Findings from an Innovative Teen Pregnancy Prevention Program

Evaluation of Need to Know (N2K) in South Texas

Final Impact Report for

The University of Texas Health Science Center at San Antonio

August 15, 2015

Prepared by

Nicole A. Dierschke, MPH

Jon Gelfond, MD, PhD

Dierschke, N., Gelfond, J., Lowe, D., Schenken, R.S., and Plastino, K. (2015). "Evaluation of Need to Know (N2K) in South Texas: Findings From an Innovative Demonstration Program for Teen Pregnancy Prevention Program." San Antonio, Texas: The University of Texas Health Science Center at San Antonio.

Acknowledgements:

We would like to thank Dr. John (Jeff) Tanner for his work on this project, Angela Garcia and Tiffany Rouse for their assistance in carrying out the data collection, and Maria Peché for providing editing expertise.

This publication was prepared under Grant Number TP2AH000014 from the Office of Adolescent Health, U.S. Department of Health & Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the policies of HHS or the Office of Adolescent Health.

EVALUATION OF NEED TO KNOW (N2K) IN SOUTH TEXAS: FINDINGS FROM AN INNOVATIVE TEEN PREGNANCY PREVENTION PROGRAM

I. Introduction

A. Introduction and study overview

Despite continuous progress in lowering the teen birth rate in America, geographic, racial/ethnic, socioeconomic and healthcare disparities persist (Centers for Disease Control and Prevention, 2015). Significant geographic variations exist within the United States with the highest rates of teen births seen in the South, including the state of Texas (National Center for Health Statistics, 2012). The county in which Need to Know (N2K) occurred is a community where teen birth rates are consistently higher than national rates. In 2013, this county had a birth rate of 40 births per 1,000 females aged 15-19, which was 51% higher than the national rate (National Vital Statistics Reports, 2013).

In addition, sexually active students are at high risk for contracting and spreading sexually transmitted diseases (STDs). Prevalence estimates suggest that young people aged 15-24 years contract half of all new STIs (Satterwhite, et al 2008) and that 1 in 4 sexually active adolescent females have an STI, such as chlamydia or human papillomavirus (HPV) (Forhan, et al 2009). In 2014, teens in Texas ages 15-19 had higher rates of particular STIs than the state rate, including chlamydia (1844.7 per 100,000 vs. 475) and gonorrhea (412.7 vs. 127.7), while the rate for syphilis among 15-19 year olds approached the state rate (23.3 vs. 27.5) (Texas Department of State Health Services, 2014).

In Texas in 2013, 46% of high school students reported ever having sex and 33% reported being currently sexually active (Centers for Disease Control and Prevention, 2014). Texas ranks last in the usage of contraceptives among the 34 states that gather and report information on contraceptive usage in teens (The National Campaign to Prevent Teen and Unplanned Pregnancy, 2013). Only half of all high school students reported using a condom during their last sexual intercourse, less than 5%

reported using a moderately effective method (i.e., shot, patch, ring), and less than 2% reported using the most effective methods, a long acting reversible contraceptive, such as an IUD or implant.

In an effort to reduce teen pregnancy, sexually transmitted infections, and sexual risk behaviors, the U.S. Department of Health and Human Services (DHHS) has, for the last six years, conducted evidence reviews to identify new programs that have shown effectiveness in reducing these outcomes. Furthermore, the DHHS has been supporting teen pregnancy prevention programming, and in 2010, the University of Texas (UT) Teen Health, housed in the Department of Obstetrics and Gynecology at the University of Texas Health Science Center at San Antonio School of Medicine, received funding from the Office of Adolescent Health to develop and evaluate an innovative demonstration program designed to prevent teen pregnancy. The Need to Know (N2K) curriculum was developed as a 3-year intervention designed to promote adolescent health and wellness to foster healthy, responsible teens, and present information that is medically accurate and age-appropriate. To evaluate the newly developed N2K program, which had never before been evaluated, it was implemented in two high schools in South Texas that serve high-risk students. At study start in 2011, almost 64% of students in the two high schools were characterized as economically disadvantaged, 82% were minorities, and the graduation rate was 80.9% compared to 86% statewide (Texas Education Agency, 2012). The teen birth rate for the county in which the intervention occurred was 61.8 per 1,000 females aged 15-19 in 2011, compared to the rate for Texas of 46.9 and for the nation of 31.3 (San Antonio Metropolitan Health District, 2011; National Vital Statistics, 2011).

This report presents the methods and results of an evaluation of the N2K program. The evaluation components included: (1) N2K program impacts on high school students (impact study), and (2) implementation elements describing adherence, quality of interactions and

engagement, the experience of the comparison group, and the context in which the program was implemented (implementation study).

B. Primary research question(s)

This evaluation tested the impact of the 3-year N2K curriculum on sexual initiation and engagement in risky sexual behavior among students. Risky sexual behavior is defined as sexual intercourse in the last three months without using a condom or birth control, even once. The primary research questions were:

1. At program completion, what is the impact of Need to Know on sexual initiation (vaginal intercourse) among virgins at baseline?
2. At program completion, what is the impact of Need to Know on engagement in risky sexual behavior for the full analytical sample?

C. Secondary research questions

Four secondary research questions measured the impact of N2K on the same two behavioral outcomes at the end of the first year and at the end of the second year of programming:

1. After one year, what is the impact of Need to Know on sexual initiation (vaginal intercourse) among virgins at baseline?
2. After one year, what is the impact of Need to Know on engagement in risky sexual behavior for the full analytical sample?
3. After two years, what is the impact of Need to Know on sexual initiation (vaginal intercourse) among virgins at baseline?
4. After two years, what is the impact of Need to Know on engagement in risky sexual behavior for the full analytical sample?

II. Program and comparison programming

N2K is based on a framework of positive youth development intended to delay sexual initiation and reduce the rate of teenage pregnancy. The N2K curriculum is designed to support the students throughout their high school career with program implementation over three consecutive years from 9th through 11th grade.

A. Description of program as intended

The UT Teen Health N2K Program consists of 3 years of curriculum intended for 9th, 10th and 11th grade students to be delivered in a group or classroom setting of up to 32 students. Each year of the curriculum features 16 lessons. Eight lessons are intended to be taught in the fall semester and eight in the spring. A lesson is taught one time per week, and has between 3 and 5 topics that are intended to be covered. Each lesson is 25 minutes long, and can be taught in any class period that allows for 25 minutes of content instruction. The program was implemented by four health educators who have a bachelor's degree, at least two years of related experience, and who are trained to implement the program. The same four health educators implemented the program over the three years, and each health educator delivered the lesson to multiple class sections every week.

