Article

Early Engagement in Course-Based Research Increases Graduation Rates and Completion of Science, Engineering, and Mathematics Degrees

Stacia E. Rodenbusch,† Paul R. Hernandez,‡ Sarah L. Simmons,§ and Erin L. Dolan†*

†Texas Institute for Discovery Education in Science, College of Natural Sciences, University of Texas at Austin, Austin, TX 78712; ‡Department of Learning Sciences and Human Development, College of Education and Human Services, West Virginia University, Morgantown, WV 26506; §Howard Hughes Medical Institute, Chevy Chase, MD 20815-6789

National efforts to transform undergraduate biology education call for research experiences to be an integral component of learning for all students. Course-based undergraduate research experiences, or CUREs, have been championed for engaging students in research at a scale that is not possible through apprenticeships in faculty research laboratories. Yet there are few if any studies that examine the long-term effects of participating in CUREs on desired student outcomes, such as graduating from college and completing a science, technology, engineering, and mathematics (STEM) major. One CURE program, the Freshman Research Initiative (FRI), has engaged thousands of first-year undergraduates over the past decade. Using propensity score–matching to control for student-level differences, we tested the effect of participating in FRI on students' probability of graduating with a STEM degree, probability of graduating within 6 years, and grade point average (GPA) at graduation. Students who completed all three semesters of FRI were significantly more likely than their non-FRI peers to earn a STEM degree and graduate within 6 years. FRI had no significant effect on students' GPAs at graduation. The effects were similar for diverse students. These results provide the most robust and best-controlled evidence to date to support calls for early involvement of undergraduates in research.

INTRODUCTION

Undergraduate research experiences (UREs) are seen as integral to training the next generation of scientists, driving governmental and philanthropic agencies to invest millions of dollars annually to support undergraduate research internships (Sadler *et al.*, 2010; American Association for the Advancement of Science [AAAS], 2011; President's Council

of Advisors on Science and Technology [PCAST], 2012). A growing body of research documents the positive outcomes of UREs. Undergraduates who conduct research in science, technology, engineering, or math (STEM) report cognitive gains such as learning to "think and work like a scientist," affective gains such as finding research enjoyable and exciting, and behavioral outcomes such as increased intentions to pursue further education or careers in science (Seymour *et al.*, 2004; Laursen *et al.*, 2010; Lopatto and Tobias, 2010). An increasing number of well-controlled, large-scale, and longitudinal studies indicate that UREs can attract, retain, and improve the success of undergraduates in STEM (Estrada *et al.*, 2011; Eagan *et al.*, 2013; Hernandez *et al.*, 2013). These results have been the impetus for calls for widespread involvement of undergraduate students in research (AAAS, 2011).

The apprenticeship structure of UREs, in which an undergraduate works one-on-one with a more experienced researcher, such as a faculty member, postdoctoral scientist, or graduate student, limits the number of undergraduates who can participate in research. This limitation, coupled

with interest in expanding the availability and accessibility of research experiences and the high cost associated with apprenticeships, has driven the development of courses that engage students in doing research, also called discovery-based research courses or course-based undergraduate research experiences (CUREs; Wei and Woodin, 2011; PCAST, 2012; Auchincloss *et al.*, 2014; National Academies of Sciences, Engineering, and Medicine, 2015). CUREs involve students in addressing a research question or problem that is of interest to the scientific community in the context of a class (Auchincloss *et al.*, 2014). When compared with traditional lab courses, CUREs afford opportunities for students to make discoveries that are relevant to stakeholders outside the classroom, including practicing scientists, and to engage in iterative work such as troubleshooting, problem solving, and building off one another's progress in a way that more closely resembles the practice of STEM (Auchincloss *et al.*, 2014; Corwin *et al.*, 2015b).

One example of a national-level, upper-division, single-semester CURE is the Genomics Education Partnership, in which students enrolled in a genomics-related course finish raw *Drosophila* genome sequence data and annotate genes and other genome features as part of addressing a larger research question related to *Drosophila* genome evolution (Lopatto *et al.*, 2008; Leung *et al.*, 2010). The Science Education Alliance–Phage Hunters program is an example of a national-level, introductory, two-semester CURE in which students identify and characterize novel soil bacteriophages in the context of a two-semester introductory biology course series (Hatfull *et al.*, 2006; Jordan *et al.*, 2014). Other CURE models involve addressing a range of research questions using a common, centrally supported technology, such as high-throughput sequencing (Buonaccorsi *et al.*, 2011, 2014), and local CUREs, in which faculty members integrate an aspect of their research into courses they teach at their own colleges or universities (Bascom-Slack *et al.*, 2012; Kloser *et al.*, 2013; Harvey *et al.*, 2014).

CUREs have the potential to make research experiences available at scale, rather than to a select few who seek out research internships or are handpicked by faculty (Auchincloss *et al.*, 2014). Because CUREs can be offered at the introductory level, they have greater potential to change students' educational and career trajectories than research internships, which are mostly available to students later in their undergraduate careers, in junior or senior year. This enormous potential has led to rapid growth in the number of CUREs and recommendations for their widespread adoption (AAAS, 2011; PCAST, 2012), despite critiques that point out the dearth of evidence of their effectiveness and impact (Linn *et al.*, 2015). Most studies of CURE effectiveness or impact rely on student self-report of knowledge and skill gains or intentions to pursue graduate education in STEM or science-research related careers, rather than more direct measures of achievement and retention in STEM. However, several CUREs have been in operation long enough to examine longer-term effects for students—especially whether CURE participation influences students' persistence and success in STEM and in college in general.

