PSYCHOLOGICAL SCIENCE

Randomized evaluation of a school-based, trauma-informed group intervention for young women in Chicago

Monica P. Bhatt[1]*, Jonathan Guryan[2], Harold A. Pollack[3,4], Juan C. Castrejon[1]†, Molly Clark[1]†, Lucia Delgado-Sanchez[1]‡, Phoebe Lin[1]†, Max Lubell[1]†, Cristobal Pinto Poehls[1]‡, Ben Shaver[1]‡, Makenzi Sumners[1]†

This study explores whether a school-based group counseling program for adolescent girls, implemented at scale, can mitigate trauma-related mental health harms. In a randomized trial involving 3749 Chicago public high school girls, we find that participating in the program for 4 months induces a 22% reduction in posttraumatic stress disorder symptoms and find significant decreases in anxiety and depression. Results surpass widely accepted cost-effectiveness thresholds, with estimated cost-utility well below $150,000 per quality adjusted life year. We find suggestive evidence that effects persist and may even increase over time. Our results provide the first efficacy trial of such a program specifically designed for girls, conducted in America's third largest city. These findings suggest the promise of school-based programs to mitigate trauma-related harms.

INTRODUCTION

Decades of psychology research establish the developmental mental health harms of trauma exposure, particularly during adolescence. The coronavirus disease 2019 (COVID-19) global pandemic has sharply increased these mental health challenges: the proportion of mental health–related emergency department visits for children aged 5 to 11 and 12 to 17 years increased approximately 24 and 31%, respectively, in 2020 compared to previous years (*1*). While these experiences are not unique to young women, their burdens are often disproportionately borne by girls, particularly Black and Latinx girls. These burdens include effects on mental health, development, and schooling attainment (*2*). Young women who have been exposed to a traumatic event such as violence are more likely to report symptoms of posttraumatic stress disorder (PTSD) than their male peers (*3*); depressive disorders are more common among women (*4*); and women are more likely than men to develop an anxiety disorder over their life spans (*5*–*7*). Exposure to trauma, through its effect on mental health, also influences academic performance and increases the likelihood of high school dropout (*8*). In Chicago, for example, students' exposure to local homicides is associated with notable reductions in academic performance and scores on standardized tests (*9*).

Despite the marked increase in trauma exposure for adolescent girls and accompanying mental health challenges, there is a dearth of evidence regarding what works to address these challenges. Upon beginning this study, the influential Blueprints for Healthy Youth Development identified only six promising or model programs to address anxiety, depression, and PTSD in adolescents; none focused on girls specifically [see, e.g., (*10*)].

Existing data also suggest that investing in programming for young women may provide benefits beyond the individuals served. For example, research suggests that programs that help males tend to primarily benefit participating individuals, whereas programs that help females yield greater benefits for families and communities (*11*).

Here, we provide evidence that it is possible to reduce PTSD, anxiety, and depression among adolescent girls through a school-based, group counseling program informed by cognitive behavioral therapy (CBT), acceptance and commitment therapy (ACT), and narrative therapy. The program that we study includes a curriculum that intentionally helps student participants to develop self-awareness, build self-esteem and self-efficacy, and enhance their own individual competencies to make positive and healthy decisions.

Students randomized to treatment received programming through Youth Guidance's Working on Womanhood (WOW) program. WOW is a school-based, trauma-informed, relationship-centered, group counseling and mentoring program designed specifically by and for Black and Latinx women to disrupt the cycle of intergenerational trauma.

WOW's 39-lesson curriculum is designed around five core themes: self-awareness, emotional intelligence, healthy relationships, visionary goal setting, and leadership. The program is delivered by master-level social workers and counselors. Each of these staff members works with 50 to 55 girls across four to five groups in a school. Groups meet once per week, during the school day. For needs that cannot be addressed in a group setting, WOW counselors provide individual counseling and referrals to other services. The positive network created by WOW is posited to improve peer relationships and social supports while strengthening girls' ties with important adults, such as parents, teachers, and counselors. The program model and curriculum were refined and revised following a formal formative evaluation in 2018 (*12*).

WOW is rooted in CBT, ACT, and narrative therapy as well as key learnings from other school-based mentoring programs. CBT helps individuals recognize and understand connections among

[1]Crime Lab and Education Lab, University of Chicago, Chicago, IL, USA. [2]School of Education and Social Policy, Northwestern University, Evanston, IL, USA. [3]Crown Family School of Social Work, Policy and Practice, University of Chicago, Chicago, IL, USA. [4]University of Chicago Health Lab, University of Chicago, Chicago, IL, USA.
*Corresponding author. Email: mbhatt@uchicago.edu
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

thoughts, feelings, and behavior as well as the context within which those thoughts, feelings, and behaviors are developed. Previous research has found that CBT is effective in providing symptomatic relief for specific psychiatric disorders, including depression (*13*–*17*), PTSD (*18*), and anxiety disorders (*19*–*21*). CBT has also been found to be effective when adapted to a school-based setting for youth exposed to trauma in urban settings (*21*). There is evidence that school-based CBT can improve school grades (*22*, *23*) and reduce negative behavioral outcomes such as violent crime arrests (*24*).

ACT is used primarily to strengthen psychological flexibility, consisting of six primary components: diffusion, acceptance, self as context, contact with the present moment, values, and committed action. Psychological flexibility promotes behavior change in the service of chosen values by connecting the present moment with what the situation affords. Narrative therapy is a psychotherapeutic approach based on the notion that people construct narratives to define themselves and give meaning to their daily experiences and life events; by supporting narrating their life experiences in richer and more gratifying ways, the individual becomes empowered to consider life alternatives not previously considered. The narrative therapy process has seven core components: collaboration, externalizing the problem, identifying empowerment, identifying values, linking past story to new narrative, inviting witness to new narrative, and documenting gains (*25*, *26*). These therapies and treatments have been used to treat depression and anxiety, with no discernible differences documented to date on individual outcomes. The WOW program draws from each therapeutic tradition to form the foundation of the program curriculum (*27*).

METHODS

Research design

To scrutinize the causal impact of WOW on mental health outcomes, we conducted a cluster-randomized controlled trial in which we randomized girls to receive an offer of the WOW program or business as usual electives and services within nonrandomly preselected neighborhood high schools. Our research team and Youth Guidance together selected 10 study site Chicago Public Schools (CPS) high schools to be invited to participate in a randomized control trial during academic year 2017–2018 and 2018–2019. Each school had sufficient enrollment to provide the necessary number of treatment and control students, could accommodate WOW, and is located in an underserved community on Chicago's south or west side that has experienced high rates of violent crime (see Fig. 1).

All schools had excess demand and eligibility for the WOW program, allowing for program slots to be offered via a fair lottery given program constraints (approximately 50 girls could be served per school). In accordance to the human subject protocol approved by the University of Chicago Institutional Review Board, consent was waived for randomization into the treatment and control assigned conditions; subsequently, students were consented to participate in the program as well as for participation in baseline and follow-up surveys.

Girls who were randomized to be offered the WOW program were contacted by Youth Guidance counselors and consented into program participation. Girls who participated joined group-based counseling sessions once per week during an elective period during schools in which WOW counselors administered the WOW curriculum and built rapport with WOW groups. Girls who were not randomized to receive an offer from WOW were eligible to receive all other status quo services and elective offerings in the building.

Sample eligibility

During the first week of academic year 2017–18, we received lists of all 9th to 11th grade girls from schools participating in the study. In practice, many of these lists came with extra information—e.g., including additional grades and/or male students—and had to be cleaned before the randomization stage. We used CPS administrative data to randomize 3749 young women who met Youth Guidance's guidelines for participating in WOW, which included: (i) at least 75% overall school attendance during 2016–2017 academic year; (ii) no specific diagnosis of significant intellectual disability, per CPS records by the end of academic year 2016–2017; and (iii) no signs of proactive aggression, self-harm, or active suicidal ideation. A student is considered to have a significant intellectual disability if they were listed in any of the following six disability categories: autistic; emotional and behavior disorder; educable mental handicap; intellectual disability – profound; severe/profound handicap; trainable mental handicap. These guidelines were intended to ensure that the eligible students had a reasonable chance of engaging with and benefiting from WOW by attending a sufficient number of sessions and by being able to productively participate in and understand group activities.

The combined roster lists contained 3875 potential candidates ready to enter the randomization stage; of those, 126 individuals (about 3.25%) were excluded from the randomization because of not meeting WOW's requirements for participation. On the basis of these requirements, we randomized >95% of all 9th to 11th grade girls in schools participating in the study.

Randomization

We randomly assigned 3749 eligible 9th, 10th, and 11th grade girls to either a treatment group or control group. Because WOW programming capacity varied by school and sometimes by grade, we carried out a random assignment conditional on school-by-grade "randomization blocks" and varied the probability of assignment to the treatment condition across randomization blocks. Within each randomization block, students were randomly selected for either the treatment group, which was offered the chance to participate in WOW, or for the control group, for whom all business as usual services and electives were available. A total of 1232 young women in the study were assigned to the treatment group; 2517 women were assigned to the control group in fall 2017.