The N2K Curriculum was written by a team of professionals from the Department of Obstetrics and Gynecology at the University of Texas Health Science Center at San Antonio School of Medicine. The 9th-grade course, N2K: Basics, consists of lessons that encourage self-discovery and goal setting; discuss human growth and development, adolescent risk behaviors, communication, STDs, abstinence, contraceptives, teen dating violence, legal issues, responsible media use, and refusal skills, as demonstrated through role-playing. The 10th-grade course, N2K: Decisions, consists of lessons that increase decision making skills, clarify values, promote healthy relationships and the benefits of delaying sex, and review of anatomy, contraceptives, STDs, and legal issues. The 10th-grade course promotes the development of critical thinking

skills and application of the facts to real life situations. N2K: Decisions includes an original webisode series of 11 episodes, the first of which is viewed in class, and the others can be viewed online outside of class. The webisodes are a series of skits that show student actors (not students in N2K) implementing the skills and facts presented in class. The series runs in conjunction with the 10th-grade curriculum lessons 3-8. The 11th-grade course, N2K: Relationships, reinforces basic themes taught in prior years. It emphasizes healthy relationships, healthy living, and taking responsibility for personal health as an adolescent matures into adulthood. The final three lessons are a culminating activity, including creating collages, performing skits, writing songs, and journal entry, which allows students to express what they have learned and present it to others. The 3-year curriculum covers several content areas. About 35% of the planned 48 lessons focus on abstinence, 27% highlight positive youth development, 13% discuss risky behaviors, 12% address STDs, 8% cover contraceptives, and 5% discuss anatomy and puberty.

The first year of the program is delivered during a core subject in which all 9th-graders enroll, such as English. Teaching the N2K program during a core subject such as English helps to minimize the possibility that students outside of the 9th grade would be enrolled in that class. Health educators are assigned to teachers' classrooms where the core class is taught. The N2K program begins approximately four weeks into the semester and runs once per week. The eight lessons finish by the end of the calendar year before the holiday break. The second semester program begins two to four weeks after the holiday break, and students receive lessons weekly. All lessons finish before the end of the school year. The 10th- and 11th-grade N2K courses are implemented in the same way during another core course. In this way, the intervention is given for three consecutive years to treatment students.

A number of additional components reinforce the information provided in the classroom lesson. For example, Facebook is used to post four messages per lesson that are designed to enrich, reinforce, or answer a frequently asked question. One Facebook post stated, "Why do people call using a condom 'safer sex' and not 'safe sex'?" There is also one teen advisory board (TAB) per school which consists of 10-30 teens receiving the intervention who are nominated or volunteer and commit to serve on the board for at least one year. Health educators facilitate the TAB, which meets once a month and provides input on how to promote the N2K Curriculum. The members function as campus ambassadors. TAB members create awareness by wearing N2K t-shirts, making posters, and posting pictures of TAB activities, including such as team building exercises and volunteer work, on social media. Members can serve on the TAB for more than one year. Lastly, parental interaction is encouraged during the parent/student night, which is facilitated by the health educator and held once each year and occurs midway through the school year on each campus. During this 90-minute evening event, health educators present medical information regarding adolescent brain development, STDs, pregnancy prevention, and parent-child communication, and assist parents in discussing sensitive topics with their teens.

B. Description of counterfactual condition

The counterfactual condition involved no intervention; it represented business as usual. The N2K intervention supplemented what normally occurred in health education at two South Texas high schools. Health class was available to both treatment and comparison students but was not a requirement for graduation, and its content varied by teacher. Examples of topics covered in health class include, stress; nutrition and wellness; exercise; alcohol/tobacco/drugs.

In Texas, schools cannot talk about reproductive health without parental consent; so, health class should not cover sexual education. Students with parental consent for the survey who were in 9th grade the year prior to the treatment students served as comparison students; thus, the

counterfactual condition represented what was available to students before the intervention began. There was no additional school-based teen pregnancy prevention programming available to the treatment and comparison students. Teen pregnancy prevention activities that took place outside of the school in the community, such as at doctor's offices or faith-based organizations, were unknown.

III. Study design

A quasi-experimental design was used to estimate the impact of N2K on delaying sexual initiation and reducing engagement in risky sexual behavior. The quasi-experimental design involved a comparison group of students from the two schools who entered 9th grade in one school year, and a treatment group of similar students from the same schools who entered 9th grade the following year. This design is advantageous because the treatment group is one year behind the comparison group, making it less likely that the students in these two classes will be interacting and potentially exchanging information they learned in the N2K program.

The sections below describe the study sample recruitment, treatment assignment, data collection methods, assessment of baseline equivalence, methods for evaluating the primary and secondary research questions, and methods to assess implementation.

A. Sample recruitment

The study school district was chosen because it is a large school district with two high schools; it was believed to have a diverse student population; it was not involved in any other sex education interventions during the proposed period of intervention; and the district was willing to allow the intervention in the two high schools. The district's large size ensured a large enough sample size for sufficient study power; furthermore, the student diversity was advantageous for evaluating N2K to determine if the program could be transferred to other school districts nationwide which have diverse student populations.

Students were eligible for the study if they were in the 9th grade for the first time in the fall of the enrollment year (2011 for comparison students, and 2012 for treatment students). Students were not eligible to be a part of the evaluation sample if they were members of the LifeSkills self-contained special education group and not mainstreamed with other students.

The comparison group target population was students enrolled in 9th grade at the two high schools in the fall of 2011. Since students can enroll in school at various times, two specific time points were used to identify eligible students. In July 2011, a 9th-grade enrollment roster was obtained from the two high schools for mailing consent forms to families. In September 2011, an updated 9th-grade roster was obtained for the purpose of clarifying that the students were indeed enrolled in the 9th grade at the start of the school year. Additional students were recruited at registration events until the September surveying date. Incentives were given in appreciation for the timely return of consent forms. Students received $10 iTunes gift cards. Parents were entered in a drawing for one $100 or one of five $50 supermarket gift cards. All classroom teachers were given $25 Wal-Mart gift cards to thank them for collecting the consent forms that were returned during school hours. The same procedure was used for the following year (2012) to recruit treatment students.

The school teachers and principals were aware that N2K received approval from the school board and was occurring for three years. The principals knew that the comparison group was one year ahead of the treatment group. Parents and students were not aware of the condition for the student at the time of enrollment in the study.

While student participation in the N2K curriculum did not require active parental consent, parents were given the opportunity to opt their child out of participation in the curriculum. School administration mailed letters home to parents/guardians of students who would be offered

the N2K curriculum each fall before teaching began (beginning during the 2012-2013 school year). Parents were given the opportunity to view/inspect the N2K curriculum in the school or district office, or at a program viewing night offered each fall at each school to have any questions answered. All students whose parents did not opt them out were included in the programming. Students whose parents opted out were excused from the class during N2K programming and completed an alternative activity in the library.

Written parental consent was required for treatment and comparison students to participate in the study data collection activities. The number of comparison students eligible to participate in the study was 1,631 and the number of treatment students eligible to participate in the study was 1,491. There were 856 treatment and 760 comparison students with parental consent to participate in the evaluation.

B. Study design

This study used a quasi-experimental design. The comparison group consisted of students with parental consent who were enrolled in the 9th grade in September of 2011. The treatment group consisted of students who were enrolled in the 9th grade in September of 2012. Since the treatment and comparison groups started in different years (2012 and 2011), it is possible that there were differences in their experiences at the high schools which affected their reproductive health behaviors that were not due to the intervention. This is sometimes called a "history confound" in research and evaluation. We confirmed there were no changes in health education curricula, school leadership, school policy, or staff teaching health education. This decreases the chances that differences seen between the groups are due to something other than N2K.