The Freshman Research Initiative (FRI) at the University of Texas at Austin (UT Austin) is a CURE program that was established to improve the learning experiences of undergraduates in the College of Natural Sciences (CNS), about half of whom are life science majors. The program is described in greater detail elsewhere (Beckham *et al.*, 2015) and summarized here as context for this study. The full FRI program is a three-course series, which we refer to here as Courses 1, 2, and 3 for simplicity. FRI students first complete a research methods course (Course 1), followed by up to two semesters of course-based research (CUREs) in one of 25+ different areas, called "research streams" (Courses 2 and 3). Current research streams are offered in a range of science disciplines, including biology, biochemistry, bioinformatics, chemistry, computer science, physics, and astronomy (see https://cns.utexas.edu/fri for a complete list). Students earn three credit hours for each course, which translates to roughly 9 h of lab-related work per week. In addition, each course helps students make progress toward completing their degrees: Course 1 counts toward university requirements, Course 2 counts as an introductory lab credit, and Course 3 counts as an upper-division lab or research credit.

In Course 1, students learn to search and read scientific literature, and they design and execute one or more scientific investigations, called inquiries, which they summarize in written and oral reports. During this semester, they also participate in a matching process through which they are assigned to a stream. In Course 2, students learn about the overarching research goals for their stream, complete instructional modules to learn concepts and skills specific to the research, and begin to contribute to the stream's research. In Course 3, students become more independent, often proposing and carrying out their own independent subproject using the skills and understanding they developed in Course 2. Depending on the research, students may either work side by side on parallel projects or as a member of a team on a component of the research. As an example, after completing Course 1, students might join the Supramolecular Sensors Stream, and make use of spectroscopy, chromatography, organic synthesis, and biochemical techniques to create and utilize peptide-based sensors to differentiate wine varietals. These students can choose to earn either a general biology or general chemistry lab credit for Course 2 and either independent biology or chemistry research credit for Course 3. Courses 1 and 3 are writing intensive; students who complete these courses also complete a university writing requirement.

Each section of Course 1 enrolls 25 students and is taught by a PhD-level lecturer. Each stream (Courses 2 and 3) enrolls up to 40 students. These courses are led by a PhD-level research educator (RE), who is a hybrid of an instructor and a research scientist hired as a non–tenure-track faculty member or postdoctoral associate, and an individual or team of tenure-track or tenured principal investigators (PIs). A small number of streams enroll only 15 students per semester and are led by graduate students who serve in the role of RE. The RE role is unique and essential to FRI, because each RE mentors a team of up to 40 undergraduate researchers, which would not be practical in a more traditional research group structure. In all semesters of FRI, additional instructional support is provided by undergraduate peer mentors who previously participated in FRI and who help to create an environment that reflects the tiered expertise typical of a research group or community of practice (Wenger, 1999; Lave and Wenger, 1991). In Courses 1 and 2, a graduate or undergraduate teaching assistant provides additional research mentorship and instructional support.

FRI was launched with 40 students in 2005 and now serves ~900 students per year, which is ~40% of the incoming class in the CNS. A sufficient number of students have participated in FRI to examine its effectiveness in terms of direct, long-term student outcomes. Specifically, this analysis assessed the degree to which participation in FRI influenced students' probability of graduating with a STEM degree, probability of graduating within 6 years regardless of major, and educational performance in terms of cumulative grade point average (GPA) at graduation when compared with a matched sample of their peers.

METHODS

Participants

A sample of 4898 students was drawn from the population of students enrolling at UT Austin between 2006 and 2013 ($N = 75,767$). This study was designed to test the intermediate- and long-term impacts of the FRI on academic performance and persistence in a STEM major. This study primarily compared students who completed all three semesters of the FRI program with a group of propensity score–matched control students. In this paper, we report data from students first year, junior year, and graduation year (typically fourth or fifth year of enrollment at UT Austin). We restricted the sample to students with complete information for the variables used in the propensity score analysis ($N = 53,603$; see *FRI Program Variables*). Students enrolled in programs that guaranteed FRI enrollment were also omitted (i.e., Biology Scholars program, Emerging Scholars program, Women in Science program, Dean's Scholars Honors program, and Public Health Honors program; $N = 52,619$). A propensity score–matching procedure was conducted on the resulting sample of FRI ($n = 2648$) and non-FRI students ($n = 49,971$). Finally, the analytical sample used in data analysis was restricted to propensity score–matched FRI and non-FRI students ($N = 4898$; $n_{FRI} = 2449$ and $n_{non-FRI} = 2449$). About 93% of FRI students had a close propensity score–matched non-FRI student and were thus included in the final analytical sample.

FRI Program Variables

The following variables measured "participation" in the FRI program. FRI is a three-course CURE program. Participation in each of the three courses was measured by enrollment data collected from the registrar's office after the add/drop period ended on the 12th class day of the semester. Participation in Course 1, which students complete in the Fall of their freshman year, was dummy coded (0 = matched control group, 1 = FRI group) for all analyses. Courses 2 and 3 represent the lower- and upper-division research courses of FRI, which students complete in the Spring of their freshman year and Fall of their sophomore year, respectively. Participation in each semester was measured by enrollment data collected from the registrar's office after the add/drop period ended on the 12th class day of the semester. Spring participation (Course 2) and Fall participation (Course 3) were each dummy coded (Course 2: 0 = did not participate, 1 = participated; Course 3: 0 = did not participate, 1 = participated) for all analyses.

Identification of Matched Samples of FRI and Non-FRI Participants

To conduct an analysis of the effect of FRI participation, we first had to identify an appropriate control group of nonparticipating students. We used a propensity score–matching procedure to calculate the probability that a student would be in FRI based on a set of observed covariates in order to correct for selection bias when creating a matched control group (West *et al.*, 2008). The propensity score model (i.e., logistic regression) included 13 variables used in the FRI admissions process to generate a propensity score (from 0 to 1) for each student in the MatchIT software program (Ho *et al.*, 2007, 2011; Thoemmes, 2011). Regarding the variables that influence admissions into FRI, the minimum requirement for entry is a passing score (70%) on a math competency test. Students in several specialty programs in the CNS, such as the Women in Natural Sciences program, are automatically admitted to FRI. These account for ~30% of the FRI population. Students from groups underrepresented in the sciences, such as those with family income less than $40,000 per year, those who are first in their families to go to college, women majoring in physical sciences, computer science, or math, and students with low SAT scores, are also selected for admission. These students account for ~40% of the FRI population. The remaining ~30% of FRI students apply to the program. Applicants are given priority based on their membership in one of the underrepresented groups described above. Finally, there is some attrition from FRI after each semester. Seats that become available in Courses 2 and 3 are filled with students from the applicant waiting list.