Two additional randomization phases followed the one described above. First, we randomly selected a sample of 626 students (of the 3749 students in the study) for recruitment for a baseline survey, conducted before the beginning of WOW programming; this sample was created in a way such that the number of students assigned to the treatment group was as similar as possible to the number assigned to the control group. The final number of participants within each group has a random component to it. We can only define ex ante the desired sampling probabilities within each randomization block. When creating a random sample for the baseline survey, we used sampling probabilities such that within each block, we expected an equal number of participants landing in

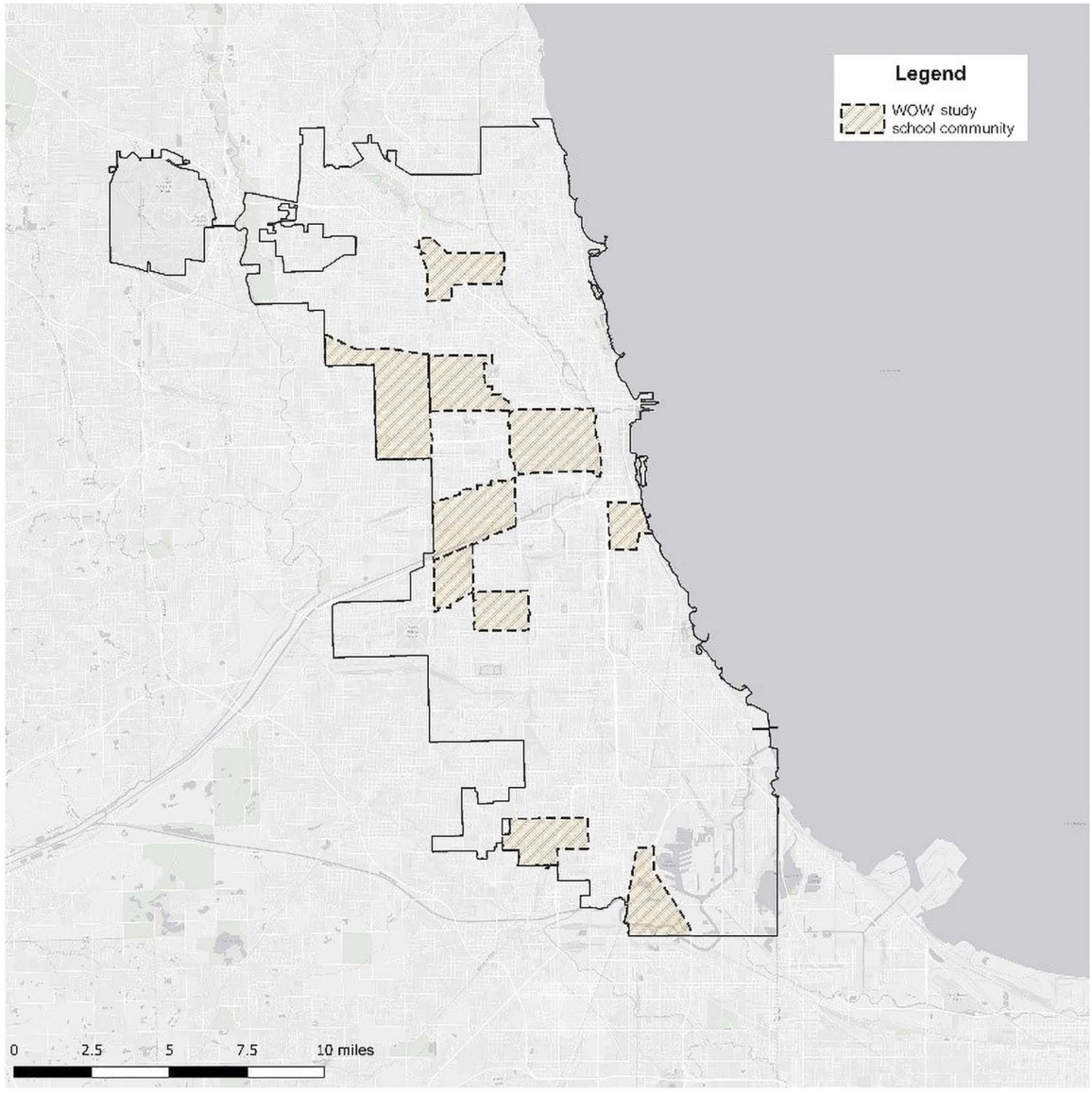

Fig. 1. WOW study school communities. WOW, Working on Womanhood.

the treatment versus control group. It just happened to be the case that we ended up with an equal number of 313 participants on each group for this case. For the follow-up survey sample, the methodology used was the same as the one used for the baseline survey sample, only that now targeting a larger sample size (~2000). The final number of participants assigned to the treatment group was that of 1008; with 1003 being assigned to the control group. From 626 randomly selected students, we had 346 responses for an effective response rate of 55%.

Second, we randomly selected a sample of 2011 students (of the 3749 students in the study) to be recruited for a follow-up survey, conducted toward the end of the first year of WOW programming; this sample was also created in a way such that the number of students in the study assigned to the treatment group was as similar as possible to the number assigned to the control group. A total of 1487 selected students responded to the follow-up survey of the 2011 offered, with an effective response rate of 84%.

We do not see differential outcome missingness for survey respondents for either survey across treatment and control groups. In the baseline survey, of the 313 students in the study assigned to the treatment group, 175 (~56%) were effectively surveyed; and of the 313 students assigned to the control group, 171 (~55%) were effectively surveyed. All other students in the study were either unable to be located or declined to participate in the baseline survey. In the case of the follow-up survey, of the 1008 students in the study assigned to the treatment group, 759 (~76%) were effectively surveyed; and of the 1003 students assigned to the control group, 728 (~73%) were effectively surveyed. Eight students assigned to the follow-up survey sample declined to participate in the study before the follow-up survey began and were excluded from the targeted sample altogether. Five of 1008 participants

assigned to the treatment group in the follow-up survey belong to the group of eight that were excluded from the follow-up survey ex ante. Within the control group, 3 of 1003 cases belong to the group of eight that were excluded from the follow-up survey ex ante. The consort diagram (see Fig. 2) shows the overall structure at each point, depicting the analytic sample for administration of a baseline survey as well as a follow-up survey (see "Data collection" section for more details on survey administration).

Analytic samples

We include analysis from two different analytic samples assigned based on the randomization procedures above here. The follow-up survey student sample is a randomly selected subset of the cohort I sample. We describe each analytic sample below, for reference:

Cohort I sample

Our cohort I sample includes girls randomized in fall 2017 (N = 3749). Eighty-five percent of young women in our study are Black and Hispanic/Latinx. About one-half of young women in the study reported speaking Spanish as their primary language at home (Table 1). The majority of students in cohort I are free and reduced price lunch recipients. We find a slightly higher proportion of students with limited English proficiency (LEP) in the treatment group and a somewhat higher proportion of youth in the control group who are flagged as having a learning disability in cohort I.

Follow-up survey sample: Selected students

The first follow-up survey sample included 2011 study youth randomly sampled for survey participation. Girls selected to be part of the follow-up survey sample were representative of girls in the cohort I sample (Table 1). Of the 2011 youth, we found that eight girls had declined survey participation at baseline and were excluded from the study; therefore, 1003 treatment youth and 1000 control youth were assigned to the survey. We find that the randomly selected follow-up survey sample students are similar to the cohort I sample and are balanced across treatment and control groups (Table 1).

Follow-up survey sample: Respondents

In addition, we find a balanced sample among follow-up survey respondents as well (Table 2). A total of 2003 students in the study were offered the opportunity to complete a follow-up survey between June and November 2018. We received 1487 completed surveys, with an effective response rate of 84%, with 759 treatment youth and 728 control youth completing the survey. A total of 1487 participants responded to at least one question of either the Student Health Survey (SHS) or Behavioral Assessment System for Children (BASC-3) questionnaires. We received 1484 completed surveys. Nonresponders in the follow-up were more likely to be Hispanic, averaged an additional day of unexcused absence, and had slightly lower grade point averages (GPAs) than respondents did. We find balance on observed covariates within this sample, with no statistically significant differences in response rates across students assigned to treatment and control (Table 2).