The quasi-experimental design resulted in an imbalance between the treatment and the comparisons groups in terms of characteristics that could affect outcomes. The initial examination of equivalence on the full study sample showed statistically significant differences

between the treatment and comparison groups with respect to the percentage of students who engaged in risky sexual behavior at baseline. The percentage students reporting sexual activity without a condom in the last 3 months in the treatment group was 1.7% and 4.5% in the comparison group ($p = .01$).

Propensity score matching was used to create a balanced analytical sample from the treatment and comparison groups. The variables used in the calculation of the propensity score were the demographic variables (age, gender, race and ethnicity) plus any variables that were statistically significantly different between the groups at an alpha < .05 level. Only observations from students who completed surveys at baseline and the end of 9th, 10th, and 11th grades were considered in the matching algorithm. Procedures included calculating a propensity score for each treatment and comparison group subject. Students' propensity scores were matched using nearest neighbor matching, after removing outlying propensity scores. The final subset of comparison group members was created by selecting those with propensity scores similar to that of the treatment subjects. Propensity score matching was used to construct two analytical samples: (1) students with data at all four data points that reported they were virgins at baseline and (2) students with data at all four data points, irrespective of their sexual experience at baseline. See Appendix D for more details.

C. Data collection

Data to assess program impacts were collected via paper survey at four time points: baseline, end of 9th grade, end of 10th grade, and end of 11th grade. Data regarding implementation were collected via fidelity and observation logs, a survey to examine sexual education in school, and an examination of contextual factors, all collected throughout the study period.

1. Impact evaluation

Data were obtained using a paper survey instrument, which included demographic information, behavioral, knowledge, and attitude questions. The same survey was used at each time point. Students required an average of 15 minutes to complete the survey. The baseline survey was administered at the beginning of the 9th grade year (first week of September). The spring survey always occurred the third week of April, which for the treatment cohort was at the end of the spring component of the program. All data were collected using the same methods across treatment and comparison groups. Appendix A includes additional details on the data collection schedule (Table A.1) and data collection procedures (Table A.2).

2. Implementation evaluation

Instruments were created to assess fidelity of program implementation, including adherence, quality, experiences of the comparison group, and context. Methods, data sources, frequency of collection, and staff responsible for collection of implementation elements are found in Appendix B. Measures of adherence included the type and number of sessions offered, what sessions were received, and what content was delivered to students. Additionally, program staff collected information on topics covered as well as health educators' credentials and training to deliver the program. Using the observation log, measures of quality included staff-student interaction and student engagement with the program.

Experiences of the comparison group were collected using the paper survey administered to both the treatment and comparison groups. We also gathered data on comparison students who were enrolled in a health class that covered a topic related to pregnancy prevention during the implementation time frame. Furthermore, to document context, program staff conducted periodic interviews with school administration and attended monthly School Health Advisory Committee (SHAC) meetings to provide information on any teen pregnancy prevention programming

available to both treatment and comparison groups as well as any external events affecting implementation. Finally, health educators reported any unplanned changes in program staff, setting, or delivery.

D. Outcomes for impact analyses

The two behavioral outcomes for the primary impact analyses were measured using three survey questions. The first outcome was sexual initiation, which was measured with a single-item dichotomous measure from the survey question: *"Have you ever had sexual intercourse?"* This survey item serves as an outcome but was also used to identify students who were virgins at baseline, which is the analytical sample for research questions focused on sexual initiation. The second outcome was engagement in risky sexual behavior, which was measured using three additional single-item dichotomous measures from the survey questions: "*In the past 3 months, have you had sexual intercourse?"* "*In the past 3 months, have you had sexual intercourse without you or your partner using a condom, even once?"* and *"In the past 3 months, have you had sexual intercourse without using an effective method of birth control[1], even once?"* If students indicated they had never had sex or that they had not had sex in the past 3 months, they were coded as 0 as they had not engaged in risky sexual behavior. If students indicated they had sex, the other two items were used to determine their value on this outcome. If they had not had sex without a condom and had not had sex without an effective form of birth control they were coded 0 as they had not engaged in risky sexual behavior. If students answered "yes" to either of these two survey questions, they were considered to have engaged in the risky sexual behavior and were coded 1. A detailed description of primary outcomes and timing of the measurements

[1] Effective birth control methods listed in the survey included: condoms, birth control pills, the shot (Depo Provera), the patch, the ring (NuvaRing®), IUD (Mirena® or Paragard® or Skyla®), and an implant (Implanon® or Nexplanon®).

are provided in Tables III.1. Outcomes for the secondary research questions were constructed in the same manner but used data from (1) baseline (fall 2011/fall 2012) and the first follow-up (spring 2012/spring 2013); (2) baseline (fall 2011/fall 2012) and the second follow-up (spring 2013/spring 2014).

Table III.1. Behavioral outcomes used for primary impact analyses research questions

Outcome name	Description of outcome	Timing of measure relative to program
Sexual initiation of virgins at baseline	Have you ever had sexual intercourse?	Two times - baseline (fall 2011/fall 2012) and final survey (spring 2014/spring 2015)
Engaged in risky sexual behavior	Constructed based on response to four items: 1) Have you ever had sexual intercourse? 2) In the past 3 months, have you had sexual intercourse? 3) In the past 3 months, have you had sexual intercourse without you or your partner using a condom, even once? 4) In the past 3 months, have you had sexual intercourse without using an effective method of birth control* even once? Students are coded 0 (did not engage in risky sexual behavior) if: (1) they indicate they have never had sexual intercourse; or (2) they indicate they have not had sexual intercourse in the past 3 months; or (3) they indicate they have not had sexual intercourse in the past 3 months without a condom; and (4) they indicate they have not had sexual intercourse in the past 3 months without an effective form of birth control. Students are coded 1 (engaged in risky sexual behavior) if they responded they were sexually active and in the past 3 months had sex without a condom at least once OR in the past 3 months had sex without an effective form of birth control at least once.	Two times - baseline (fall 2011/fall 2012) and final survey (spring 2014/spring 2015)

E. Study sample

The analytical sample for the first primary research question was defined as virgins (those not sexually initiated at baseline) with complete data on the outcome variables at baseline, spring of year 1, spring of year 2, and spring of year 3. For the second primary research question, the analytical sample was defined as all students (including virgins and nonvirgins) with complete

data on the outcome variables at baseline, spring of year 1, spring of year 2, and spring of year 3. For the secondary research questions, the analytical samples were the same as the corresponding samples used in the primary research question analyses.

Table C1. in Appendix C describes the flow of the sample from the start of the study to each follow-up survey. As described in section III.B, the total sample included 1,616 students (760 comparison students and 856 treatment students). Baseline surveys were collected from 1,568 students (822 treatment and 746 comparison). Of those students contributing baseline surveys, 1,064 were virgins at baseline (579 intervention and 485 comparison) There were 901 (476 treatment and 425 comparison) students that completed all four surveys and thus were included in the propensity score matching process. After propensity score matching, the final virgin analytical sample was 628 (314 treatment and 314 comparison) and the full analytical sample was 784 total (392 treatment and 392 comparison).