We used the following sociodemographic characteristics as matching variables, because they are associated with admission into FRI and persistence in STEM: gender, race/ethnicity, parental education levels, parental income level, and Pell grant eligibility (Schneider *et al.*, 1997; Riegle-Crumb *et al.*, 2012; Supplemental Table S1). We also included variables that have been shown to be associated with enrollment in FRI and students' choice to major in STEM: SAT total score or ACT equivalent as a measure of prior academic achievement, number of high school science credits earned as a measure of science preparation, and number of high school math credits earned as a measure of math preparation (Wang, 2013). We included the following additional variables in the matching procedure, because they affected students' likelihood of enrolling in FRI and thus may have resulted in a selection bias: whether students graduated from a Texas or out-of-state high school, the first year students enrolled at UT Austin (e.g., 2006), the first semester students enrolled at UT Austin (entry in Fall is on cycle with FRI admissions), the first college students entered at UT Austin (CNS students are prioritized), and enrollment in the Texas Interdisciplinary Program, a community-building program in the college.

We used FRI students' propensity scores to identify comparable non-FRI control students (see Supplemental Material for details). The propensity score–matching procedure resulted in two groups of equal size (FRI group $n = 2449$ and matched control group $n = 2449$). The percent bias reduction on the matching covariates was 98% in the matched sample (Supplemental Figure S1 and Supplemental Table S2). The following analysis was restricted to matched pairs in which

the FRI student participated in Course 1 alone ($n = 416$), both Courses 1 and 2 ($n = 882$), or the complete FRI program (i.e., Courses 1, 2, and 3; $n = 1151$), and the non-FRI student participated in no FRI courses. In addition, analysis was restricted to matched pairs in which both students had scores on the outcome and complete data on all predictors.

Outcomes

The following variables measured outcomes relevant to participation in FRI.

Earned Baccalaureate Degree in STEM. Students who had graduated earned degrees in a variety of colleges (e.g., natural sciences, engineering). The college of earned degree variable was recoded into a STEM degree dummy-coded variable (0 = non-STEM college, 1 = STEM college), with only the colleges of natural sciences and engineering coded as STEM colleges. Mathematics and computer science degrees are earned from the CNS.

Earned Any Degree within 6 years of Entry. Student graduation from UT Austin within 6 years of entry was measured by coding graduation versus nongraduation by Spring of 2015. This variable was dummy coded to represent graduation or nongraduation (0 = had not graduated within 6 years of entry, 1 = graduated with a degree within 6 years of entry). Because our focus was on students who had the opportunity to graduate within 6 years, this analysis was restricted to students in our data set entering UT Austin on or before 2009 (i.e., we had graduation data for students up to Spring 2015).

Cumulative GPA at Graduation. Cumulative college GPA was measured at graduation. Cumulative GPA was measured on a scale from 0 to 4.

Cumulative GPA at Midpoint of College Tenure. Cumulative college GPA was measured at the midpoint of the undergraduate college tenure (i.e., Fall of junior year). Cumulative GPA was measured on a scale from 0 to 4.

Control Variables. All covariates used in the propensity score–matching process were also used as variables in the regression analyses to control for chance imbalances across groups (Schafer and Kang, 2008). Control variables included: gender (female, male), race/ethnicity (Asian, Hispanic, white, or other), enrollment in Texas Interdisciplinary Program (yes, no), SAT total score (or ACT equivalent), Pell grant eligibility (yes, no), number of units of science on high school transcript, number of units of math on high school transcript, how students were initially accepted into UT Austin (Texas high school, other), first year enrolled at UT Austin (2006, 2007, 2008, 2009, 2010, 2011, 2012, or 2013), maternal and paternal education levels (less than college degree, college degree [2 or 4 years], or advanced degree), parental income level (≤ $39,999, $40,000 to $79,999, $80,000 to $99,999, or ≥ $100,000 per year), and first college entered at UT Austin (STEM-related college = natural sciences or engineering, or non–STEM-related college such as education or business).

Treatment of Missing Data

For the sample of 4898 participants, 12.7% of participants were missing data on their midcollege cumulative GPA, 43.0% were missing data for their cumulative GPA at graduation (i.e., had not yet graduated by Summer 2015), 41.6% were missing data on their major at graduation and time to degree completion (i.e., had not yet graduated by Summer 2015). Our matched sample included both those who had time to graduate (i.e., 4 years for traditional students or 2 years for transfer students) and a smaller number of those who did not (i.e., their first year enrolled was 2012 or 2013). Although those who did not have time to graduate (and their matched control) did not contribute to analyses related to any of the graduation outcomes (i.e., cumulative GPA, STEM degree, 6-year graduation rate), they were retained because they contributed to the analysis of the FRI effect on midpoint GPA, which was a suspected mediator of the FRI effect on cumulative GPA (see Supplemental Material for details).

To ensure unbiased estimates of the effect of FRI, we only used whole linked pairs of participants in which both the FRI and matched control participant provided data for the analysis. This approach restricted our analytical sample of the STEM degree and cumulative GPA outcomes to cases in which both members of the matched pair (i.e., both the FRI student and the matched counterpart) graduated in or before Summer 2015 (regardless of the number of years to degree). This approach also restricted our analytical sample of the 6-year graduation outcome to cases in which both the FRI student and the matched counterpart started at UT Austin on or before 2009 and thus had the opportunity to graduate within 6 years (e.g., for those starting in Fall 2006, graduation by Summer 2013; for those starting in Fall 2009, graduation by Summer 2015). To account for missing data and to account for chance imbalances on covariates used to estimate the propensity scores, we controlled for all covariates used in the propensity score matching in our regression models of the FRI treatment effect (Enders, 2010; Pan and Bai, 2015). Finally, it is important to note that, even with missingness, all of our analyses were more than adequately powered to detect small effects. An a priori power analysis indicated that the sample size required to detect a small effect (i.e., odds ratio = 1.50) of FRI on STEM degree and 6-year graduation was $N = 778$, while the sample size required to detect a small effect (i.e., $R^2 = 0.02$) on cumulative GPA was $N = 476$ (Faul *et al.*, 2007; Chen *et al.*, 2010).

