Official title: Evaluating a School-based Social and Material Needs Identification System to Prevent Youth Violence Involvement
NCT number: 05986188
Date updated: November 21, 2023

Contents

1. Study Protocol 3

a. Data Sources 3

Pathways to Potential program (P2P) data 3

Michigan education data 4

Michigan Child Protective Services data system 4

Success coach survey 4

b. Linking Data Sources 5

2. Statistical Analysis Plan 6

a. Measures 6

Primary exposure. 6

Primary outcomes. 6

Secondary outcomes. 7

Moderating variables (Aim 3). 7

Descriptive variables 7

Negative control outcomes 8

b. Aim 1 9

Multiple-baseline and -location interrupted time series (ITS) 9

Generalized synthetic control method 12

Auxiliary analyses: G-computation with a super learner machine 12

c. Aim 2 13

d. Aim 3 13

3. References 14

1. Study Protocol

Social and material needs identification systems have existed in schools for decades, but these systems' effects on youth violence is unclear. A Michigan Department of Health and Human Services (MDHHS) program, Pathways to Potential (P2P), began in 2012 as a novel approach to human services delivery. P2P seeks to improve school communities' social conditions by identifying and reducing the level and concentration of risk factors for chronic absenteeism. As of 2021, 274 Michigan public schools were implementing P2P. P2P stations MDHHS caseworkers, called success coaches, in local public schools (grades K-12). Coaches work closely with school staff to monitor school attendance and address needs that are barriers to attendance. After identifying needs, coaches connect students and families to resources and public assistance and work with the school to address school-wide needs. School-wide P2P implementation seeks to alleviate social and material needs among the entire school community.

Given that P2P addresses students' and their families' social and material needs and reduces risk factors for multiple types of youth violence, the investigators expect P2P to reduce the proximal outcome of chronic absenteeism and the distal outcomes of peer aggression and child maltreatment. Our study will link longitudinal P2P program data to state administrative records and school disciplinary data, and the investigators will combine this with a survey of success coaches that assesses P2P implementation, to:

Aim 1: Evaluate the impact of school P2P participation on chronic absenteeism, peer aggression expulsions, and child maltreatment

Aim 2: Assess if chronic absenteeism mediates the relationships of school P2P participation with peer aggression expulsions and child maltreatment

Aim 3: Examine school and implementation factors that moderate the relationships between P2P participation and chronic absenteeism, peer aggression expulsions, and child maltreatment

a. Data Sources

Pathways to Potential program (P2P) data. P2P program data is a longitudinal data set that includes school building and district codes and names for schools that participate in P2P (codes are standardized and able to link to Michigan education data), years the schools were/are active in P2P, number of success coaches per school per year, and

success coach contact logs, which contains information on contacts P2P success coaches make with and on behalf of students and families to address needs. The team will access these data through a data use agreement (DUA) with MDHHS.

Michigan education data. The State of Michigan collects discipline data for all reported expulsion incidents in Michigan schools from 2001 to present, housed within the MI School Data website, the State of Michigan's official public portal for education data. Through the MI School Data website, the study team will access publicly available information for public school disciplinary data, including school year, school and district code, incident ID, incident date, and type of incident, type and length of disciplinary action, and any administrative follow up to address the incident.

In addition to this publicly available data, the team will access school-level and student-level education variables via an agreement with the Michigan Education Research Institute (MERI). School-level variables include school district, school locale (from large city to rural), if the school is a family resource center, and school grades served. Student-level variables include student grade, gender, race, if the student is chronically absent, if the student is economically disadvantaged, if the student was retained in grade, if the student graduated on time, if the student is in special education programming, if the student is an English learner, if the student attends the school as a school of choice, and student assessment data. The team will aggregate student-level data to the school level; as such, all variables will describe the school and/or student body and be at the school-level (rather than the individual level).