The primary research questions were addressed using the virginal analytical sample of 628 students. The average age was just under 15 years old (14.7 for both the treatment and comparison groups). The sample included slightly more girls than boys (55.4% for the treatment and 54.5% for the comparison group). The majority of students identified as “other” for their race, which includes students who did not indicate they were White or Black, may have been more than one race, or did not indicate their race (57% intervention and 58.3% comparison). Similarly, the majority of the sample identified as Hispanic (53.2% treatment and 54.5% comparison). There were no differences on sexual behaviors as the students were virgins.

There were 784 students in the analytical sample testing the effects on engagement in risky sexual behavior (the full analytical sample). The average age was just under 15 years old (14.7 for both the treatment and comparison groups). The sample included slightly more girls than

boys (51.5% for the treatment and 51.0% for the comparison group). As in the virginal analytical sample, the majority of students identified as "other" for their race (56.6% treatment and 55.9% comparison). Similarly, the majority of the sample identified as Hispanic (54.9% treatment and 55.4% comparison). There were no significant differences between treatment and comparison groups in sexual behaviors within this matched sample.

F. Baseline equivalence

Diagnostic procedures were performed to assess baseline equivalence of the treatment and comparison groups on age, gender, race, Hispanic ethnicity, and sexual activity (sexual initiation, sex in the last three months, and engagement in risky sexual behavior). Baseline equivalence was assessed using the t-test for continuous variables (age), and for discrete variables (e.g., gender), the t-test was used to compare proportions at each value (e.g., proportion males). Following the identification of an analytical sample using propensity score matching methods, baseline equivalence was demonstrated in Table III.3.A for the analytical sample testing the effects on sexual initiation (those who were virgins at baseline) and in Table III.3.B for the analytical sample testing the effects on engagement in risky sexual behavior (the full analytical sample). These tables give the summary statistics of key baseline measures for the propensity score matched primary analysis described in Appendix D.

Table III.3.A Summary statistics of key baseline measures for matched virginal analytical sample

Baseline measure	Treatment mean or % (standard deviation)	Comparison mean or % (standard deviation)	Intervention versus comparison mean difference	Treatment versus comparison *p*-value of difference
Age (years)	14.7 (0.6)	14.7 (0.6)	0.0	0.73
Gender (female)	55.4%	54.5%	1.0%	0.81
Race: White	21.0%	20.1%	1.0%	0.77
Race: Black	22.0%	21.7%	0.3%	0.92
Race: Other*	57.0%	58.3%	-1.3%	0.75
Ethnicity: Hispanic	53.2%	54.5%	-1.3%	0.75
Ever had sex	0.0%	0.0%	0.0%	NA
Engaged in risky sexual behavior	0.0%	0.0%	0.0%	NA
Sample size	314	314		

*Race: Other includes those students that did not specify race as White or Black and describes students that identify as race American Indian/Alaskan native, Asian, Native Hawaiian/other Pacific Islander, more than one race, and Unknown/not reported.

Table III.3.B Summary statistics of key baseline measures for the matched sample to address engagement in risky sexual behavior

Baseline measure	Treatment mean or % (standard deviation)	Comparison mean or % (standard deviation)	Treatment versus comparison mean difference	Treatment versus comparison *p*-value of difference
Age (years)	14.7 (0.6)	14.7 (0.6)	0.0	0.80
Gender (female)	51.5%	51.0%	0.5%	0.89
Race: White	20.9%	20.9%	0.0%	1.00
Race: Black	22.4%	23.2%	-0.8%	0.80
Race: Other*	56.6%	55.9%	0.8%	0.83
Ethnicity: Hispanic	54.8%	55.4%	-0.5%	0.89
Ever had sex	11.7%	12.0%	-0.3%	0.91
Engaged in risky sexual behavior	2.0%	2.3%	-0.3 %	0.81
Sample size	392	392		

*Race: Other includes those students that did not specify race as White or Black and describes students that identify as race American Indian/Alaskan native, Asian, Native Hawaiian/other Pacific Islander, more than one race, and Unknown/not reported.

G. Methods

1. Impact evaluation

For both primary and secondary research questions, linear probability regression was used to compare outcomes between the treatment and comparison groups adjusting for baseline (first survey) demographic and behavioral variables. Linear probability regression was used instead of logistic regression because of ease of interpretation. The linear probability model was fit for each outcome and each year independently. The models used the following covariates: treatment, baseline status of the outcome, age at baseline, gender, race, ethnicity, baseline measure of sexual activity within the last 3 months (in the full analytical sample including virgins and non-virgins), and the baseline measure of engagement in risky sexual behavior (in the full analytical sample). Equations for the models are given in Appendix E.

Missing data were not imputed and only complete data were used in the primary and secondary analyses. Findings were determined to be significant at the two-sided .05 alpha level. Bonferroni's correction was used to adjust the *p*-values for the multiplicity of primary hypotheses. For the primary analyses the Bonferroni correction multiplied the *p*-values by 2.0. All analyses were performed using R Version 3+ (Vienna, Austria).

Sensitivity analyses were performed comparing estimates using linear probability modeling, a sample that was larger as missing covariates were imputed, and estimates from logistic regression. See Appendix F.

2. Implementation evaluation

Elements of adherence, quality, counterfactual, and context were used to assess implementation of the program, and descriptive statistics were used to describe overall implementation. Program adherence was measured by: average weekly sessions offered, average percentage of sessions attended, percentage of topics covered, and percentage of health educators trained to deliver the program. The quality of staff-student interactions was measured using observation data obtained by evaluation staff. Health educator knowledge, level of enthusiasm, poise and confidence, rapport and communication with students, and effectiveness in addressing questions/concerns were used to rate interactions. Calculations are reported as a percentage of observed interactions for each domain where the score was a 4 or a 5 on a scale of 1-5 where 1 = "Poor" and 5 = "Excellent." The quality of student engagement with the program was measured using two elements from the observation data: "To what extent did the students appear to understand the material?" and "How actively did the group members participate in discussion and activities?" Calculations are reported as a percentage of observed engagements where the score was a 4 or a 5 on a scale of 1-5 where 1 = "Little understanding" and 5 = "Good understanding" and 1 = "Little participation" and 5 = "Active participation," respectively. A

complete description of each implementation element and operationalization is found in Table D.1 in Appendix G.

Counterfactual experiences were calculated using the following survey question: "Have you ever been in a sex education class at school?" A dummy variable was created where students who answered "this school year", "last school year or before", or "last school year and this school year" are coded as 1 and those who responded "no, never participated" are coded as 0. The percentage of comparison students who indicated participation is reported.

Program staff participated in SHAC meetings and compiled a list of other teen pregnancy prevention programming that was made available to the treatment and comparison cohorts, including associated implementation dates and content delivered.

IV. Study findings

The N2K evaluation assessed program effectiveness in delaying sexual initiation and decreasing engagement in risky sexual behavior and used implementation findings to understand the overall context of program implementation.

A. Implementation evaluation findings

The implementation study evaluated four elements of the program: fidelity to program through adherence to curriculum standards and high quality delivery, experiences of the comparison group, and context of the environment in which the program was offered. Results indicate that there was very high implementation fidelity, attendance was consistently above 75% for the treatment group for each year of the program, and there was high quality staff-student interactions and student engagement in the program.

