these sorts of short-term stressors (Shiffman, Stone, & Hufford, 2008). Preliminary studies have found EMA to be generally acceptable and feasible for use in adolescents to collect information on suicidal ideation following psychiatric hospitalization (Czyz, King, & Nahum-Shani, 2018). Studies that incorporate information about short-term mental states and experiences, so-called precipitating risk factors, in addition to predisposing risk and protective factors could improve both predictive modeling for suicidal behavior and our understanding of the etiology of suicide.

REFERENCES

Administration for Children and Families. (2019). *The AFCARS Report #26*. Retrieved from https://www.acf.hhs.gov/sites/default/files/cb/afcarsreport26.pdf

Administration for Children and Families. (2020). *Child Maltreatment 2018*. Retrieved from https://www.acf.hhs.gov/sites/default/files/cb/cm2018.pdf

Achenbach, T. M. (2009). *The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Amini, P., Ahmadinia, H., Poorolajal, J., & Moqaddasi Amiri, M. (2016). Evaluating the High Risk Groups for Suicide: A Comparison of Logistic Regression, Support Vector Machine, Decision Tree and Artificial Neural Network. *Iran J Public Health, 45*(9), 1179-1187.

Anderson, H. D. (2011). Suicide ideation, depressive symptoms, and out-of-home placement among youth in the U.S. child welfare system. *J Clin Child Adolesc Psychol, 40*(6), 790-796.

Angelakis, I., Austin, J. L., & Gooding, P. (2020). Association of Childhood Maltreatment With Suicide Behaviors Among Young People: A Systematic Review and Meta-analysis. *JAMA Netw Open, 3*(8), e2012563.

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med, 34*(28), 3661-3679.

Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., . . . Reis, B. Y. (2017). Predicting Suicidal Behavior From Longitudinal Electronic Health Records. *Am J Psychiatry, 174*(2), 154-162.

Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., . . . Skopp, N. A. (2019). Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry, 76*(6), 642-651.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1983). *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Bridge, J. A., Goldstein, T. R., & Brent, D. A. (2006). Adolescent suicide and suicidal behavior. *J Child Psychol Psychiatry, 47*(3-4), 372-394.

Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *J Affect Disord, 245*, 869-884.

Burns, B. J., Phillips, S. D., Wagner, H. R., Barth, R. P., Kolko, D. J., Campbell, Y., & Landsverk, J. (2004). Mental health need and access to mental health services by youths involved with child welfare: a national survey. *J Am Acad Child Adolesc Psychiatry, 43*(8), 960-970.

Centers for Disease Control and Prevention. (1988). Operational criteria for determining suicide. *MMWR Morb Mortal Wkly Rep, 37*(50), 773-774, 779-780.

Centers for Disease Control and Prevention. (2003). *Web-based Injury Statistics Query and Reporting System (WISQARS) [Online]*. Retrieved from: www.cdc.gov/ncipc/wisqars

Centers for Disease Control and Prevention. (2020). *Youth Risk Behavior Survey Data Summary & Trends Report 2009-2019*. Retrieved from https://www.cdc.gov/healthyyouth/data/yrbs/pdf/YRBSDataSummaryTrendsReport2019-508.pdf

Centers for Disease Control and Prevention. (2021). *Data and Statistics on Children's Mental Health.* Retrieved from https://www.cdc.gov/childrensmentalhealth/data.html