Michigan Child Protective Services data system. Through a DUA with MDHHS, the team will access administrative data for the state Child Protective Services system, MiSACWIS. MiSACWIS contains hundreds of variables, including dates of investigations into child maltreatment, and names of alleged victims. The investigators will link these data to the education data provided by MERI to generate the total number of child maltreatment investigations per school per year.

Success coach survey. The study team will survey P2P success coaches to assess variability in adherence to P2P key components. Representatives from the P2P Central Office will provide in-kind contributions of time and effort to collaborate on the development and distribution of a web-based survey to all P2P success coaches throughout the state to assess school-based implementation of P2P. The team will also have research assistants to serve as success coach outreach specialists. These outreach specialists will interface with the approximately 200 success coaches throughout the state and their supervisors stationed at local MDHHS offices to support the recruitment and completion of the developed success coach survey. The survey will

collect demographic information on success coaches in addition to assessing success coaches' adherence to three core components of P2P: student connection, family connection, and school connection. The investigator team will work with P2P Central Office staff and former success coaches to develop the survey, and the team will use information from the survey to develop overall P2P implementation scores that incorporate the three core components of student connection, family connect, and school connection. The investigators will recruit participants in collaboration with P2P Central Office in addition to local MDHHS offices. Specifically, the P2P Central Offices will support the recruitment of survey participants through disseminating the online survey via their success coach email listserv, distributing the survey during bi-monthly meetings with directors, and sharing the survey during the monthly virtual success coach call (which typically brings around 100 success coaches per month). Assuming 80% completion, the investigators anticipate 160 success coaches to complete the survey. The team will use Qualtrics ® to administer the online survey. The survey will display a research information sheet to participants as the first screen of the survey, which will detail the survey purpose and potential benefits and harms of participating.

Main analyses will exclude school years 2020-2021 and 2021-2022, as disruptions from the COVID-19 pandemic caused most K-12 schooling to transition online during this time. Sensitivity analyses will include these COVID-19 affected school years (2020-2021 and 2021-2022) to evaluate the impacts of P2P for all intervention years.

b. Linking Data Sources

The investigators will store all data in the secure MEDC enclave system. Sensitive data housed by MEDC reside solely on file, database or computational servers hosted within U-M data centers in Washtenaw County, MI. These servers are highly secure and approved for use with FERPA, Export Controlled (ITAR, EAR), PII, HIPAA and Sensitive Human Subject Research data. All servers are monitored 24/7 for network and physical intrusion and regularly patched by U-M Information Technology Services. Group-based access controls ensure data access follows the principle of least privilege. Separate server instances are used to ensure identifiable and de-identified data are not co-mingled. Virtual data enclaves requiring a VPN connection and two-factor authentication allow approved researchers to analyze data without removing it from protected data centers.

Our unit of analysis for all analyses will be the school, and our final linked administrative dataset will include multiple rows per school—one row per school per year from 2004-2024. The investigators will first link Michigan education data—including school disciplinary data—and P2P data using school and district codes. They will then

calculate the number of child maltreatment investigations per school per school year using the Child Protective Services data, and they will append this variable to the linked Michigan education data and P2P data using the school name identifier. Data for this project come from administrative data sources, plus newly collected survey data from P2P success coaches. The research team will conduct all administrative data linkages in the MEDC enclave system and store the final merged dataset in a password protected file in accordance with the data sharing agreements. All final data will be school-level aggregate data.

2. Statistical Analysis Plan

a. Measures

Primary exposure. The primary exposure at the school-level will be participation, and years of participation, in P2P. The team will obtain these data via a DUA with MDHHS.

Primary outcomes.

- *Chronic absenteeism (Aim 1 & 3 outcome; Aim 2 mediator)*. A student is considered chronically absent if they miss 10% or more of school days in which they are enrolled at the school. For each school, the investigators will calculate the total number of students that were chronically absent via the data provided by MERI.