RESULTS

Graduation with STEM Degree

We assessed students' attainment of a STEM degree based on descriptive statistics (Table 1) and bivariate correlations (Table 2) and found a raw difference favoring the FRI group. However, raw differences between FRI and non-FRI groups may be untrustworthy, as they do not control for chance imbalances on the matching covariates and they use data from unlinked members of matched pairs (e.g., one member of the matched pair graduated with a STEM degree [STEM degree = 1], but the other member of the matched pair had not yet graduated [STEM degree = missing]). Therefore, we conducted a logistic regression analysis on all matched pairs with graduation data (both pairs graduated; STEM degree: 0 = non-STEM college; 1 = STEM-related college) to determine the effect of FRI participation on students' probability of graduating with a STEM degree. We used a hierarchical approach in the logistic regression analysis (not

Table 1. Summary of descriptive statistics of outcomes and key predictors as a function of FRI status

Variable[a]	FRI				Matched control			
	N	%	M	SD	N	%	M	SD
STEM degree	1482	81			1377	68		
6-year graduation[b]	1082	79			1104	75		
Cumulative GPA	1482		3.39	0.45	1374		3.34	0.43
FRI course work	2449	100			2449	100		
Course 1 only[c]	416	17			0	0		
Courses 1 and 2[c]	882	36			0	0		
Courses 1, 2, and 3[c]	1151	47			0	0		
Midpoint GPA	2188		3.28	0.55	2086		3.20	0.56

STEM degree codes: 0 = non-STEM, 1 = STEM; 6-year graduation codes: 0 = did not graduate, 1 = graduated within 6 years. Course variables (e.g., Course 1 only) were dummy coded to indicate level of participation in FRI = 1 versus otherwise = 0 (reference group was the non-FRI matched control group).

[a]For dichotomous variables (e.g., STEM degree: 0 = non-STEM degree, 1 = STEM degree).

[b]Six-year graduation represents graduation rate of those who started at UT Austin on/before 2009.

[c]Sample size for FRI course work (e.g., FRI group $N = 2449$) broken down by subgroup (e.g., Course 1 only, $n = 416$).

to be confused with hierarchical linear models or multilevel models), such that matching variables were entered in step 1 and FRI variables were entered in step 2. This approach also allowed us to identify whether students experienced different outcomes as a result of participating in one, two, or all three FRI courses.

First, we regressed STEM graduation on all variables used to estimate propensity scores to control for chance imbalances on any of the matching covariates (step 1), followed by three dummy-coded variables indicating level participation in FRI (step 2; Course 1 only, Course 1 and 2, Course 1, 2, and 3 [reference category was the non-FRI group]). The results indicated that FRI membership has a statistically significant effect on the probability of graduating with a STEM degree over and above control variables (Table 3). Because our analysis focused on a set of three related outcomes, we adopted a Bonferroni-corrected alpha level ($\alpha = 0.05/3 = 0.017$) to control type I error rate inflation in assessing statistical significance.

Parameter estimates in the final step of the logistic regression model revealed that students who participated in all three semesters of FRI (Courses 1, 2, and 3) were significantly more likely to graduate with a STEM degree compared with the non-FRI control group (O.R.$_{Courses123}$ = 6.08, 98.3% CI [3.66, 10.12]; see Supplemental Table S3 for complete details).

To make these findings more concrete, we calculated the predicted probability of earning a STEM degree for students in the non-FRI control and FRI groups. After controlling for other factors in the model, non-FRI students had a 71% predicted probability of graduating with a STEM degree compared with 94% for FRI students who completed all three courses (47% of FRI students completed all three courses; Figure 1A). Students who only participated in Course 1 (17% of FRI students completed only Course 1) or Courses 1 and 2 (36% of FRI students completed Courses 1 and 2) were just as likely to graduate with a STEM degree as non-FRI students (O.R.$_{Course1}$ = 0.69; O.R.$_{Courses12}$ = 1.37).

Graduation within 6 years

Descriptive statistics and bivariate correlations indicated a slight raw difference in students' 6-year graduation rate favoring the FRI group (Tables 1 and 2). To assess an unbiased effect of FRI participation on students' probability of graduating within 6 years of entering college regardless of major, we conducted a logistic regression analysis on matched pairs in which both had the opportunity to graduate within 6 years: FRI students and matched controls who both enrolled at UT Austin on or before 2009. We used the same

Table 2. Summary of bivariate correlations among outcomes and key predictors

Variable	1	2	3	4	5	6	7
1 STEM degree	1	0.13**	0.08**	−0.09**	0.01	0.21**	0.10**
2 6-year graduation		1	0.09**	−0.10**	−0.01	0.14**	0.45**
3 Cumulative GPA			1	−0.09**	−0.01	0.12**	0.98**
4 Course 1 only				1	−0.14**	−0.17**	−0.11**
5 Courses 1 and 2					1	−0.26**	−0.01
6 Courses 1, 2, and 3						1	0.15**
7 Midpoint GPA							1

STEM degree codes: 0 = non-STEM, 1 = STEM; 6-year graduation codes: 0 = did not graduate, 1 = graduated within 6 years. Course 1 codes: 0 = matched control, 1 = FRI; Courses 2 and 3 codes: 0 = did not participate, 1 = participated.