Connell, A. M., McKillop, H. N., & Dishion, T. J. (2016). Long-Term Effects of the Family Check-Up in Early Adolescence on Risk of Suicide in Early Adulthood. *Suicide Life Threat Behav, 46 Suppl 1*, S15-22.
Curtin, S. C. (2020). State Suicide Rates Among Adolescents and Young Adults Aged 10-24: United States, 2000-2018. *Natl Vital Stat Rep, 69*(11), 1-10.
Czyz, E. K., King, C. A., & Nahum-Shani, I. (2018). Ecological assessment of daily suicidal thoughts and attempts among suicidal teens after psychiatric hospitalization: Lessons about feasibility and acceptability. *Psychiatry Res, 267*, 566-574.
Degenhardt, F., Seifert, S., & Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform, 20*(2), 492-503.
DelPozo-Banos, M., John, A., Petkov, N., Berridge, D. M., Southern, K., K, L. L., . . . Travieso, C. M. (2018). Using Neural Networks with Routine Health Records to Identify Suicide Risk: Feasibility Study. *JMIR Ment Health, 5*(2), e10144.
Diamond, G. S., Wintersteen, M. B., Brown, G. K., Diamond, G. M., Gallop, R., Shelef, K., & Levy, S. (2010). Attachment-based family therapy for adolescents with suicidal ideation: a randomized controlled trial. *J Am Acad Child Adolesc Psychiatry, 49*(2), 122-131.
Dowd, K., Dolan, M., Smith, K., Orin, D., Keeney, J., Wheeless, S., & Biemer, P. (2014). *NSCAW II Combined Waves 1-3 Restricted Release Version Data File User's Manual*. Ithaca, NY: National Data Archive on Child Abuse and Neglect, Cornell University.
Dugoff, E. H., Schuler, M., & Stuart, E. A. (2014). Generalizing observational study results: applying propensity score methods to complex surveys. *Health Serv Res, 49*(1), 284-303.
Duprey, E. B., Oshri, A., & Liu, S. (2020). Developmental pathways from child maltreatment to adolescent suicide-related behaviors: The internalizing and externalizing comorbidity hypothesis. *Dev Psychopathol, 32*(3), 945-959.
Esposito-Smythers, C., Spirito, A., Kahler, C. W., Hunt, J., & Monti, P. (2011). Treatment of co-occurring substance abuse and suicidality among adolescents: a randomized trial. *J Consult Clin Psychol, 79*(6), 728-739.
Evans, R., White, J., Turley, R., Slater, T., Morgan, H., Strange, H., & Scourfield, J. (2017). Comparison of suicidal ideation, suicide attempt and suicide in children and young people in care and non-care populations: Systematic review and meta-analysis of prevalence. *Child Youth Serv Rev, 82*, 122-129.
Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., . . . Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychol Bull, 143*(2), 187-232.
Fulginiti, A., He, A. S., & Negriff, S. (2018). Suicidal because I don't feel connected or vice versa? A longitudinal study of suicidal ideation and connectedness among child welfare youth. *Child Abuse Negl, 86*, 278-289.
Gabrielli, J., Hambrick, E. P., Tunno, A. M., Jackson, Y., Spangler, A., & Kanine, R. M. (2015). Longitudinal Assessment of Self-Harm Statements of Youth in Foster Care: Rates, Reporters, and Related Factors. *Child Psychiatry Hum Dev, 46*(6), 893-902.
Garcia de la Garza, A., Blanco, C., Olfson, M., & Wall, M. M. (2021). Identification of Suicide Attempt Risk Factors in a National US Survey Using Machine Learning. *JAMA Psychiatry*.
Goldstein, T. R., Ha, W., Axelson, D. A., Goldstein, B. I., Liao, F., Gill, M. K., . . . Birmaher, B. (2012). Predictors of prospectively examined suicide attempts among youth with bipolar disorder. *Arch Gen Psychiatry, 69*(11), 1113-1122.
Gradus, J. L., Rosellini, A. J., Horvath-Puho, E., Street, A. E., Galatzer-Levy, I., Jiang, T., . . . Sorensen, H. T. (2020). Prediction of Sex-Specific Suicide Risk Using Machine Learning and Single-Payer Health Care Registry Data From Denmark. *JAMA Psychiatry, 77*(1), 25-34.

Harrell, F. E., Jr., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med, 15*(4), 361-387.
Hawton, K., Saunders, K. E., & O'Connor, R. C. (2012). Self-harm and suicide in adolescents. *Lancet, 379*(9834), 2373-2382.
Heneghan, A., Stein, R. E., Hurlburt, M. S., Zhang, J., Rolls-Reutz, J., Fisher, E., . . . Horwitz, S. M. (2013). Mental health problems in teens investigated by U.S. child welfare agencies. *J Adolesc Health, 52*(5), 634-640.
Heneghan, A., Stein, R. E., Hurlburt, M. S., Zhang, J., Rolls-Reutz, J., Kerker, B. D., . . . Horwitz, S. M. (2015). Health-risk behaviors in teens investigated by U.S. Child Welfare Agencies. *J Adolesc Health, 56*(5), 508-514.
Hill, R. M., Oosterhoff, B., & Kaplow, J. B. (2017). Prospective identification of adolescent suicide ideation using classification tree analysis: Models for community-based screening. *J Consult Clin Psychol, 85*(7), 702-711.
Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*, 199-236.
Huang, X., Ribeiro, J. D., & Franklin, J. C. (2020). The Differences Between Individuals Engaging in Nonsuicidal Self-Injury and Suicide Attempt Are Complex (vs. Complicated or Simple). *Front Psychiatry, 11*, 239.
Huque, M. H., Carlin, J. B., Simpson, J. A., & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Med Res Methodol, 18*(1), 168.
Ilgen, M. A., Downing, K., Zivin, K., Hoggatt, K. J., Kim, H. M., Ganoczy, D., . . . Valenstein, M. (2009). Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. *J Clin Psychiatry, 70*(11), 1495-1500.
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer Texts in Statistics.
Janitza, S., Celik, E., & Boulesteix, A.-L. (2018). A computationally fast variable importance test for random forests for high-dimensional data. *Advances in Data Analysis and Classification, 12*(4), 885-915.
Joiner, T. E. (2005). *Why people die by suicide*. Cambridge, MA: Harvard University Press.
Karmakar, C., Luo, W., Tran, T., Berk, M., & Venkatesh, S. (2016). Predicting Risk of Suicide Attempt Using History of Physical Illnesses From Electronic Medical Records. *JMIR Ment Health, 3*(3), e19.
Karpievitch, Y. V., Hill, E. G., Leclerc, A. P., Dabney, A. R., & Almeida, J. S. (2009). An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. *PLoS One, 4*(9), e7087.
Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status. *Am J Psychiatry, 159*(9), 1548-1555.
Kessler, R. C., Hwang, I., Hoffmire, C. A., McCarthy, J. F., Petukhova, M. V., Rosellini, A. J., . . . Bossarte, R. M. (2017). Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *Int J Methods Psychiatr Res, 26*(3).
Kessler, R. C., Stein, M. B., Petukhova, M. V., Bliese, P., Bossarte, R. M., Bromet, E. J., . . . Army, S. C. (2017). Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry, 22*(4), 544-551.
Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., . . . Army, S. C. (2015). Predicting suicides after psychiatric hospitalization in US Army soldiers: the