- *Peer aggression expulsions (Aims 1-3 outcome):* The state of Michigan requires all public schools to report expulsions for all students to the state, including the incident type(s) which led to the expulsion. Expulsions are instances in which the local educational agency removes a student from their school for disciplinary purposes for the remainder of the school year or longer. The investigators will consider expulsions for peer aggression as expulsions which included any of the following: 1) firearm possession-handgun; 2) firearm possession-rifle/shotgun; 3) firearm possession-other firearm; 4) other weapon possession; 5) bomb or similar threat; 6) arson; 7) physical violence with injury, and/or; 8) physical violence without injury. For each school, the investigators will calculate the total number of peer aggression expulsions per school year.

- *Child maltreatment (Aims 1-3 outcome)*: The Michigan Statewide Automated Child Welfare Information System (MiSACWIS) includes dates of investigations into child maltreatment, names of alleged victims, and date of birth of alleged victims for all child maltreatment investigations in the state of Michigan. The

investigators will link the MiSACWIS data to individual student data from MERI that includes information on student name, date of birth, schools of enrollment, and dates of school enrollment enrolled schools. The investigators will use this linked dataset to calculate the number of child maltreatment investigations per school per school year.

Secondary outcomes.

- *Firearm-involved expulsions (Aim 1)*: In the state of Michigan, firearm possession on school grounds by a student requires the student to be expelled from the school. Expulsions for firearm possession on school grounds can be recorded as one or more of the following incident types: 1) firearm possession-handgun; 2) firearm possession-rifle/shotgun; 3) firearm possession-other firearm. The investigators will consider any expulsions that included one of these types as an expulsion reason as a firearm-involved expulsion. For each school, the team will calculate the total number of firearm-involved expulsions per school year.

Moderating variables (Aim 3).

- *P2P dose.* Ratio of success coaches to economically disadvantaged students within the school

- *Diversity index.* Diversity index of the school, calculated as the likelihood that any two randomly selected students within the same school identify as having different racial backgrounds

- *School level.* School type as elementary, middle, high school, or mixed

- *Pathways to Potential implementation score* (available for 2023-24). Success coach implementation score calculated from success coach survey

Descriptive variables.

The team will calculate the median and inter-quartile range (continuous variables) or percent (categorical variables) for the following variables, stratified by P2P status. Some descriptive variables will be offset or control variables in subsequent analytic models (see Aim 1, Aim 2, and Aim 3, below, for details).

- *Total enrollment.* Number of students enrolled in the school per school year

- *School locale.* School's geographic status on an urban continuum, from "large city" to "rural"

- *Grade distribution of student body.* Number or percent of student body in each grade (K-12) in the school per school year

- *Gender distribution of student body.* Number or percent of student body that identifies as male and as female in the school per school year

- *Race distribution of student body.* Number or percent of student body that is American Indian or Alaska Native, Asian, African American or Black, Native Hawaiian or Pacific Islander, White, Hispanic or Latino, or Two or More Races in the school per school year

- *Economically disadvantaged.* Number or percent of student body that is economically disadvantaged in the school per school year, i.e., eligible for free or reduced-price meals via locally gathered and approved family applications under the National School Lunch program, are in households receiving food (SNAP) or cash (TANF) assistance, are homeless, are migrant, or are in foster care

- *Grade retention.* Number or percent of student body that did not advance to the next grade level at the end of the year in the school per school year

- *Graduation.* Number or percent of the 4-year graduation cohort that graduates within 4 years in the school per school year (for high schools, only)

- *Special education.* Number or percent of student body that is eligible for Special Education services in the school per school year

- *English learners.* Number or percent of student body that is eligible to receive Limited English Proficient (LEP) services in the school per school year

- *School of choice.* Number or percent of student body that is school of choice services in the school per school year

- *Student assessment.* Average student assessment indicators per school per year

Negative control outcomes

The investigators will include two negative control outcomes that are hypothesized not to be affected by P2P implementation. These control series are:

- *Average annual salary of school staff*. Average full-time base salary for employees within the school per school year