Data collection

Survey measures

In consultation with Youth Guidance and our team's expert mental health researcher (K. Grant, DePaul University), we compiled a comprehensive survey that contained several rigorous diagnostic instruments. We selected the nationally normed and clinically validated BASC-3 (*28*) to measure anxiety and depression (among several

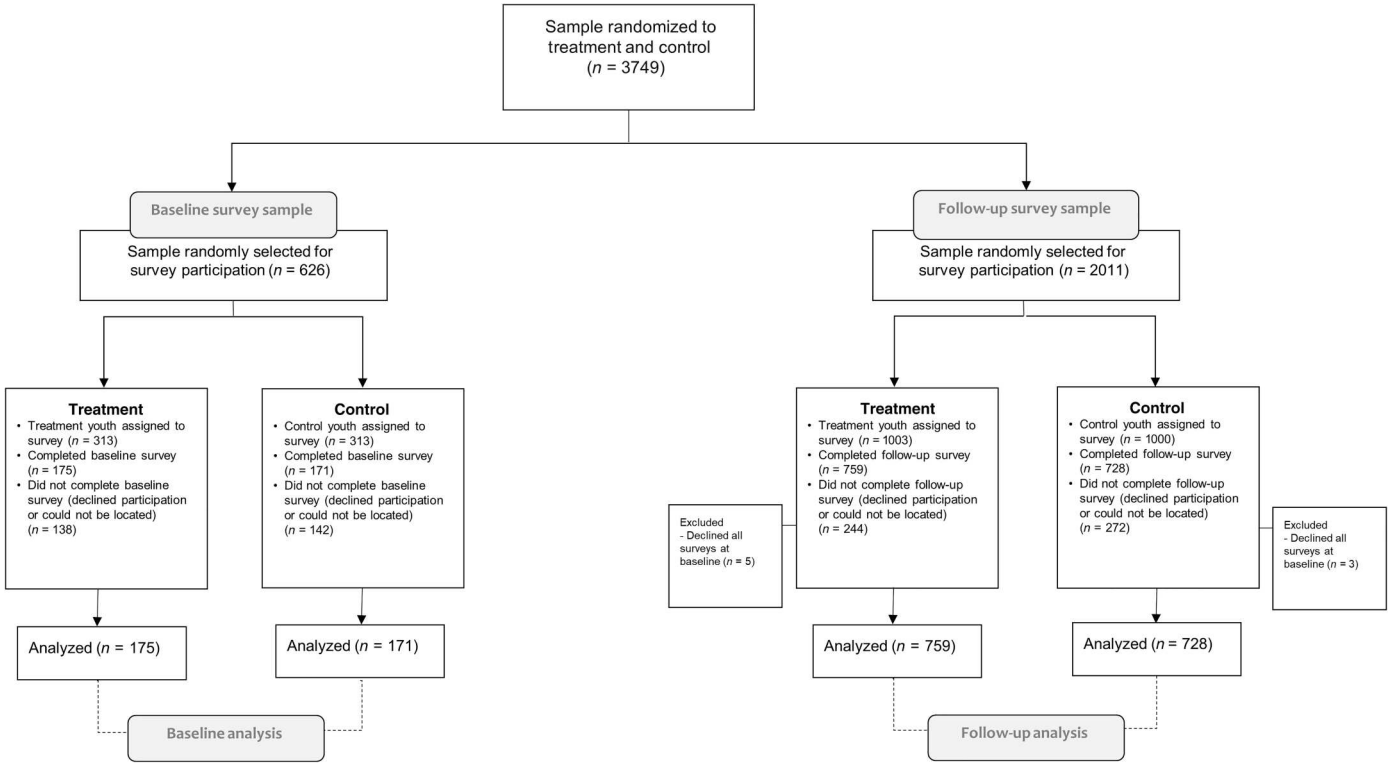

Fig. 2. WOW study consort diagram: Cohort I.

Table 1. Balance table: Cohort I and follow-up survey sample. Significance between the mean differences for the treatment versus control groups was estimated using a linear regression with block-level fixed effects. All missing cases were imputed using block-level averages. A joint *F* test was performed on demographic/academic covariates grouped together to test their joint significance. *P* values estimated using heteroscedasticity robust SEs. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; +$P < 0.1$. GPA, grade point average; LEP, limited English proficiency. Obs., observations.

	Cohort I			Follow-up survey sample		
Variable	**Treat mean**	**Control mean**	***P* value**	**Treat mean**	**Control mean**	***P* value**
Old for grade	0.149	0.122	0.135	0.151	0.127	0.118
Age as of randomization	15.15	15.12	0.256	15.16	15.13	0.19
Learning disability	0.096	0.073	0.099+	0.091	0.081	0.393
Free/reduced lunch recipient	0.952	0.951	0.929	0.946	0.954	0.402
Black	0.393	0.328	0.705	0.402	0.407	0.871
Hispanic	0.582	0.64	0.859	0.574	0.57	0.944
Out-of-school suspensions	0.067	0.07	0.319	0.072	0.072	0.985
In-school suspensions	0.101	0.081	0.729	0.109	0.107	0.91
Unexcused absences	8.92	9.43	0.412	8.96	9.37	0.417
Excused absences	4.1	4.47	0.26	4.03	4.36	0.161
Days present at school	162.7	162.8	0.62	162.6	162.9	0.631
GPA	2.86	2.81	0.235	2.86	2.84	0.532
Disciplinary incidents	0.287	0.25	0.898	0.298	0.278	0.567
Violent disciplinary incidents	0.025	0.027	0.312	0.027	0.025	0.805
Any disciplinary incidents?	0.157	0.14	0.861	0.162	0.156	0.666
LEP indicator	0.11	0.099	0.063+	0.102	0.097	0.703
Primary language Spanish	0.497	0.538	0.397	0.491	0.474	0.403
Obs.		3749			2003	
Joint *F* tests		$F_{17, 3702} = 1.044$			$F_{17, 1956} = 0.732$	
		Prob > *F* = 0.406			Prob > *F* = 0.772	

other mediators and secondary outcomes), enabling us to compare our study sample to the national distribution of young women aged 15 to 18. We also elected to use the well-established Child and Adolescent Trauma Screen (CATS) to measure trauma exposure and PTSD symptomology. We relied on the BASC-3 to gather self-reported data on our hypothesized mechanisms of self-esteem and self-efficacy. The self-efficacy measurement was supplemented using the Self-Efficacy Questionnaire for Children drawn from Muris (*29*).

We drew upon a variety of validated instruments and other survey questions to complement these instruments with a range of additional survey items that explored physical and sexual health, risk-taking behavior, self-image, future orientation, relationships, and more. We used the Difficulties in Emotional Regulation Scale–Short Form (*30*) to measure emotional regulation and the Future Outlook Inventory (*31*–*33*) to capture future orientation. Our surveys also included items from the National Longitudinal Study of Adolescent to Adult Health (Add Health) (*34*), which provided information on reproductive health outcomes and substance use.

Baseline survey

We administered a preprogram survey in the fall of school year 2017–2018 to obtain mental and physical health outcome measures not well captured in administrative data. We used this baseline survey data to test for and establish balance between treatment and control groups, improve the precision of our estimates, and provide information on the overall mental health characteristics of our population of interest.

Because of funding constraints, we could not conduct a baseline survey of our entire sample to complement the administrative data to which we had access for the full study sample. We were able to draw a random subset of 626 students from our full sample, blocking on school and grade, for baseline survey. Of the 626, we successfully consented, assented, and surveyed 346 to achieve a response rate of 55.3%. As shown in Tables 1 and 2 below, we achieved good balance in our surveyed sample on demographic and other variables, including our primary outcomes and proposed mechanisms.

Follow-up survey

Our research team conducted a follow-up survey from May to October 2018 to capture first-year post-program impacts on key outcomes for our first study cohort, which received services from fall 2017 to spring 2019. We worked with the Survey Research Center at the University of Michigan's Institute for Social Research (ISR) to administer the follow-up survey to a random sample of approximately 2000 students using the survey instruments that we used for the baseline survey. We used a two-phase sampling to obtain responses from a representative subsample of hard-to-locate respondents, with a goal of an effective response rate of 85%, which ISR had achieved in our previous studies of CPS

Table 2. Balance table: Follow-up survey respondents. Significance between the mean differences for the treatment versus control groups was estimated using a linear regression with block-level fixed effects. All missing cases were imputed using block-level averages. A joint *F* test was performed on demographic, academic, and mental health covariates grouped together to test their joint significance. Obs(*w*) refers to the weighted sample size, i.e., the sum of the sample weights used in the statistical tests. *P* values estimated using heteroscedasticity robust SEs. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; +$P < 0.1$. PTSD, posttraumatic stress disorder.

Variable	Treat	Control	Est	*P* value	Obs	Obs(*w*)
Mental health indicators						
Anxiety raw score	15.893	15.969	−0.066	0.938	329	329
Anxiety per score	59.25	58.565	0.603	0.839	329	329
Anxiety at risk or clinically significant	0.208	0.267	−0.064	0.165	329	329
Depression raw score	8.754	8.553	0.337	0.677	326	326
Depression per score	61.395	60.535	1.363	0.678	326	326
Depression at risk or clinically significant	0.257	0.277	−0.016	0.745	326	326
PTSD raw score	12.641	13.38	−0.372	0.805	333	333
PTSD at risk or clinically significant	0.388	0.362	0.049	0.364	333	333
Demographic indicators						
Old for grade	0.15	0.133	0.011	0.541	1487	1621.4
Age as of randomization	15.179	15.112	0.008	0.687	1487	1621.4
Learning disability	0.093	0.066	0.024+	0.085	1487	1621.4
Free/reduced lunch recipient	0.945	0.945	−0.002	0.87	1487	1621.4
Black	0.435	0.422	0.001	0.944	1487	1621.4
Hispanic	0.55	0.561	0.001	0.941	1487	1621.4
Out-of-school suspensions	0.07	0.071	−0.004	0.84	1487	1621.4
In-school suspensions	0.119	0.116	−0.007	0.824	1487	1621.4
Unexcused absences	8.589	8.925	−0.422	0.425	1487	1621.4
Excused absences	3.935	4.108	−0.324	0.217	1487	1621.4
Days present at school	163.327	163.874	−0.146	0.867	1487	1621.4
GPA	2.883	2.883	0.02	0.587	1487	1621.4
Disciplinary incidents	0.309	0.272	0.025	0.614	1487	1621.4
Violent disciplinary incidents	0.023	0.028	−0.006	0.481	1487	1621.4
Any disciplinary incidents?	0.171	0.14	0.027	0.143	1487	1621.4
LEP indicator	0.106	0.093	0.012	0.463	1487	1621.4
Primary language Spanish	0.473	0.469	0.012	0.564	1487	1621.4
	Number of obs. =	326	[Obs(w) = 326]			
Joint *F* test I (mental health and demographic)	$F_{25, 271}$ =	1.03				
	Prob > *F* =	0.428				
	Number of obs. =	1487	[Obs(*w*) = 1621.4]			
Joint *F* test II (demographic only)	$F_{17, 1440}$ =	0.918				
	Prob > *F* =	0.553				

students (*24*). We use the follow-up survey items in two ways: (i) to assess program impact and (ii) to identify potential mediators of treatment effects.