Army Study To Assess Risk and rEsilience in Servicemembers (Army STARRS). *JAMA Psychiatry, 72*(1), 49-57.
Kidd, S., Henrich, C. C., Brookmeyer, K. A., Davidson, L., King, R. A., & Shahar, G. (2006). The social context of adolescent suicide attempts: interactive effects of parent, peer, and school social relations. *Suicide Life Threat Behav, 36*(4), 386-395.
Klaus, N. M., Mobilio, A., & King, C. A. (2009). Parent-adolescent agreement concerning adolescents' suicidal thoughts and behaviors. *J Clin Child Adolesc Psychol, 38*(2), 245-255.
Leavitt, R. A., Ertl, A., Sheats, K., Petrosky, E., Ivey-Stephenson, A., & Fowler, K. A. (2018). Suicides Among American Indian/Alaska Natives - National Violent Death Reporting System, 18 States, 2003-2014. *MMWR Morb Mortal Wkly Rep, 67*(8), 237-242.
Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. *PLoS One, 6*(3), e18174.
Lee, J., & Little, T. D. (2017). A practical guide to propensity score analysis for applied clinical research. *Behav Res Ther, 98*, 76-90.
Leslie, L. K., Hurlburt, M. S., James, S., Landsverk, J., Slymen, D. J., & Zhang, J. (2005). Relationship between entry into child welfare and mental health service use. *Psychiatr Serv, 56*(8), 981-987.
Leslie, L. K., James, S., Monn, A., Kauten, M. C., Zhang, J., & Aarons, G. (2010). Health-risk behaviors in young adolescents in the child welfare system. *J Adolesc Health, 47*(1), 26-34.
Lipari, R. N., Sarra, H., Blau, G., & Rubenstein, L. (2016). *Adolescent mental health service use and reasons for using services in specialty, educational, and general medical settings*. Retrieved from https://www.samhsa.gov/data/sites/default/files/report_1973/ShortReport-1973.html
Maalouf, F. T., Atwi, M., & Brent, D. A. (2011). Treatment-resistant depression in adolescents: review and updates on clinical management. *Depress Anxiety, 28*(11), 946-954.
Masi, G., & Liboni, F. (2011). Management of schizophrenia in children and adolescents: focus on pharmacotherapy. *Drugs, 71*(2), 179-208.
Masi, G., Perugi, G., Millepiedi, S., Mucci, M., Pfanner, C., Berloffa, S., . . . Akiskal, H. S. (2010). Pharmacological response in juvenile bipolar disorder subtypes: A naturalistic retrospective examination. *Psychiatry Res, 177*(1-2), 192-198.
Miller, A. B., Esposito-Smythers, C., Weismoore, J. T., & Renshaw, K. D. (2013). The relation between child maltreatment and adolescent suicidal behavior: a systematic review and critical examination of the literature. *Clin Child Fam Psychol Rev, 16*(2), 146-172.
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Ruch, D. A., Steelesmith, D. L., Warner, L. A., Bridge, J. A., Campo, J. V., & Fontanella, C. A. (2021). Health Services Use by Children in the Welfare System Who Died by Suicide. *Pediatrics, 147*(4), e2020011585.
Ryu, S., Lee, H., Lee, D. K., & Park, K. (2018). Use of a Machine Learning Algorithm to Predict Individuals with Suicide Ideation in the General Population. *Psychiatry Investig*, *15*(11), 1030-1036.
Sandri, M., & Zuccolotto, P. (2008). A bias correction algorithm for the Gini variable importance measure in classification trees. *J Computational Graphical Stats, 17*(3), 611-628.
Sellers, C. M., McRoy, R. G., & O'Brien, K. H. M. (2019). Substance use and suicidal ideation among child welfare involved adolescents: A longitudinal examination. *Addict Behav, 93*, 39-45.
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu Rev Clin Psychol, 4*, 1-32.