- *Average educator effectiveness*. Average Educator Effectiveness ranking within the school per school year, a locally determined label assigned as a result of annual educator evaluations

b. Aim 1

Aim 1: Evaluate the impact of school P2P participation on chronic absenteeism, peer aggression expulsions, and child maltreatment

The Aim 1 analysis plan will use a two-pronged approach to estimate the effect of school Pathways to Potential (P2P) participation on the proximal outcome of (1) chronic absenteeism and the distal outcomes of (2) child maltreatment and (3) peer aggression expulsions. Our main analyses will use a controlled, multiple-baseline and -location interrupted time series (ITS), which we will supplement with a comparative case study design, extended for multiple treated units and staggered adoption, using generalized synthetic control methodology.

For any proposed models that are inestimable due to outcome distributions or trends, we will use the G-computation with a super learner machine for drawing inferences of the effects of school P2P participation on dichotomized outcome variables.

Multiple-baseline and -location interrupted time series (ITS)

The team based this approach on Degli Esposti and colleagues' analyses of "stand your ground laws" on firearm homicides and their corresponding analysis pre-registration.[1]

The team will use a controlled, multiple-baseline and -location interrupted time series (ITS) design, using natural variation in the timings and locations of P2P to estimate its effects. The team will model changes in (1) chronic absenteeism, (2) child maltreatment, and (3) peer aggression expsulsions using Poisson regression analyses within a generalized linear mixed model (GLMM) framework. Analyses will include all schools that implemented P2P at any point between 2011 and 2018 and schools that did not implement P2P during the study period, 2004-2022.

Time-varying confounders. ITS designs account for typical time-varying confounders that remain constant or change gradually over time, such as school size and grade, sex, race, and age distributions. In an ITS design, analysts should not control for these confounders if they do not (or are unlikely to) change from year to year (above and beyond long-term trends that are already identified from modelling the underlying trends). Analyses should control for factors that might change abruptly and unpredictably from year to year and coincide with P2P implementation (i.e., "history bias"), as excluding such factors from analyses may bias estimated effects and threaten internal validity. The team identified potential time-varying covariates that may confound the relationship between P2P implementation and our main outcomes. A potential confounder is the changing definition of student absence over time. Prior to school year 2017-18, a student was counted as absent only if they missed a full day of school; beginning with 2017-18, that rule was changed to count students as absent if they miss more than 50 percent of the school day to align with the Every Student Succeeds Act. Analyses will adjust for the changing definition of "student absence" over time.

Analyses. The investigators will use segmented Poisson regression within a GLMM framework to evaluate the impact of P2P participation on trends in the proximal outcome of (1) chronic absenteeism and the distal outcomes of (2) child maltreatment and (3) peer aggression expulsions. In line with recommendations from the American Statistical Association, the investigators will evaluate effects based on the magnitude of estimates and corresponding 95% confidence intervals rather than on the p-value.[2] The team theorizes that the impact of P2P participation will follow a step-change (i.e., abrupt sustained change in level) and therefore will use a simple dummy variable to code the implementation of P2P (pre-intervention period (coded 0) and the post-intervention period (coded 1)). This is because the team does not anticipate any lag or temporary change following P2P implementation. Schools that never implement P2P will be encoded as 0 throughout since no cross-over effect of the intervention is anticipated. The investigators will model time in school years (July 1st-June 30th) as an independent variable. In each GLMM model, analyses will adjust for existing temporal patterns by modelling year as a random effect and, if appropriate, non-linear trends will be modelled using polynomials or splines. If significant residual autocorrelation is detected ($p < 0.10$), analyses will use a sandwich estimator to generate robust standard errors, which provide more conservative estimates of uncertainty.

Analyses will estimate a random intercept, slope, and P2P effect model, allowing the outcomes, the effect of time, and the effect of implementing P2P to vary across schools.