Adherence elements included how much and how often sessions were offered, what and how much was received, what content was delivered, and who delivered the material. The total intended dosage was 16 lessons per year for a total of 48 lessons for three years.

Each year, 16 lessons were delivered to all selected classes. A total of 2,783 lessons were taught over the 3-year period with health educators averaging between 53 and 66 lessons a week. On average, each lesson included four topics and lasted 25 minutes. Health educators reported unplanned adaptations due to insufficient time in 0.08% of lessons and running longer than 25 minutes in 0.4% of lessons. Health educators reported covering 98.7% of topics in 9th grade, 95.7% of topics in 10th grade, and 99.1% of topics in 11th grade.

The dosage received by the treatment students in the analytical sample consistently exceeded 75% for each year of treatment. The average dosage received in the 9th grade was 95%, while the 10th and 11th grade mean dosage was 87%. All treatment students attended some portion of the program over the 3-year period.

The number, type, quality, and training of health educators reflected fidelity to the program model. Four health educators delivered the program and all of them were trained to implement the program, both before they started and through annual refreshers. Health educator qualifications were met. Each health educator had a bachelor's degree, at least two years of related experience, and the same four health educators delivered the program over the three years.

Observations by evaluation staff indicated high quality staff-student interaction and student engagement with the program across the three years. The average percentage of interactions over the three years that received a 4 or a 5 out of 5 points on a five-point scale were: 99.6% for 'Knowledge of Program'; 98.3% for 'Level of Enthusiasm'; 99.1% for 'Poise and Confidence'; 99.6% for 'Rapport with Students'; and 99.7% for 'Effectively Addressed Questions'.

Survey findings for the comparison group show that 80.2% reported receiving sex education in school at baseline, and 94.9% reported receiving sex education in school by the end of the

program. While the percentage of comparison students who reported they received a sex education class at some point during their high school career was above 80%, results from a logistic regression analysis indicate there was no impact on receiving a sex education class for either of the two primary research questions (3% increase in odds of sexual initiation at three years, $p = 0.9$; <1% increase in engagement in risky sexual behavior at three years, $p = 1.0$). However, this exposure was not through a formal sex education curriculum. Within the schools, the SHAC has to recommend sex education curricula, which then has to be approved by the school board, and there were no such approvals during the study period. Therefore, no comparison students attended a health class in school that taught teen pregnancy prevention in any formal program.

Teen pregnancy activities outside of the school in the community, such as at doctor's offices or faith-based, organizations are unknown. There were no external events, such as changes in the Texas legislation budgets, regarding high school students and sex education programs.

B. Impact evaluation findings

The study found no evidence that the N2K program changed the rate of sexual initiation or the rate of engagement in risky sexual behavior at year 3 (Table IV.1). The estimated impact on sexual initiation among virgins at baseline at year 3 (11th grade) was 2.4 percentage points, which indicates more treatment students (46.4%) initiated sex than comparison students (44.0%), but this is statistically not significant ($p = 1.0$). The estimated impact of engagement in risky sexual behavior is 2.0 percentage points, which indicates more treatment students (28.4%) engaged in risky sexual behavior than comparison students (26.4%), but this impact is not statistically significant ($p = 1.0$).

The study also found no evidence that the N2K program changed the rate of sexual initiation among virgins at baseline or the engagement in risky sexual behavior at the end of program years 1 (9th grade) or 2 (10th grade) (Table IV.2).

Table IV.1. Post-intervention estimated effects using data from N2K Final Evaluation Survey at Year 3 to address the primary research questions

Outcome measure	Treatment % (standard deviation)	Comparison %	Treatment compared to comparison (*p*-value of difference)
Ever Had Sex as of Year 3 among virgins at baseline[a]	46.4%	44.0 %	2.4% (1.0)
Engaged in Risky Sexual Behavior at Year 3[b]	28.4%	26.4 %	2.0% (1.0)

Source: N2K Final Evaluation Survey at Year 3

Notes: Linear probability model adjusted for demographics and baseline sexual activity status. *P*-values are Bonferroni corrected.
[a]Sample size is 314 for both the treatment and comparison conditions.
[b]Sample size is 392 for both the treatment and comparison conditions.

Table IV.2. Post-intervention estimated effects using data from N2K Evaluation Survey Years 1-3 to address the secondary research questions

Outcome measure	Treatment mean or % (standard deviation)	Comparison mean or % (standard deviation)	Treatment compared with comparison Mean difference (*p*-value of difference)
Ever Had Sex as of Year 1 among virgins at baseline[a]	13.9%	12.3%	1. 7% (0.54)
Ever Had Sex as of Year 2 among virgins at baseline[a]	29.6%	26.0%	3.6% (0.31)
Engaged in Risky Sexual Behavior at Year 1[b]	9.5%	7.8%	1.7% (0.38)
Engaged in Risky Sexual Behavior at Year 2[b]	15.6%	17.3%	-1.8% (0.50)

Source: N2K Evaluation Survey follow-up administered in spring semester at Years 1 and 2.

Notes: Linear probability model adjusted for demographics and baseline sexual activity status. *P*-values are nominal and not Bonferroni corrected.
[a]Sample size is 314 for both the treatment and comparison conditions.
[b]Sample size is 392 for both the treatment and comparison conditions.

V. Conclusion

This was the first evaluation study of N2K, an innovative teen pregnancy prevention program for adolescents in the 9th, 10th, and 11th grades. Though the N2K intervention was delivered with high quality and fidelity, shifts in knowledge seen throughout the program did not translate into measurable behavioral change among this high school population. Among the analytical sample, mean attendance was consistently near or above 90% across all three years. In

terms of impacts, treatment students who were virgins at baseline were no less likely to initiate sexual intercourse than those virgins in the comparison group by the end of the program. Also, there was no impact on engagement in risky sexual behavior for both virgin and non-virgin treatment students relative to the comparison group at the end of the program. This report concludes there were no effects on sexual initiation or engagement in risky sexual behavior in the interim years 1 or 2, or the final year 3.

Limitations of this study include the non-randomized design that recruited comparison students from the freshman class of 2011 and treatment students from the freshman class of 2012. Due to this study design, the treatment and comparison groups could have differed in unmeasured ways that could have affected the results. Additionally, since treatment and comparison students were in the same schools, there may have been contamination if comparison students interacted with treatment students. Our study design and measures were not able to take this potential exposure into consideration. Moreover, while the percentage of comparison students who reported they received a sex education class prior to the final N2K evaluation survey was 94.9%, results from a logistic regression analysis indicate there was no impact on receiving a sex education class for either of the two primary research questions (3% increase in odds of sexual initiation at three years, $p = 0.9$; <1% increase in birth control non-use at three years, $p = 1.0$).

Future efforts with N2K could modify curriculum content to focus more on desired outcomes – such as abstinence and contraceptive use. Future efforts could also shift the target population to either (1) focus on starting the intervention younger (middle school) and continuing through high school might prove to be more impactful in promoting the desired

behavior changes in late adolescence; or (2) focus on those students at highest risk for engaging in risky sexual behavior (those who have initiated sex prior to beginning high school).

VI. References

Centers for Disease Control and Prevention (CDC). 1991-2013 High School Youth Risk Behavior Surveillance System Data Youth Online, 2014.