**$p < 0.01$.

Table 3. Regression analysis (logistics or OLS) with graduation with a STEM degree, graduation within 6 years with any degree, and graduation cumulative GPA as outcomes

Step	Predicator	Graduation with STEM degree ($n = 1624$)			Graduation within 6 years ($n = 990$)			Graduation cumulative GPA ($n = 1510$)		
		$-2LL$	Pseudo-R^2	$\Delta\chi^2(df)$	$-2LL$	Pseudo-R^2	$\Delta\chi^2(df)$	R^2	ΔR^2	$\Delta F(df)$
1.	Controls[a]	1653.38	0.183	217.79 (24)***	1012.38	0.113	77.37 (21)***	0.216	0.216	17.00 (24)***
2.	FRI courses	1545.68	0.266	107.70 (3)***	990.18	0.143	22.20 (3)***	0.228	0.013	8.02 (3)***
3.	Midpoint GPA							0.951	0.723	21,705.91 (1)***

$-2LL = -2*$log likelihood; pseudo-R^2 = Nagelkerke R^2 estimate of effect size; $\Delta\chi^2$ = change in overall chi-square from the previous model.
[a]The list of control variables is described in the *Methods* section.
***$p \leq 0.001$.

hierarchical procedure described above, but with graduation within 6 years as the outcome (0 = did not graduate within 6 years; 1 = graduated within 6 years).
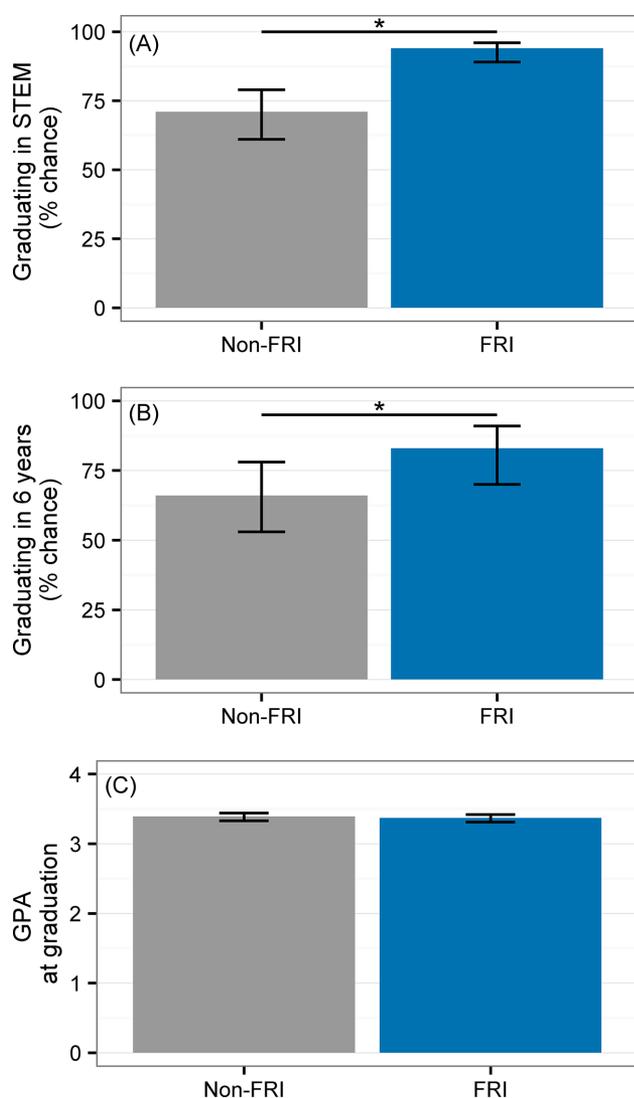

Figure 1. Participation in all three FRI courses significantly improves students' predicted probability of graduating with a STEM major (A) and graduating in 6 years (B), but does not affect students' probability of earning a higher cumulative GPA at graduation (C). Error bars represent 98.3% confidence intervals; $p < 0.017$.

The results indicated that completing the full FRI program has a statistically significant effect on students' probability of graduating within 6 years, over and above control variables (Table 3). Parameter estimates in the final step of the logistic regression model revealed that students who participated in all three semesters of FRI were significantly more likely to graduate within 6 years (O.R.$_{Courses123}$ = 2.43, 98.3% CI [1.34, 4.43]; Supplemental Table S4). To make these findings more concrete, we calculated the predicted probability of graduating within 6 years for students in the non-FRI control and FRI groups. After controlling for other factors in the model, non-FRI students had a 66% predicted probability of graduating with any degree within 6 years compared with 83% for FRI students (Figure 1B). FRI students who only participated in Course 1 or Courses 1 and 2 were just as likely to graduate within 6 years as non-FRI students (O.R.$_{Course1}$ = 0.63; O.R.$_{Courses12}$ = 1.07).

Cumulative GPA

Again, descriptive statistics and bivariate correlations indicated a slight raw difference in cumulative graduation GPA, favoring the FRI group (Tables 1 and 2). To assess an unbiased effect of FRI participation on educational performance at graduation, we conducted a regression analysis on all matched pairs with cumulative graduation GPA scores. Preliminary analysis indicated an FRI effect on midpoint GPA (Supplemental Table S5). Thus, midpoint GPA was entered in step 3 as potential mediator of the effect of participating in FRI. As above, the results indicated that FRI membership (step 2) had a statistically significant effect on cumulative GPA at graduation, over and above control variables (Table 3). FRI students who completed Courses 1 and 2 or all three courses exhibited statistically significantly higher graduation GPA compared with the non-FRI control group (step 2; $b_{Courses12}$ = 0.07 and $b_{Courses123}$ = 0.12), but students who completed only FRI Course 1 (step 2; $b_{Course1}$ = 0.01) were not significantly different from the non-FRI control group. We suspected that grades in FRI courses themselves could be influencing cumulative graduation GPA. Thus, we controlled for midpoint GPA and found that the positive effects of participating in FRI were nullified (Figure 1C and Supplemental Table S6).

Potential Race, Gender, and First-Generation Moderation Effects

We explored whether students from different backgrounds differed in their outcomes as a result of participating in FRI.