The following regression equation represents a standard ITS analysis. *T* is the time elapsed since the start of the study with the unit representing the frequency with which

observations are taken (i.e., year); X_t is a dummy variable indicating the pre-intervention period (before P2P coded 0) or the post-intervention period (after P2P coded 1); Y_t is the outcome at time t; and N_t is the student body size (i.e., total enrollment) at time t entered as an "offset" variable into the model:

Eq. 1

Standard ITS: $logY_t = \beta_0 + \beta_1 T + \beta_2 X_t + logN_t$

The investigators will extend this standard ITS model to allow for random effects (γ) and include schools which never implement P2P as part of the control group. Here, t is replaced by i, and j is schools. This new notation represents the hierarchical structure of observations over time (i) nested within schools (j). X_{ij} is now a dummy variable indicating P2P participation within and between schools, where 1 encodes the presence of P2P (in participation period in P2P schools) and 0 encodes the absence of P2P (non-P2P schools and non-participation periods in P2P schools)

Eq. 2

Random intercept, slope, P2P effect: $logY_{ij} = (\beta_0 + \gamma_{0i}) + [(\beta_1 + \gamma_{1i})T_j] + [(\beta_2 + \gamma_{2i})X_{ij}] + logN_{ij}$

The team will fit all models using Monte Carlo maximum likelihood estimation and will check model fit via absolute model fit indices (RMSEA<0.10, SRMR<0.08) and model assumptions (e.g., multicollinearity, normally distributed residuals, covariance structure, homoscedasticity). The team will conduct all analyses in R, primarily using the glmm package.

The team will carry out pre-specified analyses to check the robustness of our methods and rule out ('falsify') alternative explanations for our findings. These analyses will check our study design, our classification of P2P schools and their time course, the influence of measured confounders, and our statistical models. As already described above, the team will extend our study design from a basic ITS design to a controlled multiple-baseline and -location ITS design to help address history bias; the main threat to our study design's internal validity.

In addition, the team will use the same analytical ITS approach where the intervention series (e.g., chronic absenteeism and peer aggression expulsions) and the control series (e.g., teacher salary and effectiveness) will be analyzed separately, and then compared. The addition of the control outcome will therefore be treated as a robustness check of our main findings.

Generalized synthetic control method

The team will conduct a complementary analysis using the generalized synthetic control (GSC) method to estimate the average treatment effect of P2P on the study outcomes.[3] GSC estimates changes in outcomes for each school that implements P2P relative to a synthetic control group with the same pre-intervention trends in these outcomes, estimated using tools from interactive fixed-effects models and synthetic control methodology (SCM).[4,5] Thus, each control group will be constructed using a weighted subset of non-P2P schools in Michigan, where weights were derived using data-driven algorithms to minimize differences in the outcome from each treatment unit throughout the pre-intervention period. We will additionally match on school-level covariates (see descriptive variables above, such as total enrollment); thus, adjusting for observable and unobservable confounders (time-varying and time invariant) under the assumption that outcome trends for schools in Michigan would be the same as those for the synthetic control group throughout the post-intervention period, in the absence of P2P. A full description of GSC and its implementation is published elsewhere.[3]

GSC will offer a complementary analysis to multiple -baseline and -location ITS, placing the emphasis on comparative non-P2P schools, rather than the pre-intervention period of P2P schools. To ensure that P2P schools fall within the range of the donor pool of non-P2P schools, the convex hull assumption, the study team may be required to remove schools that are outside the range (i.e., too high) of the outcome variables and place other theoretical restrictions on the donor pool (i.e., total enrollment within the schools) to overcome computational issues.[5]

Auxiliary analyses: G-computation with a super learner machine

For any proposed models that are inestimable due to outcome distributions or trends, we will use the G-computation with a super learner machine for drawing inferences of the effects of school P2P participation on dichotomized outcomes for each year from 2012/13 to 2021/22. Analyses will include one model for each school year and dichotomized outcome. For each of the models, the treatment group will be schools in their first year of P2P implementation; the control group will be schools who never or have not yet implemented P2P. As such, the size of the control and treatment groups will change across the modeled years.