Administrative data

Master data sharing agreements with CPS and the Chicago Police Department (CPD) provided the administrative data foundation for our statistical analysis of education and arrest outcomes. Because the study sample was drawn from CPS administrative data, we had access to longitudinal student-level records for every student in the treatment and the control groups, including student-level school records on attendance, course grades, test scores, graduation and college enrollment, and school disciplinary actions. Students who switch schools or drop out are tracked and analyzed by original group assignment. Our data suggest that 3% of CPS 9th and 10th graders switch schools each year; 9% drop out or leave Chicago. These data were available before the intervention, throughout the intervention period, as well as after intervention, even for students who changed schools as long as they remained connected to CPS and were particularly valuable to

track youth for in-person surveys, as students could have changed schools, dropped out, or changed addresses during the study period.

Our team also has a master data sharing agreement with the CPD to access arrest data (*35*). These include data on the identity of the offender, date and location of the crime event, and criminal charges (for juvenile and adult offenders). We also had access to CPD victimization data covering juvenile and adult arrests citywide. We matched our study sample to these arrest records using probabilistic matching on name and date of birth. This study was approved by The University of Chicago Institutional Review Board, protocol IRB17-0585.

Data analysis

Because of our randomized experimental design, our analysis plan is straightforward. We estimate both the intent-to-treat (ITT) effect and the treatment-on-the-treated (TOT) effect. The ITT estimate comes from estimating Eq. 1

$$Y_i = \pi_0 + \pi_1 Z_i + X_i \pi_2 + B_i + \varepsilon_i \quad (1)$$

where Y_i is an outcome for participant i measured after random assignment, Z_i is an indicator for having been offered WOW programming, B_i is a full set of randomization block fixed effects, ε_i is a random error term, and X_i is a set of baseline controls to improve precision. These include age, school, and grade fixed effects; an indicator for having a learning disability, an indicator for having limited English proficiency, an indicator for being old for grade, and indicators for black and Latinx; and the following academic measures measured in the 2016–2017 baseline school year: number of each type of grade received (A to D and F), days present, days of out-of-school suspensions, and days of in-school suspensions. We also include indicators for ever being arrested for a violent, drug, or property related charge. Missingness of baseline covariates is balanced across treatment and controls groups. Missing baseline covariates were imputed using randomization block means by treatment status. For each imputed baseline covariate, we included an indicator identifying those observations that were imputed.

To estimate the TOT effect, we use random assignment (Z_i) as an instrumental variable (IV) for participation (D_i), as in Eqs. 2 and 3 (*36*, *37*). The first-stage equation is

$$D_i = \gamma_0 + \gamma_1 Z_i + X_i \gamma_2 + B_i + \mu_i \quad (2)$$

where D_i is an indicator for having participated in WOW programming (defined as having participated in at least one WOW group session), the γs are parameters to be estimated, μ_i is a random error term, and all other variables are defined as above. The relationship of interest is

$$Y_i = \beta_0 + \beta_1 D_i + X_i \beta_2 + \beta_i + \vartheta_i \quad (3)$$

The identifying assumption here is that treatment assignment has no effect on the outcomes of those assigned to treatment who do not participate. The IV estimate for the parameter B_i (Eq. 3) is essentially a ratio of two ITT estimates—the ITT effect on the outcome of interest in the numerator and the ITT effect on participation in the denominator.

RESULTS

Baseline descriptive statistics

We used the baseline survey to test for and establish balance between treatment and control groups, improve the precision of our estimates, and describe the overall mental health characteristics of our population of interest. Of our 346 survey respondents at baseline, we observe no statistically significant differences in our 13 baseline measures.

At baseline, young women in the study had suffered, on average, at least two serious traumatic experiences in their lifetimes; nearly 30% had personally witnessed someone being attacked, stabbed, shot at, hurt badly, or killed. More than 45% had someone close die suddenly or violently. Twenty-four percent scored within the clinical range or the at-risk range for anxiety; 27% scored within the clinical range or the at-risk range for depression (*38*).

The baseline survey also helps us understand the baseline prevalence of depression, anxiety, and PTSD within this population (N = 346). Approximately 17% of girls surveyed at baseline appeared at risk for depression and anxiety. About 10% appeared within the clinically significant range for depression and about 7% for anxiety. Findings on trauma exposure do not account for experiences of direct physical abuse, sexual abuse, or sexual assault due to mandatory reporting requirements and may be a conservative estimate. Our baseline surveys indicate that 38% of 10th and 11th grade young women in our study sample exhibit signs of PTSD; the observed prevalence of probable PTSD among these young women is twice that of service members returning from Iraq and Afghanistan. Our depression and anxiety measures are drawn from the BASC-3 SRP-A. The clinically significant range is defined as 2 SDs above the mean when scores are normalized to the appropriate national norm. In our case, we use general female, aged 15 to 18. Almost 30% scored within the probable PTSD range on the CATS, and on average, girls reported 2.5 exposures to traumatic events.

Follow-up survey outcomes

We observed large and statistically significant treatment-control differences in PTSD, depression, and anxiety scores (see Figs. 3 to 5).

We find that participation in WOW causes a 22% decrease in PTSD symptom severity scores, which measures the frequency and intensity of PTSD symptoms, and a 38% decrease in scores that indicate "moderate trauma-related distress." We also find that WOW participation causes significant decreases in measures of anxiety (9.77%) and depression (14.1%) (see Table 3).

We analyze academic outcomes in the administrative data for the full sample of 3749 girls randomized to treatment and control conditions. We observed no clinically or statistically significant differences between treatment and control participants in overall attendance, GPA, or freshman on-track status. Treatment and control participants displayed virtually identical outcomes on these measures (see Table 4).

Economic policy analysis (cost-utility results)

Our main results suggest that WOW improved mental health outcomes for young women participating in the program relative to those who did not. Building on these findings, we used cost-utility analysis to examine the economic value of such mental health improvements relative to WOW's associated program costs. We focused on observed treatment effects in reducing

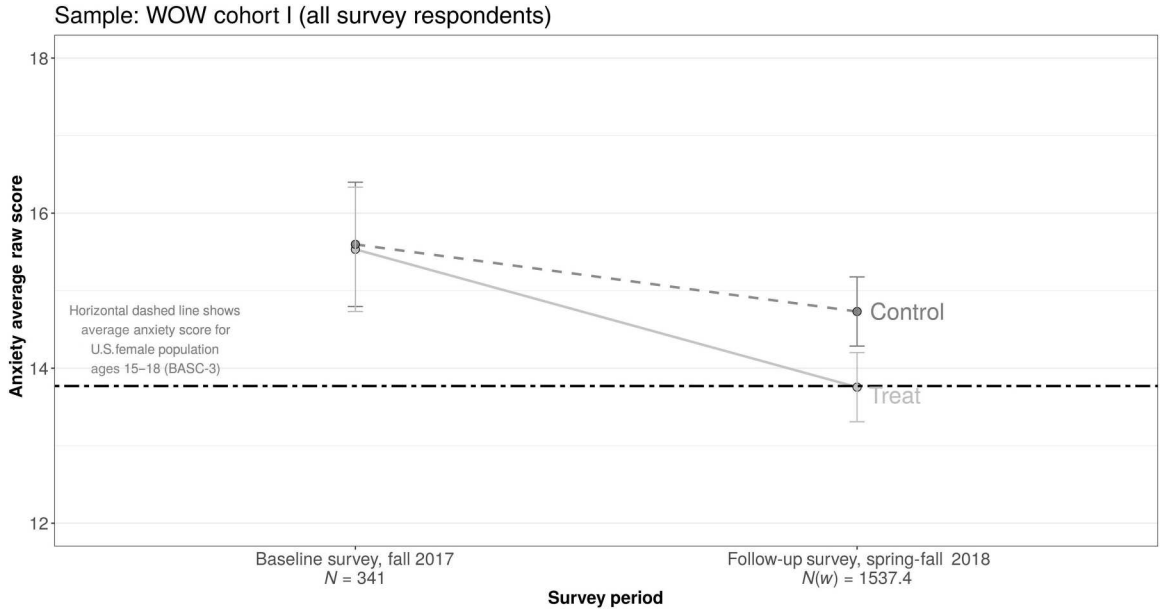

Fig. 3. Mental health trajectory for anxiety (effect of being offered WOW). Note that control means shown in the plot are regression adjusted for both the baseline and follow-up samples. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. BASC-3, Behavioral Assessment System for Children. *N*(*w*), weighted sample size.

depression and PTSD symptoms, converting these mental health benefits to changes in quality adjusted life year (QALY) units gained through the intervention. Briefly summarized, QALYs are constructed by multiplying the length of time that a program effect is presumed to occur by a utility weight that corresponds to the quality of life of individual. One year in perfect health is equivalent to one QALY as is 2 years in a very poor health state with the assigned utility weight of 0.5 (1 year * 1 = 1 QALY = 2 years * 0.5).