Specifically, we tested whether students' race/ethnicity, gender, or first-generation college status moderated the effect of FRI on the outcomes. Exploratory moderated regression analyses (logistic and OLS) indicated that students' sociodemographic characteristics did not moderate the effects of FRI on outcomes. Given the analytical sample sizes, number of predictors in our models, and the adjusted alpha level, our exploratory analyses were more than adequately powered to detect small moderating effects (i.e., O.R. = 1.50 or $R^2 = 0.02$; power > 0.99).

DISCUSSION

To the best of our knowledge, this is the largest and most carefully controlled analysis to date of the effects of participating in a CUREs on long-term student outcomes that are of high interest to students and institutions alike. Specifically, the data reported here indicate that participation in early CUREs significantly increases students' likelihood of graduating with a STEM degree and graduating within 6 years. After controlling for other variables, the outcomes of participating in the full FRI program were the same regardless of students' gender, race/ethnicity, and first-generation in college status, showing that these effects were robust for diverse students. Results from these analyses demonstrate the importance of using quasi-experimental techniques for controlling for selection bias in determining the effects of research experiences, since the data show that the variables that influenced entry into FRI had statistically significant effects on all of the outcomes we examined.

The effects of FRI differed depending on whether students completed Courses 1, 2, and 3, which could be due to the nature of the courses or to time spent in the program. In Course 1, students have total freedom to define their own investigations, from posing questions to investigate to designing studies to collecting and analyzing data to constructing and evaluating scientific arguments. Courses 2 and 3 are more similar to UREs, because students engage in conducting novel studies that build on and contribute to a faculty member's ongoing research, with the potential to yield publishable results as well as methods, data, and other products (e.g., inventions, companies) that are of interest to communities outside the classroom. Thus, the problem space has been defined to some extent. Students carve out their own aspect of the research to pursue and must collect and analyze data and construct and evaluate arguments but may not have complete latitude to select their research questions or methods. This study provides a preliminary test of whether having full intellectual responsibility posing research questions is important for students to achieve desired outcomes (National Academies of Sciences, Engineering, and Medicine, 2015). The parameter estimates from our regression models (Supplemental Tables S4–S6) indicated that Course 1 alone did not have a significant effect on any of the outcomes we examined, yet model fit was improved by including Course 1 in all three models (Table 3). These results suggest that investigatory courses like Course 1 may have distinct positive effects on graduating with a STEM degree when compared with research courses (i.e., Courses 2 and 3). Alternatively, it may be that the independent effects of each FRI course on students' probability of graduating with

a STEM degree can simply be attributed to longer exposure to a learning environment that is more motivating than traditional lab course experiences (Graham *et al.*, 2013).

The distinct, significant effects of Courses 2 and 3 on students' likelihood of graduating in 6 years and graduating with a STEM degree indicate that the duration of students' involvement in CUREs is important for their outcomes. Specifically, the data indicate that a one-semester research course is sufficient to achieve these outcomes to some extent but that participation in additional semesters is important for maximally realizing these outcomes. This finding adds to those from Shaffer and colleagues (2014), who found that students who spent more time on their CURE work reported increased learning and greater interest in STEM courses and in STEM in general. These results are likely to be conservative estimates of the effect of participating in CUREs, because the bivariate correlations show that participation in Courses 1, 2, and 3 are all fairly highly correlated. It is likely that collinearity between participating in each course suppresses the independent effects of each course. Larger samples of students who participate in Course 1 only or Courses 1 and 2 only are needed to confirm this.

These analyses were conducted with data from a CURE program that has involved enough students for a sufficient length of time to examine long-term outcomes such as graduation rates and majors. The extent to which these results will apply to other CUREs, especially CUREs that enroll students later in their undergraduate degrees, needs to be determined by conducting similar, carefully controlled studies. Given that many CUREs are small in scale or have more finite life spans, this may prove difficult. An alternative approach would be for studies of CUREs to report long-term outcomes of participating and nonparticipating students such that meta-analyses can be done in the future to identify effects across research course experiences.

These findings are arguably the most robust evidence to date that CUREs improve the outcomes of undergraduate STEM students. We have statistically controlled for background variables related to academic motivation and preparation (e.g., prior achievement, math and science preparation, parental education) and controlled for initial entry into FRI. This lends confidence that the outcomes reported here can be attributed to CURE participation. However, there are likely to be other variables not included in our analysis that may predict FRI participation and cause the outcomes of interest. We are currently collecting data on psychological variables that may predict students' participation in FRI and their persistence in college and in STEM (e.g., motivation, interest in research; Hernandez *et al.*, 2013) in order to more fully understand the effects of CURE participation per se.

These results do not yield insights into the features of CUREs that lead to these outcomes. There are many structural differences between FRI and traditional lab courses that could be leading to the outcomes reported here (Auchincloss *et al.*, 2014). For example, Courses 2 and 3 meet in dedicated lab spaces that become a sort of scientific home for students. Typically, two wet-lab FRI groups meet in a single large lab space, such that up to 80 students are cycling in and out of the space over the course of the week. Students working on computational projects meet in regularly scheduled conference-style classrooms or a robotics lab and also work online at a distance. The lab spaces are open to students and staffed

by REs, graduate or undergraduate teaching assistants, or peer mentors throughout the day. FRI lab spaces often become a place where students not only conduct research but also study for classes and spend time more informally. The involvement of undergraduate mentors gives students access to near peers who have recent experience learning the research and who can provide general advice on navigating the first 2 years of college. Class size is not likely to be a major factor, since enrollments are similar between FRI courses and standard laboratory courses, and most FRI courses enroll up to 35 students, which is larger than the typical 24-person introductory lab course. Different versions of FRI that make use of curricular and instructional staffing models are now being implemented at universities across the country. Cross-site study of student outcomes has the potential to yield insight into which FRI design elements are necessary and sufficient to achieve the results reported here.