Centers for Disease Control and Prevention. Featured. Article: Teen Birth Rates Drop but Disparities Persist, 2015.

Forhan SE, Gottlieb SL, Sternberg MR, Xu F, Datta SD, McQuillan GM, et al. Prevalence of sexually transmitted infections among female adolescents aged 14 to 19 in the United States. Pediatrics. 2009;124(6):1505-12 doi: 10.1542/peds.2009-0674. Epub 2009 Nov 23.

Honaker, J., King, G., Blackwell, M. (2011). Amelia II: A Program for Missing Data. Journal of Statistical Software, 45(7), 1–47.

Martin, J.A., Hamilton, B.E., Ventura, S.J. et al. (2013) Births, Final Data for 2011, National Vital Statistics Reports 62(1).

Martin, J.A., Hamilton, B.E., Ventura, S.J. et al. (2013) Births, Final Data for 2011, National Vital Statistics Reports 62(1).

National Vital Statistics Reports, Births: Final Data for 2013; 64(1).

San Antonio Metropolitan Health District. San Antonio & Bexar County Health & Demographic Statistics: Health Profiles, 2011. (http://www.sanantonio.gov/LinkClick.aspx?fileticket=zAREK7rofAc%3d&tabid=1663&portalid=0&mid=9638 accessed 7/25/2015).

Satterwhite CL, Torrone E, Meites E, Dunne EF, Mahajan R, Ocfemia MC, et al. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. Sex Transm Dis. 2013;40(3):187-93.

Sekhon, J.S. (2011) "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R." *Journal of Statistical Software* 42(7): 1-52.

Texas Department of State Health Services. Texas STD Surveillance Report, 2014.

Texas Education Agency. 2011-12 Academic Excellence Indicator System. District Report, 2012.

Appendix A: Impact evaluation data collection

Table A.1. Data collection efforts used for the impact analysis of N2K and their timing (month/year)

Data collection effort	Comparison	Treatment
Baseline survey	09/11	09/12
First Follow-up	04/12	04/13
Second Follow-up	04/13	04/14
Final Follow-up	04/14	04/15

Table A.2. Summary of impact data collection procedures used in impact analysis of N2K

Data collection effort	Parental Consent	Baseline	First Follow-up	Second Follow-up	Final Follow-up
Survey mode*	Written signature	Self-administered paper survey in school setting	Self-administered paper survey in school setting	Self-administered paper survey in school setting	Self-administered paper survey in school setting
Incentive	$10 iTunes gift card for students One $100 gift card and five $50 gift cards raffled for parents $25 gift card for teachers	$10 iTunes gift card	$10 iTunes gift card	$10 iTunes gift card	$10 iTunes gift card
Staff involved	Program staff	Trained evaluation staff and proctors	Trained evaluation staff and proctors	Trained evaluation staff and proctors	Trained evaluation staff and proctors

*Surveys were mailed to all students not present during testing, using their last known address, and were accepted for up to one month after in-school surveying during each surveying period.

Appendix B: Implementation evaluation data collection

Table B.1. Data used to address implementation research questions

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Adherence: How often were sessions offered? How many were offered?	The number and frequency of sessions were captured with HanDBase iPad program and converted into a fidelity log Excel database.	Data were collected once a week for each class taught (16 lessons in each curriculum manual).	Program Staff (Health Educators)
	Each session was intended to be 25 minutes and was taught during half of a 50 minute period. Health educators were instructed to complete the session in the 25 minutes, as written.	Length was not recorded; however, deviations from the 25-minute length (+/- 5 min) were reported by the health educators in the fidelity log. Health educators were specifically trained to deliver the curriculum in the allotted time.	
Adherence: What and how much was received?	The number and type of sessions were captured with HanDBase iPad program and converted into an attendance log Excel database.	The number and type of sessions were recorded at the end of every session and were collected once a week for each class taught.	Program Staff (Health Educators)
	Attendance data were used to calculate the average number of sessions attended as the dosage received.	The attendance at the individual level was collected for every session and was entered into HanDBase iPad each day.	
Adherence: What content was delivered to youth?	The number of topics covered was self-reported by health educators immediately after each session. The health educator completed a fidelity log on each and every session they delivered using the HanDBase iPad program. Logs were emailed to the evaluator at the end of each week of teaching.	Data were collected once a week for each class taught	Program Staff (Health Educators)
	The fidelity log for each session listed the intended topics. The health educator recorded compliance with the topics and documented any unplanned adaptations. If content was not delivered during a session, it was not covered in any future sessions.		
Adherence: Who delivered material to youth?	A list of health educators hired and trained to deliver the curriculum, along with their background qualifications and health educator training records was maintained. There were four health educators and one program coordinator trained to deliver the curriculum.	Health educator training data were collected before implementation and any training record updates were maintained by program coordinator. Annual refreshers on adolescent sexual health were provided before the start of a new school year.	Program Coordinator and Program Director

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Quality: Quality of staff-student interactions	Observations of staff-student interaction quality using protocol developed by evaluation staff was collected in HanDBase and converted into a quality Excel database.	Systematic sampling of 10% of classroom sessions was conducted once a week. Each health educator was sampled in each class at least once.	Evaluation Staff
Quality: Quality of youth engagement with program	Observations of student engagement using protocol developed by evaluation staff was collected in HanDBase and converted into a quality Excel database.	Systematic sampling of 10% of classroom sessions was conducted once a week. Each health educator was sampled in each class at least once.	Evaluation Staff
Counterfactual: Experiences of comparison condition	Survey item #7 on baseline and follow up assessments, which asks "Have you ever been in a sex education class at school?"	Data collection occurred at baseline, end of the first year, end of the second year, and final survey time points.	Evaluation Staff
	Telephone calls with high school health teachers were made to obtain class content and to see if their curriculum had changed since last year.	Calls were conducted before program implementation and the beginning of every school year during the course of the program.	Program Coordinator
	Information was gleaned from participation in school district's monthly School Health Advisory Council (SHAC). The SHAC reviews and recommends all health education information, including teen pregnancy prevention programs that will be implemented in the school, including health classes. Participating in this council allowed UT Teen Health to know if any changes regarding teen pregnancy prevention would be occurring in the intervention schools.	Monthly meetings	Program Coordinator
	School rosters were used to track the comparison student with parental consent who were enrolled in health class.	Rosters were collected at the beginning of each semester.	Evaluation Staff

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Context: Other TPP programming available or offered to study students (both intervention and comparison)	Information was gleaned from participation in school district's monthly School Health Advisory Council (SHAC) where any TPP programming is discussed, recommended to the school board, or refused according to state law.	Monthly meetings	Program Coordinator
	Health class rosters were obtained from high school administration.	Data were collected at the beginning of each semester to document comparison and treatment students' participation in a health class that offers TPP programming.	Evaluation Staff
Context: External events affecting implementation	Information involving the school district from local news outlets; national news outlets	Ad hoc	Program Coordinator
	Meetings with school administration	Monthly updates with administrator by phone, email, or meeting.	Program Coordinator
	Ongoing contact with city-wide, federally funded clinics, healthcare officials, and physicians	Ad hoc	Program Director
	Phone calls with clinics surrounding the school district obtaining updated information on services and clinic hours, which is provided to the school nurses and counselors	Annually	Program Coordinator and Program Staff
	Information gleaned from the school district's School Health Advisory Council (SHAC)	Monthly SHAC meetings	Program Coordinator
	Texas legislation related to high school students and sex education	Ad hoc	Program and Evaluation Staff
Context: Substantial unplanned adaptation(s)	Fidelity monitoring logs specifically identify the topics referred during each session and any unplanned adaptations are documented in these logs	After each and every session for all educators, then compiled and reviewed by evaluation staff on a weekly basis	Program Director, Program and Evaluation Staff