Simon, G. E., Johnson, E., Lawrence, J. M., Rossom, R. C., Ahmedani, B., Lynch, F. L., . . . Shortreed, S. M. (2018). Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records. *Am J Psychiatry, 175*(10), 951-960.
Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics, 8*, 25.
Timmermans, S. (2005). Suicide determination and the professional authority of medical examiners. *Am Socio Rev, 70*(2), 311-333.
Tran, T., Luo, W., Phung, D., Harvey, R., Berk, M., Kennedy, R. L., & Venkatesh, S. (2014). Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry, 14*, 76.
Turecki, G., & Brent, D. A. (2016). Suicide and suicidal behaviour. *Lancet, 387*(10024), 1227-1239.
Valdiviezo, H. C., & Aelst, S. V. (2015). Tree-based prediction on incomplete data using imputation or surrogate decisions. *Information Sciences, 311*, 163-181.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *J Statistical Software, 45*(3).
Verona, E., Sachs-Ericsson, N., & Joiner, T. E., Jr. (2004). Suicide attempts associated with externalizing psychopathology in an epidemiological sample. *Am J Psychiatry, 161*(3), 444-451.
Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci, 5*(3), 457-469.
Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *J Child Psychol Psychiatry, 59*(12), 1261-1270.
Wanner, B., Vitaro, F., Tremblay, R. E., & Turecki, G. (2012). Childhood trajectories of anxiousness and disruptiveness explain the association between early-life adversity and attempted suicide. *Psychol Med, 42*(11), 2373-2382.
Wasserman, D., Hoven, C. W., Wasserman, C., Wall, M., Eisenberg, R., Hadlaczky, G., . . . Carli, V. (2015). School-based suicide prevention programmes: the SEYLE cluster-randomised, controlled trial. *Lancet, 385*(9977), 1536-1544.
Yap, M. B., & Jorm, A. F. (2015). Parental factors associated with childhood anxiety, depression, and internalizing problems: a systematic review and meta-analysis. *J Affect Disord, 175*, 424-440.

Appendix A – NSCAW II Variables

Note: the information provided here is a summary of the information provided in the NSCAW II Restricted Release Data File User's Manual Appendix 2. Interested readers are encouraged to consult the official DFUM for full details.

VARIABLES COLLECTED FROM CHILD INTERVIEW

Construct	**Measure**	**Information Gathered**
Child characteristics	NSCAW-developed questions	Child's demographic information, BMI
Developmental/Cognitive status	Kaufman Brief Intelligence Test (K-BIT): Expressive Vocabulary, Definitions, and Matrices	Standardized assessment tool comprised of two subsets: Vocabulary (expressive vocabulary and definitions) & Matrices (ability to perceive relationships & complete analogies)
Academic achievement	Woodcock-Johnson III Tests of Achievement (WJ-III)	Standardized test of academic achievement; four subtests used: Letter-Word Comprehension, Passage Comprehension, Calculation, and Applied Problems
School engagement	Drug Free Schools (DFSCA) Outcome study questions	School achievement; student's disposition towards learning and school
Peer relationships, including social rejection	Loneliness and Social Dissatisfaction Questionnaire for Young Children	Success in making and keeping friendships; school adjustment
Protective factors	Resiliency Scale - LongSCAN	Resources that a child has that facilitate resiliency
Parental monitoring	Supervision-Child Scale from Fast Track Project	Extent to which the caregiver monitors the child's activities
Independent Living	NSCAW-developed questions	Life skills the youth may have developed where (s)he learned the skills
Satisfaction with caseworker services	NSCAW-developed questions	Degree of satisfaction with CPS caseworker services
Future expectations	Expectations About Employment, Education, and Life Span section from the Adolescent Health Survey (adapted)	Expectations related to child's life experiences
Mental health - depression	Children's Depression Inventory	All aspects of well-being, including behavior problems

Mental health - trauma	Trauma Symptom Checklist for Children – PTSD section (adapted)	Indicators of post-traumatic stress disorder
Participation in activities	Youth Self Report – Social Competence Scale	Involvement in activities which may promote social skills or cognitive development
Behavior problems	Youth Self Report – Syndrome and Total Problems Scale	Magnitude of aggressive behavior and impulse control
Relationship with parents and other significant adults	Rochester Assessment Package for Schools (RAPS) -and- National Longitudinal Study of Adolescent Health, In-Home questionnaire	Degree of supportive relationships between chil and adult
Loss, violence, and other stressors in and out of the home	Violence Exposure Scale (VEX-R) – Home Set	Violence observed and experienced in the home
CPS services received	NSCAW-developed questions	Factors the affect the CPS service provision process
Substance abuse	Youth Risk Behavior Survey (YRBS) -and- The CRAFFT -and- National Longitudinal Study of Adolescent Health	Misuse of controlled substances as associated with depression and maltreatment
Sexual behavior	LongSCAN	Early sexual activity
Delinquency	Modified Self Report of Delinquency -and- Denver Youth Survey	Participation in delinquent or criminal activities
Maltreatment	Parent-Child Conflict Tactics Scale (adapted)	Additional maltreatment information in order to better understand the effects of the severity and specific type of abuse
Deviant peer affiliation	Deviant Peer Affiliation Scale	Involvement with peers who engage in risky or deviant behaviors
Maltreatment - injuries	Child Health and Illness Profile – Adolescent Edition, Injury questions	Nature and extent of injuries in the past 12 months