To generate an estimate of the cost per QALY gained through WOW, we applied utility weights for depression and PTSD, drawn from existing research, to the responses in the follow-up survey administered to the young women in the study. For depression, we assumed that an individual who is depression-free has a utility weight of 1 (perfect health) and that an individual who satisfies screening criteria for clinically significant depression has a utility weight of 0.59 (*39*). We also assumed that an individual at risk for clinically significant depression has a utility weight between 0.59 and 1 proportionate to their BASC-3 depression *t* score. The assignment of utility weights for PTSD followed a similar pattern. We assumed an individual who is PTSD-free has a utility weight of 1 and that an individual with clinically significant PTSD has a utility weight of 0.61 (*40*). Those at risk for PTSD receive a utility weight between 0.61 and 1 proportionate to their survey CATS PTSD score.

The program cost was estimated to be $2300 per actual participant in WOW. This number was obtained from Youth Guidance, the WOW program provider. That cost figure is based on a caseload of 55 youths served per counselor. It includes the direct salary and benefit costs of counselors and the curriculum specialists who provide training and coaching. It also includes further expensing of program management, other direct programming expenses, and overhead costs for information technology, human resources, and Youth Guidance financial and program evaluation/quality assurance required to operate the program.

We performed an ITT analysis to compare mental health outcomes among all youth invited to participate in WOW with those observed in the control group. On the basis of the program take-up rate of 65%, we converted our program cost data to be comparable to the ITT measure. The cost per program participant in the ITT treatment-assigned group was thus $2300*0.6525 or $1501.

Following the medical cost-effectiveness literature, we then computed the cost per QALY gained by computing an incremental cost-effectiveness ratio (ICER) for each outcome

$$\left(\frac{\text{Cost}_T - \text{Cost}_C}{\text{QALY}_T - \text{QALY}_C}\right) = \left(\frac{\$1501 - \$0}{\text{QALY}_T - \text{QALY}_C}\right)$$

where Cost_T refers to the average cost of implementing WOW calculated within the full treatment-assigned group (e.g., $2300*0.6525 = ~$1500 for the PTSD case). For the cost-benefit analysis, we use the average program cost across the entire treatment group; because not everyone in the treatment group enrolls in WOW, the per-person program cost of $2300 needs to be scaled down by the participation rate before we can plug it into the ICER's equation. The percentage of study participants from the treatment group enrolled in the WOW program is 65.3%. Note that this percentage varies slightly across the different mental health constructs, as not everyone responded to all questions in the survey (e.g., it is 64.92% for depression, 65.25% for PTSD, etc.). We take an estimate of β_1 from the following regression model as an estimate of QALYT − QALYC

$$\text{QALY}_i = \beta_0 + \beta_1 \text{WOW}_1 + BX + BXmiss + \mu_i + \varepsilon_i$$

where

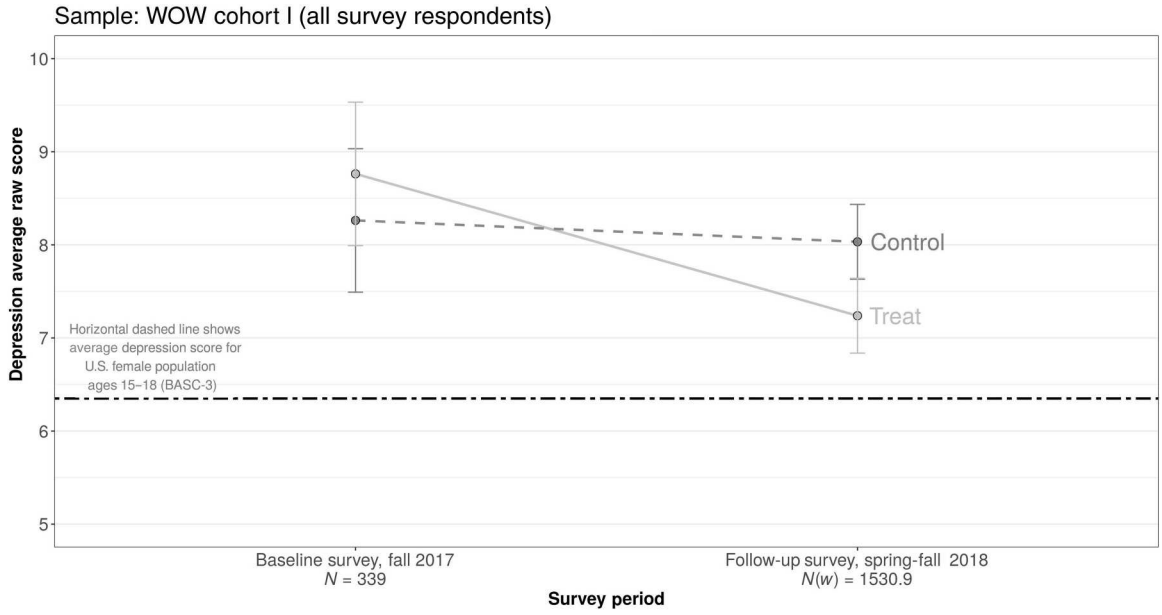

Fig. 4. Mental health trajectory for depression (effect of being offered WOW). Note that control means shown in the plot are regression adjusted for both the baseline and follow-up samples. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level.

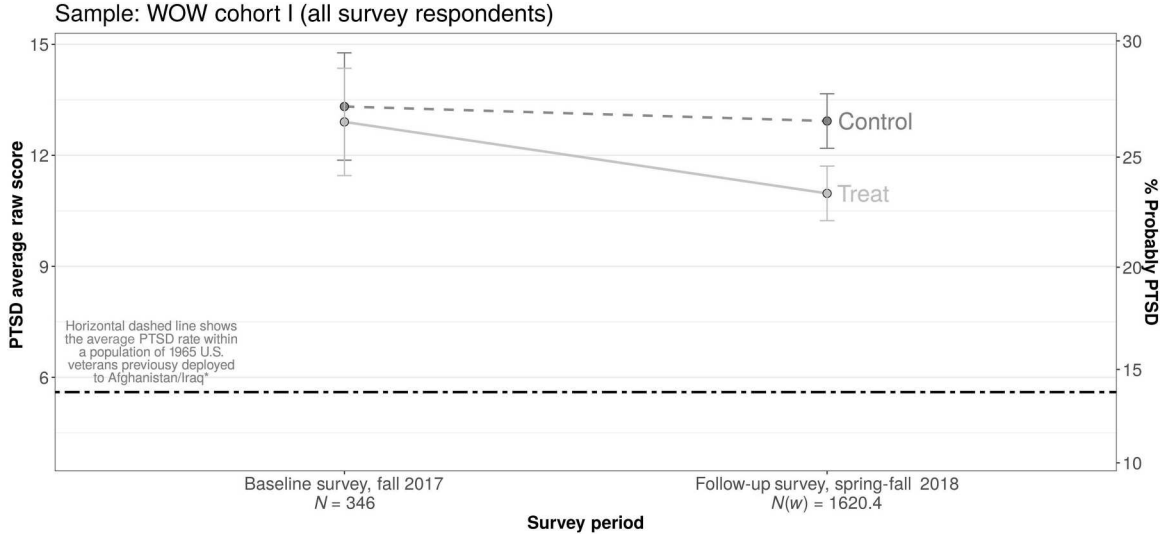

Fig. 5. Mental health trajectory for PTSD (effect of being offered WOW). Note that control means shown in the plot are regression adjusted for both the baseline and follow-up samples. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. The left-hand *y* axis indicates average posttraumatic stress disorder (PTSD) raw scores [Child and Adolescent Trauma Screen (CATS)]; the right hand one shows the percentage with probably PTSD. The horizontal line showing the probable PTSD incidence for the veterans' population should only be read using the right-hand *y* axis; the average PTSD raw scores (CATS) scale is only valid for the WOW sample (*38*).

1) $QALY_i$ is an individual's utility weight multiplied by an assumed effect duration of 365 days. (We perform sensitivity analyses of assumed effect duration below.)
2) WOW_i is a WOW treatment binary indicator.
3) X is a vector of baseline covariates.
4) *Xmiss* is a dummy indicator identifying those cases where the baseline covariates included in X were imputed, using randomization block-level means.
5) μ_i is block-level fixed effects.
6) ε_i is normally distributed random error term.

We compute an ICER for depression and PTSD separately, where the cost per QALY refers to the cost per QALY gained in terms of just that outcome.