The team will include covariates in the models that plausibly cause the focal outcomes including total enrollment; school type (elementary, middle, high school); school locale

(rural, town, city, suburban); student body distribution by race, gender, and grade; school academic performance; school economically disadvantaged student rate; grade retention rate; graduation rate; proportion of students enrolled in special education; proportion of English learners; and proportion of students attending as school of choice. Because P2P may influence these covariates, analyses will consider school values of these covariates for the prior school year, pre-P2P implementation (e.g., the 2012 model will consider 2011 values for these covariates). Outcomes and P2P status will be for the present modeled year (e.g., the 2012 model will consider 2012 P2P status and 2012 youth violence outcomes).

Analyses will estimate the marginal effect of P2P by G-computation for the binary outcome of whether the school had a peer aggression expulsion; analyses will obtain the outcome prediction (Q-model) using a Super Learner. The Super Learner includes the following machine learning techniques: logistic regression with Lasso penalization, logistic regression with Elasticnet penalization, neural network with one hidden layer, and support vector machine with radial basis. The team will use the gc.sl.binary function of the RISCA R package to estimate models. Analyses will estimate the average treatment effect on the treated (ATT).

c. Aim 2

Aim 2: Assess if chronic absenteeism mediates the relationships of school P2P participation with peer aggression expulsions and child maltreatment

The investigators will explore whether the effect of P2P implementation on peer aggression expulsions and child maltreatment is mediated by chronic absenteeism using longitudinal structural equation mediation modeling.[6] Analyses will assume contemporaneous mediating effects (i.e., mediator and outcome measured at the same point in time) given that measurement frequency is annual and mediating effects may dissipate in a year span. Analyses will control for variables listed in Descriptive variables, above.

d. Aim 3

Aim 3: Examine school and implementation factors that moderate the relationships between P2P participation and chronic absenteeism, peer aggression expulsions, and child maltreatment

Within the Aim 1 multiple baseline and location ITS models, the investigators will evaluate for heterogenous treatment effects of P2P on chronic absenteeism and child

maltreatment according to P2P dose received, school level, and student body diversity (see Aim 1, above). Analyses will add an interaction term between each of these three moderators and the P2P treatment indicator (X_{ij}, Eq. 2), evaluating each moderator in a separate model.

In addition to these analyses, the investigators will use data from success coach surveys to inform post-hoc analyses. Our population will be P2P schools (n~100) during the 2023-24 school year (the same year of the success coach survey) who had a success coach complete the survey. The team will estimate generalized linear models (with a negative binomial or Poisson distribution, as appropriate) to evaluate the associations between school adherence to P2P and peer aggression expulsions, child maltreatment, and chronic absenteeism within schools. Models will include an offset for the log of the student size to estimate rates. Analyses will control for school level covariates, as described in Descriptive variables, above.

3. References

1. Degli Esposti M, Wiebe DJ, Gasparrini A, Humphreys DK. Analysis of "Stand Your Ground" Self-defense Laws and Statewide Rates of Homicides and Firearm Homicides. *JAMA Netw Open*. 2022;5(2):e220077-e220077.

2. American Statistical Association. *American Statistical Association Statement on Statistical Significance and P-Values*.; 2016. https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf

3. Xu Y. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Polit Anal*. 2017;25(1):57-76.

4. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J Am Stat Assoc*. 2010;105(490):493-505.

5. Bonander C, Humphreys D, Degli Esposti M. Synthetic control methods for the evaluation of single-unit interventions in epidemiology: a tutorial. *Am J Epidemiol*. 2021;190(12):2700-2711.

6. Goldsmith KA, MacKinnon DP, Chalder T, White PD, Sharpe M, Pickles A. Tutorial: The practical application of longitudinal structural equation mediation models in clinical trials. *Psychol Methods*. 2018;23(2):191.