VARIABLES COLLECTED FROM CAREGIVER INTERVIEW

Construct	Measure	Information Gathered
Family composition and demographics	NSCAW-developed questions	Family composition and demographic information
Neighborhood factors	Philadelphia Family Management Study Parent Interview Schedule (adapted)	Behavior of individuals and families in terms of the environment of their community
Health and disabilities services received by child	Child and Adolescent Services Assessment (CASA) -and- National Evaluation of Family Support Programs, Child Health Questionnaire -and- Questionnaire for Identifying Children with Chronic Conditions – Revised (QuICCC-R) -and- National Survey of Children with Special Health Care Needs II (SLAITS) -and- Services Assessment for Children and Adolescents (SACA) -and- The National Early Intervention Longitudinal Study (NEILS) -and- National Comorbidity Study (NCS) -and- National Health Interview Survey (NHIS) -and- National Survey of America's Families (NSAF) -and- NSCAW-developed questions	History of health, injury, and disability status of child; services received by child

Adaptive skills	Vineland Adaptive Behavior Scale (VABS) Screener – Daily Living Skills and Socialization Skills	Regular behaviors the child exhibits
Global social competence	Social Skills Rating System – Social Skills Scale	Level of development of social skills possessed by the child
Behavior problems	Child Behavior Checklist	Degree to which child exhibits different types of behaviors
Income	NSCAW-developed questions	Financial resources available to the child's household
Services received by caregiver	NSCAW-developed questions	Frequency and duration that services were received
Social support and other family resources, including assistance with child-rearing	Duke-UNC Functional Social Support Questionnaire (adapted)	Perceived social support for child and family
Physical health	Short-Form Health Survey (SF-12)	Caregiver's physical health status
Mental health - depression	Composite International Diagnostic Interview Short Form (CIDI-SF)	Caregiver experiences that indicate symptoms of depression
Alcohol dependence	The Alcohol Use Disorders Identification Test (AUDIT)	Caregiver symptoms that indicate symptoms of alcohol dependence
Drug dependence	Drug Abuse Screening Test (DAST)	Caregiver symptoms that indicate symptoms of drug dependence
Criminal involvement of parents	NSCAW-developed questions	Caregiver criminal history and involvement with the criminal justice system
Behavioral monitoring and discipline	Parent-Child Conflict Tactics Scale (CTSPC) Neglect and Substance Abuse questions	Methods and frequency of discipline measures used by the caregivers with the child during the last 12 months
Domestic violence in the home	Conflict Tactics Scale (CTS2) – Physical Assault Subscale	Type and frequency of violence occurring in the home and directed toward female caregiver in the last 12 months, and subsequent use of services
Satisfaction with caseworker	NSCAW-developed questions	Satisfaction level with services received from CPS caseworker

VARIABLES COLLECTED FROM CPS CASEWORKER INTERVIEW

Construct	Measure	Information Gathered
Case investigation	NSCAW-developed questions	Circumstances surrounding the investigation report; background of the caseworker
Alleged abuse	Modified Maltreatment Classification System (MMCS)	Details about the specific nature of the alleged abuse or neglect
Risk assessment	NSCAW-developed questions	Factors determining case decisions, including prior history of abuse or neglect, caregiver substance abuse, domestic violence in the home, caregiver mental health problems, poor parenting skills, excessive discipline
History since case report	NSCAW-developed questions	Child's history with the child welfare system since the case report that resulted in the child's selection for NSCAW
Caseworker involvement with the child/family	NSCAW-developed questions	Caseworker's individual involvement with case, including referrals made for family members, caseworker contact with siblings, number of contacts with service providers and family, and attitudes about service to family
Services to parents	NSCAW-developed questions	Service referrals and status for parents
Services to child	NSCAW-developed questions	Service referrals and status for child
Caseworker background	NSCAW-developed questions	Demographic information about caseworker, employment and education history, and attitudinal questions about work
Climate of CPS agency	NSCAW-developed questions	Culture, climate, and social context of the CPS agency

Geoffrey D. Kahn

PERSONAL DATA

Johns Hopkins Bloomberg School of Public Health
624 N Broadway, office 802
Baltimore, MD 21205

Phone: (301) 704-7409
E-mail: gkahn@jhu.edu

EDUCATION AND TRAINING

Doctor of Philosophy/2021 — Johns Hopkins University, Mental Health

Master of Science in Public Health/2009 — Emory University, Global Epidemiology

Bachelor of Science/2006 — Emory University, Biology, Spanish

PROFESSIONAL EXPERIENCE

Research Associate: May 2014-Sept. 2016 — Johns Hopkins University, Moore Center for the Prevention of Child Sexual Abuse

- Maintained databases for digital survey responses administered through SurveyGizmo
- Cleaned data for internal use and external partners, datasets of up to hundreds of thousands of observations and dozens of variables
- Performed data analysis for multiple projects examining the effect of juvenile sex offender registration, including simple tests of association, and ARIMA and survival analysis modeling
- Performed ad hoc analyses for a variety of projects, including descriptive statistics for interim progress reports to funders and power/sample size calculations for grant applications