We also compute the cost per QALY gained for the combination of PTSD and depression. To avoid complex assumptions about the

Table 3. Mental health primary outcomes: Anxiety, depression, and PTSD. Baseline covariates include: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A to D and F); dummies for ages 14 to 15, 15 to 16, and 17+; and indicators for having a learning disability, having LEP, being in 9th or 10th grades, being old for grade, being Black, being Hispanic, and ever being arrested in each of the three arrest categories: violence, drug, and property. *N*(*w*) refers to the weighted sample size, i.e. the sum of the sample weights used in the statistical tests. *P* values estimated using heteroscedasticity robust SEs. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; +$P < 0.1$. IV, instrumental variable.

Sample	*N*(*w*)	Control mean	Intention to treat	Effect of participation (IV)	Control compiler mean	*N*(*w*)	Control mean	Intention to treat	Effect of participation (IV)	Control compiler mean	*N*(*w*)	Control mean	Intention to treat	Effect of participation (IV)	Control compiler mean
			Anxiety raw score					**Depression raw score**					**PTSD raw score**		
Baseline	341	15.98	−0.064 (0.821)	−0.096 (1.178)	16.01	339	8.53	0.5 (0.789)	0.752 (1.132)	8.5	346	13.77	−0.416 (1.485)	−0.62 (2.116)	14
Year one FU	1537.4	14.83	-. 976* (0.455)	−1.513* (0.692)	15.48	1530.9	8.02	−0.794+ (0.41)	−1.227* (0.625)	8.7	1620.4	12.93	−1.956** (0.752)	−3.016** (1.14)	13.91
			Anxiety at risk					**Depression at risk**					**PTSD at risk**		
Baseline	341	0.18	−0.029 (0.041)	−0.044 (0.059)	0.21	339	0.17	−0.013 (0.042)	−0.019 (0.06)	0.16	346	0.09	0.053 (0.033)	0.079+ (0.047)	0.06
Year one FU	1537.4	0.15	0.006 (0.021)	0.01 (0.032)	0.16	1530.9	0.13	−0.009 (0.019)	- 0.013 (0.028)	0.13	1620.4	0.13	−0.042* (0.017)	−0.065* (0.026)	0.17
			Anxiety clinically significant					**Depression clinically significant**					**PTSD clinically significant**		
Baseline	341	0.08	−0.033 (0.026)	−0.049 (0.037)	0.11	339	0.1	0.001 (0.034)	0.001 (0.049)	0.13	346	0.29	−0.009 (0.049)	−0.013 (0.07)	0.3
Year one FU	1537.4	0.04	−0.003 (0.011)	−0.005 (0.017)	0.05	1530.9	0.09	−0.019 (0.017)	−0.029 (0.026)	0.12	1620.4	0.27	−0.035 (0.024)	−0.055 (0.036)	0.28
			Anxiety at risk or clinically significant					**Depression at risk or clinically significant**					**PTSD at risk or clinically significant**		
Baseline	341	0.27	−0.062 (0.045)	−0.094 (0.065)	0.32	339	0.28	−0.012 (0.049)	−0.018 (0.071)	0.29	346	0.37	0.044 (0.053)	0.066 (0.076)	0.36
Year one FU	1537.4	0.2	0.003 (0.023)	0.004 (0.035)	0.21	1530.9	0.22	−0.027 (0.023)	−0.042 (0.035)	0.25	1620.4	0.39	−0.078** (0.026)	−0.12** (0.04)	0.44

Table 4. Academic primary outcomes: Cohort I. Baseline covariates and randomization block-fixed effects included in all models. Baseline covariates include: days; present number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A to D and F); dummies for ages 14 to 15, 15 to 16, and 17+; and indicators for having a learning disability, having LEP, being in 9th or 10th grades, being old for grade, being Black, being Hispanic, and ever being arrested in each of the following arrest categories: violence, drug, and property. Year 1(2) outcomes for cohort I correspond to Chicago Public Schools (CPS) academic year 2017–2018 (2018–2019). Misconduct data from CPS was only available up to academic year 2018–2019 at the moment this table was constructed. PSAT/SAT test are not administered to senior year students; i.e., the sample used for estimates does not include participants that were enrolled in 12th grade during the testing year. The freshman on-track indicator is only applicable for cohort I CPS students enrolled in 9th grade at the beginning of academic year 2017–2018 (N = 1,199). *P* values estimated using heteroscedasticity robust SEs. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; +$P < 0.1$.

Sample	*N*	Control mean	Intention to treat	Effect of participation (IV)	Control compiler mean	Control mean	Intention to treat	Effect of participation (IV)	Control compiler mean	Control mean	Intention to treat	Effect of participation (IV)	Control compiler mean
			GPA				Failed any course				Credits earned		
Cohort I (year 1)	3749	2.61	0.003 (0.026)	0.005 (0.042)	2.67	0.29	0.01 (0.015)	0.017 (0.024)	0.23	6.65	0.007 (0.031)	0.011 (0.05)	6.78
Cohort I (year 2)	3749	2.62	−0.009 (0.025)	−0.015 (0.041)	2.69	0.28	0.003 (0.014)	0.006 (0.023)	0.24	6.64	0.025 (0.033)	0.042 (0.054)	6.72
			Days present				School persistence index				Any school incidence		
Cohort I (year 1)	3749	154.68	0.892 (0.809)	1.469 (1.32)	158.35	0	0.015 (0.02)	0.025 (0.032)	0.13	0.15	0.002 (0.012)	0.004 (0.02)	0.18
Cohort I (year 2)	3749	148.59	−0.504 (0.949)	−0.828 (1.55)	152.08	0	−0.018 (0.018)	−0.029 (0.029)	0.06	0.12	0.018 (0.012)	0.029 (0.019)	0.14
			Any L4-6 school incident				In-school suspension				Out-of-school suspensions		
Cohort I (year 1)	3749	0.08	0.004 (0.01)	0.006 (0.016)	0.09	0.08	0.008 (0.015)	0.012 (0.025)	0.1	0.11	0.004 (0.015)	0.007 (0.024)	0.11
Cohort I (year 2)	3749	0.06	0.004 (0.008)	0.007 (0.013)	0.06	0.06	0.033* (0.015)	0.054* (0.025)	0.05	0.06	0.005 (0.01)	0.008 (0.017)	0.06
			PSAT/SAT math score (SD)				PSAT/SAT math score (SD)				Freshman on-track		
Cohort I (year 1)	3749	0	−0.019 (0.029)	−0.031 (0.047)	0	0	−0.044 (0.029)	−0.072 (0.048)	0.02	0.86	0.025 (0.018)	0.043 (0.031)	0.88
Cohort I (year 2)	2595	−0.03	−0.053 (0.035)	−0.09 (0.06)	0.04	−0.02	−0.031 (0.034)	−0.053 (0.058)	0.03				

cumulative impact of co-occurring conditions, we assigned each student in the study a utility weight that corresponds to the minimum utility weight that they scored across depression and PTSD. Although anxiety is one of our primary outcomes, we do not include it in the main CBA analysis, because we were unable to find utility weights in the literature that mapped well onto the anxiety questions asked in the BASC-3. Table 5 reports the cost per QALY gained for each outcome.

The Institute for Clinical and Economic Research reports two standard thresholds to use for economic evaluations of pharmaceutical pricing as well as medical and public health interventions in the United States: $150,000 and $100,000 per QALY (*41*). Comparing the cost per QALY gained from WOW to these thresholds, WOW appears to be highly cost-effective when evaluated solely in terms of its impact on PTSD or on the combined PTSD/depression outcome. It is borderline cost-effective when evaluated solely as a depression prevention intervention.

To assess the sensitivity of these results to variance in the estimates of the program effects, we performed a bootstrap analysis to examine the variation in our computed ICER for the combined outcome using 100,000 repetitions. We find that WOW meets the $150,000 threshold in 92.4% of samples and the $100,000 threshold in 81.2% of samples. Figure S8 is the cumulative distribution function of the bootstrapping results.

As a further robustness check, we conducted a bounding exercise where we generate utility weights for a third mental health outcome, anxiety, using the utility weights for depression and students′ responses to anxiety questions of the BASC-3. This does not change our fundamental result. When we combine utility weights using the minimum quality-of-life weights for an individual across their PTSD, depression, and imputed anxiety scores and construct an ICER, we find that the cost per QALY gained to be $67,505. Assuming the quality of life for anxiety is no better than the quality of life for depression, this ICER provides an estimate of the cost per QALY gained across our three primary outcomes.

Table 5. Cost per QALY gained. Showing cost per QALY gained assuming a 365-day benefit. We also show the number of posited benefit days required to meet relevant cost-benefit thresholds. QALY, quality adjusted life year.

Outcome	Cost per QALY gained presuming a 365-day benefit	Days of posited benefit required to meet the $150,000 per QALY threshold	Days of posited benefit required to meet the $100,000 per QALY threshold
Depression as the only valued outcome	$180,522 per QALY	439	659
PTSD as the only valued outcome	$64,274 per QALY	156	235
Combined depression and PTSD	$58,025 per QALY	141	212

A second uncertainty (see Table 5) relates to the posited duration of program effects. If the true program effect were shorter than 365 days, then we may overstate program cost-utility. Examining the combined depression/PTSD outcome, we find that the intervention meets the $100,000 benchmark with a posited intervention duration of 235 days of benefit. The intervention meets the $150,000 benchmark with a posited benefit duration of 156 days.