Future research on CUREs should focus on using research and theory from social sciences, including situated learning (Brown *et al.*, 1989), communities of practice (Wenger, 1999; Lave and Wenger, 1991), and knowledge integration (Linn *et al.*, 2015), to understand the features of CURE design and implementation that lead to these long-term outcomes (Corwin *et al.*, 2015a). Recent research aimed at distinguishing CUREs from traditional lab courses indicates that the extent to which students have opportunities to make discoveries that are of broad interest, engage in iterative work (e.g., troubleshooting, revising based on feedback, building off one another's findings), and have opportunities to develop a sense of ownership of their research projects may be particularly important design features (Hanauer *et al.*, 2012; Hanauer and Dolan, 2014; Corwin *et al.*, 2015b). In addition, study of CUREs indicates that more proximal outcomes, including the development of scientific self-efficacy and scientific identity and internalization of scientific values, are important predictors of persistence in science research–related education and career paths (Estrada *et al.*, 2011; Hernandez *et al.*, 2013; Robnett *et al.*, 2015). CUREs should be examined for their potential to foster student growth in these domains, ideally using a model-based approach that links CURE design features to students' short- and long-term outcomes (Corwin *et al.*, 2015a). Future research on CUREs should also follow the advice of calls for the next generation of discipline-based education research, aimed at understanding not simply what works for students but for whom and in what contexts (Singer *et al.*, 2012; Freeman *et al.*, 2014; Dolan, 2015).

These results should be useful on a national level for tailoring allocation of funds to CUREs versus UREs according to the intended goals. CUREs, especially those offered as part of introductory course work, are likely to be a more fruitful investment when stakeholders are interested in increasing graduation rates and retention in STEM majors. Investment in research internships may be better suited to helping students confirm their career interests, explore graduate education, and further develop their scientific expertise. These results lay an important foundation for conducting cost–benefit analyses regarding the value of CUREs in terms of yielding additional tuition dollars and increasing the earning potential of STEM majors, especially for students from underrepresented or underserved backgrounds, for whom FRI was equally effective.

The effects of FRI on graduation rates and STEM retention have been and continue to be an important factor in driving institutional investment in the program. Currently, ~65% of the costs are borne jointly by the university instructional budget and college-level administrative funds, and 35% are covered by funds from grants, gifts, and endowment. Based on the results presented here, the CNS aspires for all first-year undergraduates in the college to participate if they are interested. About 200 students per year are on the waiting list, a number that has remained steady even as the program has grown. There is also a waiting list of faculty who would like to lead streams. The main limiting factors are space to accommodate the open lab structure of the program and funds to support the unique instructional staffing model, mainly the inclusion of the PhD-level RE and undergraduate peer mentors.

In his letter to the U.S. president, John P. Holdren noted,

> *Economic forecasts point to a need for producing, over the next decade, approximately 1 million more college graduates in STEM fields than expected under current assumptions. Fewer than 40% of students who enter college intending to major in a STEM field complete a STEM degree. Merely increasing the retention of STEM majors from 40% to 50% would generate three-quarters of the targeted 1 million additional STEM degrees over the next decade.* (PCAST, 2012)

FRI represents a scalable, affordable way to meet this demand. According to predicted probabilities in this study, out of every 100 students who enter college, 17 more will complete an undergraduate degree if they complete FRI. For every 100 students who graduate, 23 more will stay in a STEM major if they complete FRI. A rough estimate of the total per-student cost of FRI is ~\$500 for Course 1 and ~\$1000 each for Courses 2 and 3. Although this cost is higher than the typical ~\$500 per-student cost of a standard introductory lab course at UT Austin, the cost is low compared with the typical ~\$5000 per student for 8–10 wk Summer research internships and to the tuition dollars lost when students leave college. Costs could be lowered further by scaling up some of the cost-saving measures that we have implemented at UT Austin, such as offering peer mentors relevant course credit instead of pay, hiring senior undergraduates instead of graduate students as teaching assistants, or hiring graduate students as REs. Other models should also be tested, such as tenure-track or tenured faculty serving as the RE as part of their standard teaching responsibilities.

Given that FRI boosted retention among students regardless of their background, the diversity of students enrolled in the program provides the additional benefit of diversifying to the STEM workforce. In the long term, growing a more diverse STEM workforce has the potential to produce more creative, effective, and feasible ideas than would be accomplished by homogenous groups (McLeod *et al.*, 1996). In the near term, FRI can be a model for addressing the massive attrition of undergraduate students from STEM disciplines and ensuring that *all* students have the potential to earn higher wages and experience lower unemployment rates associated with STEM-related jobs (U.S. General Accounting Office, 2005; Langdon *et al.*, 2011; PCAST, 2012).

ACKNOWLEDGMENTS

REFERENCES

American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.

Auchincloss LC, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, Lawrie G, McLinn CM, Pelaez N, Rowland S, *et al.* (2014). Assessment of course-based undergraduate research experiences: a meeting report. CBE Life Sci Educ *13*, 29–40.

Bascom-Slack CA, Arnold AE, Strobel SA (2012). Student-directed discovery of the plant microbiome and its products. Science *338*, 485–486.

Beckham JT, Simmons S, Stovall GM, Farre J (2015). The freshman research initiative as a model for addressing shortages and disparities in STEM engagement. In: Directions for Mathematics Research Experience for Undergraduates, ed. MA Peterson and YA Rubinstein, Singapore: World Scientific, 181–212.

Brown JS, Collins A, Duguid P (1989). Situated cognition and the culture of learning. Educ Res *18*, 32–42.

Buonaccorsi VP, Boyle MD, Grove D, Praul C, Sakk E, Stuart A, Tobin T, Hosler J, Carney SL, Engle MJ, *et al.* (2011). GCAT-SEEKquence: Genome Consortium for Active Teaching of undergraduates through increased faculty access to next-generation sequencing data. CBE Life Sci Educ *10*, 342–345.

Buonaccorsi V, Peterson M, Lamendella G, Newman J, Trun N, Tobin T, Aguilar A, Hunt A, Praul C, Grove D, *et al.* (2014). Vision and change through the Genome Consortium for Active Teaching using next-generation sequencing (GCAT-SEEK). CBE Life Sci Educ *13*, 1–2.