SHAC = School Health Advisory Council, TPP = Teen Pregnancy Prevention

Appendix C: Study sample

Table C.1. Student sample sizes by intervention status, for the full sample

Number of students	Time Period	Total sample size	Treatment sample size	Comparison sample size	Total response rate	Treatment response rate	Comparison response rate
Assigned to condition		1,616	856	760			
Contributed a baseline survey	Fall 2011 / Fall 2012	1,568	822	746	97%	96%	98%
Virgins who contributed a baseline survey	Fall 2011 / Fall 2012	1114	604	510	NA	NA	NA
Contributed a follow-up survey (First follow-up, 9th grade)	Spring 2012 / Spring 2013	1,452	752	700	90%	88%	92%
Virgins who contributed a follow-up survey (First follow-up, 9th grade)[a]	Spring 2012/ Spring 2013	1,043	566	477	94%	94%	94%
Contributed a follow-up survey (Second follow-up, 10th grade)	Spring 2013 / Spring 2014	1,224	662	562	76%	77%	74%
Virgins who contributed a follow-up survey (Second follow-up, 10th grade)[a]	Spring 2013 / Spring 2014	884	490	394	79%	81%	77%
Contributed a follow-up survey (Final, 11th grade)	Spring 2014 / Spring 2015	1,165	621	544	72%	73%	72%
Virgins who contributed a follow-up survey (Final, 11th grade)[a]	Spring 2014 / Spring 2015	878	486	392	83%	84%	81%
Contributed all four surveys		901	476	425	57%	56%	57%
Virgins who contributed all four surveys[a]		724	397	327	64%	66%	64%
Propensity Score Matching Sample: Full Analytical sample		784	392	392	NA	NA	NA

Number of students	Time Period	Total sample size	Treatment sample size	Comparison sample size	Total response rate	Treatment response rate	Comparison response rate
Propensity Score Matching Sample: Virginal Analytical Sample [a]		628	314	314	NA	NA	NA

[a]The denominator for the response rate for the virgins samples is the number of virgins with baseline data since a student's status as a virgin was not known until the baseline surveys were completed.

Appendix D: Propensity score matching and baseline equivalence

Because of a lack of baseline equivalence, we performed propensity score matching for the benchmark analyses to answer the primary and secondary research questions. The propensity score model was a logistic regression of the outcome of treatment allocation and baseline covariates including variables defining engagement in risky sexual behavior such as sexual activity, sex without condom, and sex without contraception within the last 3 months, sexual initiation, age, gender, race, and ethnicity. The logistic regression model for the propensity score of treatment assignment probability p_i for the i^{th} student is given by $logit(p_i) = \sum_{k=1}^{6} \beta_k x_{ik}$ where $logit(\ )$ is the log-odds transformation of the outcome probability, β_k is the effect of the k^{th} baseline covariate, and x_{ik} is the value of the k^{th} covariate for the i^{th} student. The propensity scores that were produced were used to inform the matching process—that is students with similar propensity scores were matched together.

This matching analysis was performed using individuals that had complete analytic data at the baseline time point. Outlying propensity scores of greater than 80% or less than 20% treatment assignment probability were removed. The nearest neighbor approach was implemented using the Matching R package (Sekhon, 2011). The nearest neighbor for each treatment individual was a comparison individual with propensity scores were within a standardized distance caliper of 0.3. This caliper excludes matches of treatment and comparison students that were far apart. In addition to using a nearest neighbor approach for identifying matches, we also created a matched sample using a weighted exact match approach. This exact match method identified individuals who were identical across treatment and control conditions, and allowed for one-to-many matches (which was analyzed using sample weights). The method

of matching used did not affect the conclusions in the study – that is, the results were robust across the two different approaches.

Appendix E: Linear probability model

To increase the precision of the estimates of the effect of intervention and to account for any differences in baseline variables remaining after propensity score matching, we performed linear probability regression for the benchmark analyses to answer the primary and secondary research questions. The linear probability model used was a multiple linear regression of the binary (yes=1, no=0) outcomes sexual initiation (yes/no) and engagement in risky sexual behavior (yes/no). This model adjusts for effect of baseline variables or covariates that are related to outcomes. The linear probability regression model for all research questions are given by $p_i = \beta_0 + \sum_{k=1}^{6} \beta_k x_{ik} + \beta_I x_{Ii} + \in_i$ where p_i is outcome status (1 for yes and 0 for no) of the i^{th} student, β_k is the effect of the k^{th} baseline covariate, and x_{ik} is the value of the k^{th} covariate for the i^{th} student and β_I is the effect of the intervention with x_{Ii} the indicator of the treatment status of the i^{th} student. The term $\in_i$ represents the Guassian random error. The covariates included in the model were age at baseline, gender, race, ethnicity (Latino or not) and baseline measure of sexual activity. Baseline sexual activity was only a covariate in the analysis of the full analytical set (including virgins and non-virgins). Both sexual activity (sexual activity in the last 3 months) and baseline engagement in risky sexual behaviors were included as covariates in the full analytical sample. Race was coded as three groups 1) African American, 2) Other and the reference group 3) White. Sexual activity at baseline (yes/no) was included in the model for the engagement in risky sexual behavior outcome, but excluded for the analysis of sexual initiation because the baseline status of the sexual activity was “no” for all subjects (virgins) in the analytical sample of sexual initiation.

Appendix F: Sensitivity analyses

This appendix evaluates the sensitivity of estimates to the various methodological decisions. The motivation for this is that the selection of the statistical models could affect the conclusion drawn in answering the primary or secondary research questions. If the results from the alternative analytical approaches are consistent with the primary analysis then the results are shown to be robust.

The benchmark analysis for the primary research questions was compared with two alternative analytical approaches. The benchmark analysis used a propensity score matched sample that was restricted to subjects with complete data at baseline and at years 1, 2, and 3. For the first sensitivity analysis, we performed a multiple imputation alternative that imputed the missing baseline covariates using a bootstrap EM algorithm on incomplete data with the R package Amelia (Honaker, 2011). This approach was inclusive of subjects that were missing intermediate time points (such as years 1 and 2), and hence, took advantage of all of the subjects with outcome data collected at the final time point when performing propensity score matching. The outcome data were not imputed, so that observations with missing outcomes were not matched. For the primary research question, this approach resulted in an analytical sample of 744 virginal students (372 treatment and 372 comparison) and 956 students (478 treatment and 478 comparison).

Second, we compared the linear probability model to logistic regression. Logistic regression gives an odds ratio for the treatment effect instead of a percentage difference. The advantage of logistic regression is that it assumes multiplicative effect of variables on the odds of a binary outcome, which is appropriate for the binary primary endpoints. The sample sizes for the logistic regression sensitivity analysis were the same as they were for the benchmark analysis.