Study limitations

Our analysis has several study limitations that must be considered in evaluating our results. Within-school randomization creates some potential for spillovers, which could have led us to understate program effects. Spillovers occur if control students directly enroll in WOW, if WOW staff provides informal mentoring or other supports to control-group students, or if treatment group services indirectly benefit control-group students in other ways.

Our mental health outcomes were based on the BASC-3 instrument rather than a diagnostic clinical interview. Our results thus should not be taken as providing clinical diagnosis for any specific individual. Because we could not conduct diagnostic clinical interviews, we adopted a conservative approach to the quality-of-life impact of co-occurring depression and PTSD symptoms, in which we assumed that the quality of life was equivalent to the minimum value obtained for each condition independent of comorbidity. To the extent that the quality-of-life impact of co-occurring symptoms is worse than the impact of either condition alone, we may understate the harms associated with co-occurring conditions.

Given study exigencies arising from the COVID-19 pandemic, we could not gather data on long-term mental health outcomes among WOW participants or controls. Our findings are thus based on respondents′ mental health status as captured in the follow-up survey. We were therefore not able to measure whether WOW induced benefits that lasted beyond the study period. We also could not observe intervention benefits that may have faded out for some participants before the follow-up survey was administered. WOW was also designed to be a 2-year program. We may understate the program′s full benefits given the timing of our mental health surveys.

Given these limitations, we conducted sensitivity analyses to examine the duration of program benefit required for WOW to pass the standard cost-effectiveness threshold. As noted above, WOW′s observed mental health benefits associated with reduced PTSD and depression symptoms would need to last approximately 157 days (roughly 5 months) for WOW to be deemed cost-effective, based on the $150,000-per-QALY threshold.

DISCUSSION

Young women, particularly those attending school in low-income, predominantly minority communities experience high rates of depression, PTSD, and other mental health challenges. Designing and fielding feasible and cost-effective interventions to address these challenges remains a key challenge.

The WOW intervention induces marked and statistically significant improvement in depression, anxiety, and PTSD symptoms among the young women who participate in this intervention. WOW was not designed to move academic outcomes such as

standardized test scores and grades. In addition, over the period observed, WOW did not improve these outcomes.

Evaluated on the basis of its ability to ameliorate mental health symptoms, WOW appears highly cost-effective when judged on the basis of standard cost-utility metrics used to evaluate medical and public health interventions. These benefits were achieved within the challenging real-world environment of 10 Chicago public high schools. At a per-participant cost of $2300, WOW provides one promising model that can be replicated at scale within resource-challenged public schools across the country.

The burden of mental health is often unseen and overlooked, particularly for young women of color, in part because it is not always associated with externalizing behaviors. This study marks one of the first studies of adolescent mental health for young women of color that documents prevalence of mental health challenges such as anxiety, depression, and PTSD in Chicago. It also provides rigorous evidence about how to systematically reduce the prevalence of these challenges at scale through an innovative, group-based, in-school model of therapy.

These findings are previously unknown in part because of the limited evidence base for existing interventions and programs intended to address the high rates of trauma exposure and consequent anxiety, depression, and PTSD found among young women in our study. Our results suggest that group-based, in-school therapy programs like WOW can be effective in reducing these rates and highlight the dearth of alternate services available to the young women in our study. More attention should be directed to WOW and other feasible, cost-effective interventions that support the mental health of young women.

This study was conducted in 10 schools within the real-world environment in America′s third largest city and in its third largest school district. The study included the vast majority of young women in these schools; approximately 95% of female students were randomized to either the treatment or control conditions.

Given that approximately 70% of adolescents′ mental health services are received in a school-based setting, this study marks an important advancement in quantifying the efficacy of the WOW approach of combining CBT, ACT, and narrative therapy into group-based counseling. Despite this advancement, there were young women who had more profound mental health challenges and more significant needs that were not appropriate for group-based intervention; future work should examine interventions to support the needs of these girls.

In our analysis of costs and benefits, we applied the best cost-utility metrics available. More analyses are needed to specifically study young women and to provide cost-effective and outcome measures tailored to this population. Young women experience high mental health disease burdens but are understudied in the literature.

It is also important to evaluate school-based, mental health interventions by the same yardsticks that we use to evaluate other interventions to extend life and improve health. If as a society we are willing to cover the cost of medication to support mental health through Medicaid and other payment sources, then we should also finance psychosocial interventions that prove to be cost-effective when viewed through the lens of these same cost-utility metrics.

Supplementary Materials

This PDF file includes:
Figs. S1 to S12
Tables S1 to S21
Legends for data files S1 to S3

Other Supplementary Material for this manuscript includes the following:
Data files S1 to S3

REFERENCES AND NOTES

1. R. T. Leeb, R. H. Bitsko, L. Radhakrishnan, P. Martinez, R. Njai, K. M. Holland, Mental health–related emergency department visits among children aged <18 years during the COVID-19 pandemic—United States, January 1–October 17, 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 1675–1680 (2020).
2. J. A. Fairbank, F. W. Putnam, W. W. Harris, The prevalence and impact of child traumatic stress, in *Handbook of PTSD: Science and Practice*, M. J. Friedman, T. M. Keane, P. A. Resick, Eds. (Guilford Press, ed 1, 2007). [chapter 13].
3. K. M. Fitzpatrick, J. P. Boldizar, The prevalence and consequences of exposure to violence among African-American youth. *J. Am. Acad. Child Adolesc. Psychiatry* **32**, 424–430 (1993).
4. C. J. L Murray, A. D. Lopez, The global burden of disease, in *Global Burden of Disease and Injury* (Harvard Univ. Press, Cambridge, ed. 1, 1996), pp. 117–200.
5. J. Angst, A. Dobler-Mikola, The Zurich Study–A prospective epidemiological study of depressive, neurotic and psychosomatic syndromes. IV. Recurrent and nonrecurrent brief depression. *Eur. Arch. Psychiatry Neurol. Sci.* **234**, 408–416 (1985).
6. S. E. Bruce, K. A. Yonkers, M. W. Otto, J. L. Eisen, R. B. Weisberg, M. Pagano, M. T. Shea, M. B. Keller, Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: A 12-year prospective study. *Am. J. Psychiatry* **162**, 1179–1187 (2005).
7. D. A. Regier, M. E. Farmer, D. S. Rae, B. Z. Locke, S. J. Keith, L. L. Judd, F. K. Goodwin, Comorbidity of mental disorders with alcohol and other drug abuse: Results from the epidemiologic catchment area (ECA) study. *JAMA* **264**, 2511–2518 (1990).
8. M. V. Porche, L. R. Fortuna, J. Lin, M. Alegria, Childhood trauma and psychiatric disorders as correlates of school dropout in a national sample of young adults. *Child Dev.* **82**, 982–998 (2011).
9. P. Sharkey, The acute effect of local homicides on children's cognitive performance. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11733–11738 (2010).
10. D. F. Tolin, E. B. Foa, Sex differences in trauma and posttraumatic stress disorder: A quantitative review of 25 years of research. *Psychol. Bull.* **132**, 959–992 (2006).
11. S. C. Rambaud, J. L. Pascual, R. Moro-Visconti, E. M. Santandreu, Should gender be a determinant factor for granting crowdfunded microloans? *Humanit. Soc. Sci. Commun.* **9**, 1–13 (2022).
12. R. E. Ford-Paz, L. Crown, K. Lawton, H. Goldenthal, G. Day, C. A. Coyne, T. Gill, N. Harris, S. Blakemore, C. Cicchetti, Working on Womanhood (WOW): A participatory formative evaluation of a community-developed intervention. *Eval. Program Plann.* **72**, 237–249 (2019).
13. D. A. Brent, D. Holder, D. Kolko, B. Birmaher, M. Baugher, C. Roth, S. Iyengar, B. A. Johnson, A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. *Arch. Gen. Psychiatry* **54**, 877–885 (1997).
14. G. N. Clarke, P. Rohde, P. M. Lewinsohn, H. Hops, J. R. Seeley, Cognitive-behavioral treatment of adolescent depression: Efficacy of acute group treatment and booster sessions. *J. Am. Acad. Child Adolesc. Psychiatry* **38**, 272–279 (1999).
15. P. M. Lewinsohn, G. N. Clarke, H. Hops, J. Andrews, Cognitive-behavioral treatment for depressed adolescents. *Behav. Ther.* **21**, 385–401 (1990).
16. P. Rohde, G. N. Clarke, D. E. Mace, J. S. Jorgensen, J. R. Seeley, An efficacy/effectiveness study of cognitive-behavioral treatment for adolescents with comorbid major depression and conduct disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **43**, 660–668 (2004).
17. A. Wood, R. Harrington, A. Moore, Controlled trial of a brief cognitive–behavioural intervention in adolescent patients with depressive disorders. *J. Child Psychol. Psychiatry* **37**, 737–746 (1996).
18. P. Smith, W. Yule, S. Perrin, T. Tranah, T. Dalgleish, D. M. Clark, Cognitive-behavioral therapy for PTSD in children and adolescents: A preliminary randomized controlled trial. *J. Am. Acad. Child Adolesc. Psychiatry* **46**(8), 1051–1061 (2007).
19. P. M. Barrett, A. L. Duffy, M. R. Dadds, R. M. Rapee, Cognitive-behavioral treatment of anxiety disorders in children: Long-term (6-year) follow-up. *J. Consult. Clin. Psychol.* **69**, 135–141 (2001).