Research Associate: Aug. 2010-May 2014 — Johns Hopkins University, International Vaccine Access Center

- Epidemiologic data analysis, literature review and manuscript preparation for various projects
- SAS programming for analyses related to the PERCH multi-center pneumonia etiology study

- Data analyst and technical advisor for Hib meningitis surveillance collaborators in India, included regular travel to India
- Preparation of training and advocacy materials for vaccine advocacy efforts in India, including local expert meetings and professional training courses
- Agreement for the Performance of Work with WHO Mar-Apr 2013 to prepare a review on Hib combination vaccines vs. separately administered vaccines in preparation for the SAGE April 2013 meeting

Microbiologist: Oct. 2009-Aug. 2010 Atlanta Research and Education Foundation
Contracted to Centers for Disease Control (CDC)

- Performed molecular analysis of enteric parasites (giardia, cryptosporidia, entamoeba) from humans and animals
- Trained new lab members on molecular techniques and procured their government credentials
- Maintained laboratory sample databases, shipped biological samples to collaborators, ordered supplies as needed, updated lab SOPs

Research Program Manager: June 2009-Sep. 2009 Emory University
Center for Global Safe Water

- Provided technical support to three non-governmental organizations in Mexico on data collection for use in monitoring & evaluation of school-based water and hygiene interventions
- Produced new training materials and translated existing survey instruments
- Traveled to Mexico to meet with each group and assist with the design of monitoring plans
- Attended the Regional Symposium on Water, Sanitation and Hygiene in Schools in Managua, Nicaragua

Workstudy Student: Aug. 2007-May 2009 CDC

- Performed molecular analysis of enteric parasites (giardia, cryptosporidia, entamoeba) from humans and animals
- Contributed to general laboratory maintenance

Emerging Infectious Diseases Fellow: Aug. 2006-Aug. 2007 Association of Public Health Laboratories (APHL), at CDC

- Molecular analysis of giardia genotypes infecting children in peri-urban slums outside Lima
- Traveled to Lima, Peru for 6 weeks to develop a PCR diagnostic assay for congenital Chagas disease

EDITORIAL ACTIVITIES

Peer Review Activities

Archives of Suicide Research x1
International Journal of Environmental Research and Public Health x3
Journal of Affective Disorders x1
Journal of Tropical Pediatrics x3
Vaccine x2 (plus published commentary and author response on 1 article)

HONORS AND AWARDS

Honors

2020 – Sigma Xi, The Scientific Research Honor Society, associate member

Awards

2020 – Dr. Ali Kawi Doctoral Scholarship in Mental Health Research,
Dept of Mental Health, Johns Hopkins Bloomberg School of Public Health

PUBLICATIONS

Journal Articles

Ballweber LR, Xiao L, Bowman DD, **Kahn G**, Cama VA. “Giardiasis in dogs and cats: an update on epidemiology and public health significance” Trends Parasitol. 2010 Apr;26(4):180-9.

Nundy S, Gilman RH, Xiao L, Cabrera L, Cama R, Ortego YR, **Kahn G**, Cama VA. “Wealth and its associations with enteric parasitic infections in a low-income community in Peru: Use of principal component analysis” Am J Trop Med Hyg. 2011 Jan;84(1):38-42.

Velasquez DE, Arvelo W, Cama VA, López B, Reyes L, Roellig DM, **Kahn GD**, Lindblade KA. “Molecular Insights for Giardia, Cryptosporidium, and Soil-Transmitted Helminths from a Facility-Based Surveillance System in Guatemala.” Am J Trop Med Hyg. 2011 Dec;85(6):1141-3.

Kahn GD, Fitzwater SP, Tate JE, Kang G, Ganguly NK, Nair GB, Steele D, Arora R, Chawla-Sarkar M, Parashar U, Santosham M. “Epidemiology and Prospects for Prevention of Rotavirus Disease in India” Indian Pediatrics. 2012 June;49(6):467-74.

A Chandran, **G Kahn**, T Sousa, F Pechansky, DM Bishai, AA Hyder. “Impact of Road Traffic Deaths on Expected Years of Life Lost and Reduction in Life Expectancy in Brazil” Demography. 2013 Feb;50(1):229-36.

H Nair, E Simões, I Rudan, et al. for the **Severe ALRI Working Group**. “Global and regional estimates of hospitalised severe acute lower respiratory infections in young children in 2010: a systematic analysis” Lancet. 2013 Apr;381(9875):1380-90.

S Fitzwater, P Ramachandran, K Nedunchelian, **G Kahn**, M Santosham, A Chandran. “Bacterial meningitis in children <2 years of age in a tertiary care hospital in South India: an assessment of clinical and laboratory features.” J Pediatrics. 2013 Jul;163(Suppl 1):S32-7.

S Dwivedi, B Das, S Aneja, S Sharma, M Chaturvedi, **G Kahn**, S Fitzwater, A Chandran, N Wadhwa, S Bhatnagar. “Group B streptococcal meningitis in infants beyond the neonatal period.” Indian J Pediatr. 2014 Jan; 81(1):4-8.