The results of both sensitivity analyses are given in Table F.1. The comparisons of the resulting effect estimates indicate very little sensitivity to the analytical approach. None of the approaches to the primary analyses indicate a statistically significant difference due to treatment assignment.

The benchmark analysis for the secondary research questions was also compared with the same two alternative analytical approaches. These results are given in Table F.2. The comparisons of the resulting effect estimates indicate very little sensitivity to the analytical approach. None of the approaches to the secondary analysis indicate a statistically significant difference due to treatment assignment. The sample sizes for the logistic regression sensitivity analysis were the same as they were for the benchmark analysis. Multiple imputation of baseline covariates was repeated for years 1 and 2. The analytical sets were derived using propensity score matching applied to all students who reported the endpoints at each year. The sample sizes for the multiple imputation matching analysis related to the secondary research questions were: 924 (462 treatment and 462 comparison) virginal students and 1198 (599 treatment and 599 comparison) students with survey responses in year 1, and 752 (376 treatment and 376 comparison) virginal students and 976 (488 treatment and 488 comparison) students with survey responses in year 2.

Table F.1. Sensitivity of impact analyses using data from N2K Final Evaluation Surveys to address the primary research questions

N2K compared with comparison	Benchmark approach difference	Benchmark approach *p*-value	Matched after Imputation difference	Matched after Imputation *p*-value	Logistic Regression Odds Ratio	Logistic Regression *p*-value
Ever Had Sex Year 3 among virgins	2.42%	1.0	3.59%	0.64	1.1	1.0

N2K compared with comparison	Benchmark approach difference	Benchmark approach *p*-value	Matched after Imputation difference	Matched after Imputation *p*-value	Logistic Regression Odds Ratio	Logistic Regression *p*-value
Engaged in Risky Sexual Behavior Year 3	2.03%	1.0	-0.13%	1.0	1.0	1.0

Source: N2K Evaluation Survey Year 3 responses to primary research questions.

Notes: The *p*-values represent the Bonferroni adjustment for multiple comparisons.

Table F.2. Sensitivity of impact analyses using data from the N2K Evaluation Survey to address the secondary research questions

N2K compared to Comparison	Benchmark approach difference	Benchmark approach *p*-value	Matched after Imputation difference	Matched after Imputation *p*-value	Logistic Regression Completers Odds Ratio	Logistic Regression Completers *p*-value
Ever Had Sex Year 1 among virgins	1.67%	0.54	1.21%	0.37	1.2	0.54
Ever Had Sex Year 2 among virgins	3.64%	0.31	1.74%	0.60	1.2	0.30
Engaged in Risky Sexual Behavior Year 1	1.68%	0.38	3.11%	0.06	1.3	0.33
Engaged Risky Sexual Behavior Year 2	-1.76%	0.50	-2.51%	0.30	0.87	0.48

Source: N2K Evaluation Survey Follow-up at Years 1, 2, and 3

Appendix G: Implementation evaluation methods

Table G.1. Methods used to address implementation research questions

Implementation element	Methods used to address each implementation element
Adherence: How often were sessions offered? How many were offered?	The total number of sessions was the sum of the sessions captured in HanDBase and placed into the Components Excel database. The average weekly frequency was calculated as the total number of sessions divided by the total number of weeks of implementation. The average dosage received by completers was calculated at survey completion. This was reported as the average number of sessions received and as a percentage of dosage offered.
Adherence: What and how much was received?	The average number of sessions attended per student was calculated using the attendance log that was collected at every session and uploaded weekly and converted into the Attendance excel database. This database indicates the frequency and content of sessions completed and the students' attendance. The average number of sessions attended was calculated as the average of the number of sessions that each student attended. The percentage of sessions attended was calculated by dividing the total number of sessions attended by the student by the total number of sessions offered. Additionally, the percentage of topic areas covered was calculated.
Adherence: What content was delivered to youth?	The total number of topics covered was recorded in HanDBase and exported into the Components Excel database. The proportion of topics actually covered was recorded during the evaluator's weekly observations. The proportion of topics covered was calculated by dividing the number of topics covered by the total number of topics scheduled during a session. These data were exported into the Fidelity Excel database.
Adherence: Who delivered material to youth?	The total number of educators delivering the program was a count of trained educators who implemented the program. The Program Coordinator and Program Director maintained training records for the duration of the study. The percentage of educators trained was calculated by dividing the number of educators who delivered the program by the total number of educators who were trained.
Quality: Quality of staff-student interactions	The quality of staff-student interaction was calculated using the observations of Quality made by the evaluator. Question 6 on the OAH Observation Form was used to measure quality of staff-student interactions. The interaction indicator was calculated as a percentage of observed interactions where the score of the interaction was a higher quality score (4 or 5 on a scale out of 5, where 1 = Poor and 5 = Excellent). The percentage of "higher quality" interactions was calculated by dividing the number of observations that were scored as "higher quality" by the total number of observations scored. Furthermore, staff quality was analyzed based on number and type of degrees attained for each health educator, years of teaching experience, and training in specified curriculum.
Quality: Quality of youth engagement with program	The quality of student engagement was calculated using the observation data recorded by the evaluator at observation visits. Question numbers 4 and 5 on the OAH Observation Form was used as measures of student engagement. The engagement indicator was calculated as a percentage of observed interactions where the engagement was "higher quality" (4 or 5 on a scale out of 5, where 1 = Poor and 5 = Excellent). The percentage of "higher quality" interactions was calculated by dividing the number of observations that were scored as "higher quality" by the total number of observations scored. Question 7 on the OAH Observation Form was used to assess the overall quality of the program session.
Counterfactual: Experiences of counterfactual condition	The percentage of comparison students who answered in the affirmative to question #7 on the survey, which asks "Have you ever been in a sex education class at school?", conducted at baseline, end of the first year, end of the second year, and end of the program was calculated. The percentage of comparison students who receive a health class that covered a topic related to pregnancy prevention during our implementation timeframe was documented.

Implementation element	Methods used to address each implementation element
Context: Other TPP programming available or offered to study students (both intervention and counterfactual)	Any TPP programming available to both treatment and comparison groups was monitored through the program coordinator contacting health teachers and school administrators annually, and participating in the SHAC. The date of implementation of any other TPP programming and the type of content delivered (e.g., STDs, contraception, etc.) was documented. The percentage of treatment and comparison students who received a health class that covered a topic related to pregnancy prevention during our implementation timeframe was documented.
Context: External events affecting implementation	External events in the area's health care availability, large-scale catastrophic events or factors which might affect outcomes between the comparison and treatment groups were reported. Any changes in Texas Legislation budgets regarding high school students and sex education programs were also described, using local news outlets, interviews with local healthcare providers and school administration.
Context: Substantial unplanned adaptation(s)	Any change in program staff, setting, or delivery content or dose was described. This report documents the percentage of all topics that had unplanned adaptations, and the percentage of topics that deviated -5 minutes from the 25 minute session and +5 minutes from the 25 minute session.

TPP = Teen Pregnancy Prevention; SHAC = School Health Advisory Council; STIs = Sexually Transmitted Infections