20. T. In-Albon, S. Schneider, Psychotherapy of childhood anxiety disorders: A meta-analysis. *Psychother. Psychosom.* **76**, 15–24 (2006).
21. P. C. Kendall, E. Flannery-Schroeder, S. Panichelli-Mindel, M. Southam-Gerow, A. Henin, M. Warman, Therapy for youths with anxiety disorders: A second randomized clinical trial. *J. Consult. Clin. Psychol.* **65**, 366–380 (1997).
22. B. D. Stein, L. H. Jaycox, S. H. Kataoka, M. Wong, W. Tu, M. N. Elliott, A. Fink, A mental health intervention for schoolchildren exposed to violence: A randomized controlled trial. *JAMA* **290**, 603–611 (2003).
23. S. Kataoka, L. Jaycox, M. Wong, E. Nadeem, A. Langley, L. Tang, B. Stein, Effects on school outcomes in low-income minority youth: Preliminary findings from a community-partnered study of a school trauma intervention. *Ethn. Dis.* **21**, 71–79 (2011).
24. S. B. Heller, A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, H. A. Pollack, Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *Q. J. Econ.* **132**, 1–54 (2017).
25. J. White, Wellbeing and education: Issues of culture and authority. *J. Phil. Edu.* **41**, 17–28 (2007).
26. M. White, D. Epston, in *Narrative Means to Therapeutic Ends* (Norton & Co, New York, London, 1990).
27. R. T. Lopes, M. M. Gonçalves, P. P. P. Machado, D. Sinai, T. Bento, J. Salgado, Narrative therapy vs. cognitive-behavioral therapy for moderate depression: Empirical evidence from a controlled clinical trial. *Psychother. Res.* **24**, 662–674 (2014).
28. C. R. Reynolds, R. W. Kamphaus, K. J. Vannest, BASC3: Behavior Assessment System for Children. *PsychCorp* (2015).
29. P. Muris, A brief questionnaire for measuring self-efficacy in youths. *J. Psychopathol. Behav. Assess.* **23**, 145–149 (2001).
30. E. A. Kaufman, M. Xia, G. Fosco, M. Yaptangco, C. R. Skidmore, S. E. Crowell, The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and replication in adolescent and adult samples. *J. Psychopathol, Behav. Assess.* **38**, 443–455 (2016).
31. M. F. Scheier, C. S. Carver, Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychol.* **4**(3), 219–247 (1985).
32. A. Strathman, F. Gleicher, D. S. Boninger, C. S. Edwards, The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *J. Pers. Soc. Psychol.* **66**, 742–752 (1994).
33. P. G. Zimbardo, Strategies for Coping with Social Traps: Times Perspectives Influences, paper presented at the 98th Annual Convention of American Psychological Association, Boston, August 1990.
34. K. M. Harris, J. R. Udry, National Longitudinal Study of Adolescent to Adult Health (Add Health). *Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor]*. (1994–2018).
35. P. J. Cook, S. T. Parker, H. A. Pollack, Sources of guns to dangerous people: What we learn by asking them. *Prev. Med.* **79**, 28–36 (2015).
36. J. D. Angrist, G. W. Imbens, D. B. Rubin, Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* **91**, 444–455 (1996).
37. H. S. Brown, Accounting for no-shows in experimental evaluation designs. *Eval. Rev.* **8**, 225–246 (1984).
38. T. Schell, G. Marshall, Survey of Individuals Previously Deployed for OEF/OIF, in *Invisible Wounds of War: Summary and Recommendations for Addressing Psychological and Cognitive Injuries* (RAND Corporation, ed. 1, 2008), pp. 87–116.
39. J. R. Lave, R. G. Frank, H. C. Shulberg, M. S. Kamlet, Cost-effectiveness of treatments for major depression in primary care practice. *Arch. Gen. Psychiatry* **55**, 645–651 (1998).
40. E. Gospodarevskaya, L. Segal, Cost-utility analysis of different treatments for post-traumatic stress disorder in sexually abused children. *Child Adolesc. Psychiatry Ment. Health* **6**, 15 (2012).
41. Institute for Clinical and Economic Review, Overview of the ICER value assessment framework and update for 2017-2019 (2017).

Acknowledgments: We thank our partners at Youth Guidance including L. Crown, G. Day, N. Harris, R. Clarkin, R. Langan, R. Crosby, and M. Adler-Morrison and the WOW program team. We thank J. Broom, S. Dickson, and J. Sell for their help in accessing the data that we analyzed here. We also thank K. Grant for assistance with the selection of study instruments and A. Tentner for review of ACT and narrative therapy treatment modalities. **Funding:** This paper was made possible by the support of the AbbVie Foundation; Arnold Ventures; Paul M. Angell Family Foundation; Edna McConnell Clark Foundation; Logan; MacArthur; Polk; and grant number 2016-JU-FX-0002 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. Points of view or opinions in this document are those of the authors and do not necessarily represent the views of the U.S. Department of Justice, National Institutes of Justice or any other funder. Any errors are of course our own. **Author contributions:** M.P.B.: Conceptualization, methodology, investigation, validation, resources, supervision, project administration, and writing (original draft and review and editing). J.G.: Conceptualization, methodology, investigation, validation, formal analysis, and writing (review and editing). H.A.P.: Conceptualization, methodology, investigation, funding acquisition, and writing (review and editing). J.C.C.: Project administration and writing (original draft). M.C.: Supervision, project administration, and resources. L.D.-S.: Formal analysis. P.L.: Supervision, resources, formal analysis, and project administration. M.L.: Supervision, project administration, formal analysis, and resources. C.P.P.: Data curation, investigation, formal analysis, writing (original draft and review and editing), software, visualization, and validation. B.S.: Investigation, formal analysis, software, and validation. M.S.: Supervision, project administration, resources, and formal analysis. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The data used for this study are owned and controlled by the Chicago Public Schools (CPS); the Education Lab was granted limited restricted use of these data through a data use agreement between the University of Chicago and CPS that prohibits further sharing of the data. The data are defined in contract as Confidential Information that remains the exclusive property of CPS, and the University is not permitted to share any Confidential Information with a third party such as *Science Advances* without the previous contractual consent of CPS, even if the data are deidentified. Furthermore, the data are subject to federal privacy protections outlined in the Family Educational Rights and Privacy Act (FERPA) that prohibits disseminating any data below a minimum standard of aggregation. The University retains the right to publish research results, however, including aggregated statistics from our research findings. Although the University is unable to directly share data from this research with Science, it is possible for researchers to directly access these same data from CPS through a standard process. Researchers may submit an external data request form on the CPS website, which can be found at the following link: www.cps.edu/about/district-data/conduct-primary-research. Our Master Data Use Agreement with the Chicago Police Department similarly prohibits our resharing of their confidential information, or the dissemination of any information without approval in accordance with the Illinois Juvenile Court Act. Our contract with CPD states that: "Data provision will include juvenile arrest and victimization records that, per the Illinois Juvenile Court Act (705 ILCS 405/5- 905-1-f), requires explicit approval from 'the judge of juvenile court and the chief executive of the agency that prepared the particular recording.' Judge M. P. Toomin of the Circuit Court of Cook County and Chicago Police Superintendent D. O. Brown both provided written support for the Requestor's receipt of juvenile arrest and victimization records. The Requestor acknowledges the confidential nature of the criminal justice information supplied and agrees that disclosure to anyone not directly identified in the approved application in any manner that allows the identification of the individuals whose records have been obtained from the CPD is totally prohibited under any circumstances. Should the Requestor engage in any project involving use of the records by a third party not mentioned in this document, the Requestor will notify the Chicago Police Department. Upon review and approval by CPD, third party use will be documented in a separate agreement signed by representatives of the third party. To conceal the identities of persons whose records are supplied to the Requestor, the Requestor agrees to refrain from copying any data not absolutely necessary for the use identified in the application or its security and retain such data only so long as is necessary to conduct the program described in the application to CPD." The University of Chicago Education Lab is required to permanently destroy identifying information in both administrative records and in any survey or health data collected during the study period upon conclusion of the study. Deidentified analytic datasets will be securely stored for a period of up to 10 years following project completion, unless data partners request destruction at an earlier date. After 10 years following study conclusion, all administrative data obtained for the study (that our study team does not have ownership of) will be permanently deleted from our servers. However, deidentified survey data that were collected by our research team during the study period will be published together with this paper.

Submitted 24 March 2022
Accepted 3 May 2023
Published 7 June 2023
10.1126/sciadv.abq2077

Science Advances

Randomized evaluation of a school-based, trauma-informed group intervention for young women in Chicago

Monica P. Bhatt, Jonathan Guryan, Harold A. Pollack, Juan C. Castrejon, Molly Clark, Lucia Delgado-Sanchez, Phoebe Lin, Max Lubell, Cristobal Pinto Poehls, Ben Shaver, and Makenzi Sumners

Sci. Adv. **9** (23), eabq2077. DOI: 10.1126/sciadv.abq2077