G Kahn, D Thacker, S Nimbalkar, M Santosham. “High cost is the primary barrier reported by physicians who prescribe vaccines not included in India’s Universal Immunization Program.” J Trop Pediatr, 2014 Aug; 60(4):287-91.

A Manoharan, V Manchanda, B Sundaram, S Lalwani, M Modak, S Bai, A Vijayan, A Shet, S Nagaraj, S Karande, G Nataraj, VN Yewale, SA.Joshi, RN Iyer, M Santosham, **G Kahn**, M Knoll and the Alliance for Surveillance of Invasive Pneumococci (ASIP) Study Group. "Invasive Pneumococcal Disease in Children under 5 years in India: An Observational study (2011-2015)." Lancet Infect Dis. 2017 Mar; 17(3):305-12.

Morpeth SC, Deloria Knoll M, Scott JAG, Park DE, Watson NL, Baggett HC, Brooks WA, Feikin DR, Hammitt LL, Howie SRC, Kotloff KL, Levine OS, Madhi SA, O'Brien KL, Thea DM, Adrian PV, Ahmed D, Antonio M, Bunthi C, DeLuca AN, Driscoll AJ, Githua LP, Higdon MM, **Kahn G**, Karani A, Karron RA, Kwenda G, Makprasert S, Mazumder R, Moore DP, Mwansa J, Nyongesa S, Prosperi C, Sow SO, Tamboura B, Whistler T, Zeger SL, Murdoch DR, The PERCH Study Group. "Detection of pneumococcal DNA in blood by PCR for diagnosing pneumococcal pneumonia in young children from low and middle income countries." Clin Infect Dis 2017; 64(suppl 3):S347–56.

Higdon MM, Le T, O'Brien KL, Murdoch DR, Prosperi C, Baggett HC, Brooks WA, Feikin DR, Hammitt LL, Howie SRC, Kotloff KL, Levine OS, Scott JAG, Thea DM, Awori JO, Baillie VL, Cascio S, Chuananon S, DeLuca AN, Driscoll AJ, Ebruke BE, Endtz HP, Kaewpan A, **Kahn G**, Karani A, Karron RA, Moore DP, Park DE, Rahman MZ, Salaudeen R, Seidenberg P, Somwe SW, Sylla M, Tapia MD, Zeger SL, Deloria Knoll M, Madhi SA, The PERCH Study Group. "Association of C-reactive protein (CRP) with bacterial and respiratory syncytial virus associated pneumonia among children aged <5 years in the PERCH Study." Clin Infect Dis 2017; 64(suppl 3):S378–86.

Letourneau EJ, Harris AJ, Shields RT, Walfield SM, Ruzicka AE, Buckman C, **Kahn GD**, Nair R. "Effects of juvenile sex offender registration on adolescent well-being: an empirical examination." Psychology, Public Policy, and Law 2017; 24(1):105-17.

Feder KA, McCart MR, **Kahn GD**, Mauro PM, Sheidow AJ, Letourneau EJ. "Association of mental health symptoms and peer behaviors with risk for substance use and condomless sex among youth in juvenile drug court." J Child Adol Sub Abuse 2018; 27(3):133-45.

Letourneau EJ, Shields RT, Nair R, **Kahn G**, Sandler JC, Vandiver DM. "Juvenile registration and notification policies fail to prevent first-time sexual offenses: an extension of findings to two new states." Criminal Justice Policy Review 2019; 30(7):1109-23.

Fitzwater SP, Ramachandran P, **Kahn GD**, Nedunchelian K, Suresh S, Santosham M, Chandran A. "Impact of the introduction of *Haemophilus influenzae* type b conjugate vaccine in an urban setting in southern India." Vaccine 2019; 37(12):1608-13.

Holliday CN, **Kahn G**, Thorpe, RJ, Shah R, Hameeduddin, Z. Decker MR. "Racial/ethnic disparities in police reporting for partner violence in the national crime victimization survey and survivor-led interpretation." J Racial Ethnic Health Disparities 2020; 7(3):468-80.

Shields RT, Murray SM, Ruzicka AE, Buckman C, **Kahn G**, Benelmouffok A, Letourneau EJ. "Help wanted: lessons on prevention from young adults with a sexual interest in prepubescent children." Child Abuse & Neglect 2020; 105:104416.

Kahn G & Wilcox, H. "Marijuana use is associated with suicidal ideation and behavior among US adolescents at rates similar to tobacco and alcohol." **published online Aug 11, 2020,** ***Archives of Suicide Research***

D'Agati D, **Kahn G**, Swartz K. "Preteen behaviors and sexual orientation of US high school students who report depressive symptoms, 2015-2017." Public Health Rep 2021; 136(2):132-35.

Horowitz L, **Kahn G**, Wilcox H. "The Urgent Need to Recognize and Reduce Risk of Suicide for Children in the Welfare System." **published online Mar 8, 2021,** ***Pediatrics***

Rabinowitz J, Reboussin B, Thrul J, Drabick D, **Kahn G**, Green K, Ialongo N, Huhn A, Maher B. "Early Childhood Behavioral and Academic Antecedents of Lifetime Opioid Misuse among Urban Youth." **published online Mar 10, 2021,** ***J Clin Child Adolesc Psychol***

Chan CK, Sieber FE, Blennow K, Inouye SK, **Kahn G**, Leoutsakos J-M, Marcantonio ER, Neufeld KJ, Rosenberg PB, Wang N-Y, Zetterberg H, Lyketsos CG, Oh ES. "Association of depressive symptoms with postoperative delirium and CSF AD biomarkers among hip fracture patients." **revise & resubmit, Am J Geriatr Psychiatry**

Rabinowitz J, Jin J, **Kahn G**, Kuo S, Campos A, Renteria M, Benke K, Wilcox H, Ialongo N, Maher B, Eaton W, Uhl G, Wager B, Cohen D. "Genetic Propensity for Risky Behavior and Depression Predicts Suicide Attempt Among Urban African American Males." **revise & resubmit, Am J Med Genetics**

Book Chapters

Naveen Thacker, Lois Privor-Dumm, Ranjana Kumar, **Geoff Kahn**. Role of GAVI in the Control of Vaccine Preventable Diseases in India in "Textbook of Pediatric Infectious Diseases" A Parthasarathy ed. Jaypee Brothers Medical Publishers Ltd, New Delhi. 2013

G Kahn, S Fitzwater, A Chandran, M Santosham. *Haemophilus influenzae* Type B Vaccines in "IAP Textbook of Vaccines" V Vashishtha ed. Jaypee Brothers Medical Publishers Ltd, New Delhi, 2014. & revised for 2nd ed. 2020

S Fitzwater, **G Kahn**, A Chandran, M Santosham. *Haemophilus influenzae* Type b Vaccines in "Plotkin's Vaccines" 8th ed. Elsevier, 2021

TEACHING

Classroom Instruction (Teaching Assistant)

Introduction to Behavioral and Psychiatric Genetics (330.612)
academic year 2017-18

Statistics for Psychosocial Research: Measurement (330.657.01)
academic years 2018-19, 2019-20, 2020-21

Statistics for Psychosocial Research: Structural Models (140.658.01)
academic years 2018-19, 2019-20, 2020-21

Suicide Prevention: Problem Solving Seminar (330.675.81)
academic years 2018-19, 2019-20

Other Significant Teaching

Guest lecturer for "Hib Vaccines" lecture for Child and Public Health in the Tropics (223.686), JHSPH Winter Institute 2013, 2014, and Summer Institute 2013, 2014

Gave an invited lecture at the East Carolina University Brody School of Medicine, "Suicide Etiology and Risk in Pediatric and Young Adult Populations" Sept 13, 2019

Wrote one question and accompanying grading rubric for the DrPH comprehensive written exam, academic year 2019-20

Departmental statistics tutor for Dept of Mental Health Masters students, academic year 2020-21

RESEARCH GRANT PARTICIPATION

Emory Global Field Experience (internal Emory University SPH thesis/practicum grant)
Principal Investigator: Michael Levy, PhD
Funding Level: $800

Predicting Suicide Attempts in Youth with Child Protective Services Contact (1F31MH120973-01)
Role: Principle Investigator
Funding Level: $45,016 for 2 years

PRESENTATIONS

2nd International Conference on Giardia and Cryptosporidium (2007)
  Gave 10 minute oral presentation "Genetic Diversity of *Giardia duodenalis* in Peruvian Children"

58th Annual Meeting of the American Society of Tropical Medicine and Hygiene (2009)
  Presented poster "Relationship Between Socio-Economic Factors and Time-to-Infection With *Giardia intestinalis* Among Children in Peru"

3rd Annual Johns Hopkins Vaccine Initiative Vaccine Day (2010)
  Presented poster "Multicenter bacterial meningitis sentinel surveillance in Indian infants"

8th International Symposium on Pneumococci and Pneumococcal Diseases (2012)
  Presented poster "Comprehensive review characterizing the epidemiology of *Streptococcus pneumoniae* in India"

9th International Symposium on Pneumococci and Pneumococcal Diseases (2014)
  Presented poster "Elevated CRP is associated with bacterial pneumonia in children, but may not distinguish pneumococcal from other bacterial pneumonias: PERCH Study"

35th Conference of the Association for the Treatment of Sexual Abusers (2016)
  Presented poster "Impact of Sex Crime Policies on Youth and Their Families"

9th Conference of the Mid-Atlantic Region chapter of ATSA (2017)
  Presented poster "Descriptive statistics from a mixed-methods study of non-offending pedophiles, and implications for treating adolescents with a sexual interest in children"

Contributor but not presenter:

National Conference on Health and Domestic Violence, virtual, April 2021.
  Holliday CN, **Kahn G**, Hameeduddin Z, Shah R, Miller J, Thorpe RJ, Dantzler J, Goodmark L, Decker MR. "Survivors describe the intersectional impact of race and gender inequities on discouraging police reporting for violence against women" [Oral Presentation]