Chen H, Cohen P, Chen S (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. Commun Stat Simul Comput *39*, 860–864.

Corwin LA, Graham MJ, Dolan EL (2015a). Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. CBE Life Sci Educ *14*, es1.

Corwin LA, Runyon C, Robinson A, Dolan EL (2015b). The laboratory course assessment survey: a tool to measure three dimensions of research-course design. CBE Life Sci Educ *14*, ar37.

Dolan EL (2015). Biology education research 2.0. CBE Life Sci Educ *14*, ed1.

Eagan MK, Hurtado S, Chang MJ, Garcia GA, Herrera FA, Garibay JC (2013). Making a difference in science education: the impact of undergraduate research programs. Am Educ Res J *50*, 683–713.

Enders CK (2010). Applied Missing Data Analysis, New York: Guilford.

Estrada M, Woodcock A, Hernandez PR, Wesley P (2011). Toward a model of social influence that explains minority student integration into the scientific community. J Educ Psychol *103*, 206–222.

Faul F, Erdfelder E, Lang A-G, Buchner A (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods *39*, 175–191.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014). Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci USA *111*, 8410–8415.

Graham MJ, Frederick J, Byars-Winston A, Hunter A-B, Handelsman J (2013). Increasing persistence of college students in STEM. Science *341*, 1455–1456.

Hanauer DI, Dolan EL (2014). The project ownership survey: measuring differences in scientific inquiry experiences. CBE Life Sci Educ *13*, 149–158.

Hanauer DI, Frederick J, Fotinakes B, Strobel SA (2012). Linguistic analysis of project ownership for undergraduate research experiences. CBE Life Sci Educ *11*, 378–385.

Harvey PA, Wall C, Luckey SW, Langer S, Leinwand LA (2014). The Python project: a unique model for extending research opportunities to undergraduate students. CBE Life Sci Educ *13*, 698–710.

Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, *et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genet *2*, e92.

Hernandez PR, Schultz PW, Estrada M, Woodcock A, Chance RC (2013). Sustaining optimal motivation: a longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. J Educ Psychol *105*, 89–107.

Ho DE, Imai K, King G, Stuart EA (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit Anal *15*, 199–236.

Ho DE, Imai K, King G, Stuart EA (2011). MatchIt: nonparametric preprocessing for parametric causal inference. J Stat Softw *42*.

Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, *et al.* (2014). A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. mBio *5*, e01051.

Kloser MJ, Brownell SE, Shavelson RJ, Fukami T (2013). Research and teaching. Effects of a research-based ecology lab course: a study of nonvolunteer achievement, self-confidence, and perception of lab course purpose. J Coll Sci Teach *42*, 72–81.

Langdon D, McKittrick G, Beede D, Khan B, Doms M (2011). STEM: Good Jobs Now and for the Future, ESA Issue Brief #03-11, Washington, DC: U.S. Department of Commerce.

Laursen S, Hunter A-B, Seymour E, Thiry H, Melton G (2010). Undergraduate Research in the Sciences: Engaging Students in Real Science, San Francisco, CA: Wiley.

Lave J, Wenger E (1991). Situated Learning: Legitimate Peripheral Participation, Cambridge, UK: Cambridge University Press.

Leung W, Shaffer CD, Cordonnier T, Wong J, Itano MS, Tempel EES, Kellmann E, Desruisseau DM, Cain C, Carrasquillo R, *et al.* (2010). Evolution of a distinct genomic domain in *Drosophila*: comparative analysis of the dot chromosome in *Drosophila melanogaster* and *Drosophila virilis*. Genetics *185*, 1519–1534.

Linn MC, Palmer E, Baranger A, Gerard E, Stone E (2015). Undergraduate research experiences: impacts and opportunities. Science *347*, 1261757.

Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung H-M, Du C, Eckdahl T, Goodman AL, Hauser C, Jones CJ, *et al.* (2008). Undergraduate research. Science *322*, 684–685.

Lopatto D, Tobias S (2010). Science in Solution: The Impact of Undergraduate Research on Student Learning, Washington, DC: Council on Undergraduate Research.

McLeod PL, Lobel SA, Cox TH (1996). Ethnic diversity and creativity in small groups. Small Group Res *27*, 248–264.

National Academies of Sciences, Engineering, and Medicine (2015). Integrating Discovery-Based Research into the Undergraduate Curriculum: Report of a Convocation, Washington, DC: National Academies Press.

Pan W, Bai H (2015). Propensity Score Analysis: Fundamentals and Developments, New York: Guilford.

President's Council of Advisors on Science and Technology (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics, Washington, DC: U.S. Government Office of Science and Technology.

Riegle-Crumb C, King B, Grodsky E, Muller C (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. Am Educ Res J *49*, 1048–1073.

Robnett RD, Chemers MM, Zurbriggen EL (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. J Res Sci Teach *52*, 847–867.

Sadler TD, Burgin S, McKinney L, Ponjuan L (2010). Learning science through research apprenticeships: a critical review of the literature. J Res Sci Teach *47*, 235–256.

Schafer JL, Kang J (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. Psychol Methods *13*, 279–313.

Schneider B, Swanson CB, Riegle-Crumb C (1997). Opportunities for learning: course sequences and positional advantages. Soc Psychol Educ *2*, 25–53.

Seymour E, Hunter A-B, Laursen SL, DeAntoni T (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. Sci Educ *88*, 493–534.

Shaffer CD, Alvarez CJ, Bednarski AE, Dunbar D, Goodman AL, Reinke C, Rosenwald AG, Wolyniak MJ, Bailey C, Barnard D, *et al.* (2014). A course-based research experience: how benefits change with increased investment in instructional time. CBE Life Sci Educ *13*, 111–130.

Singer SR, Nielsen NR, Schweingruber HA (2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering, Washington, DC: National Academies Press.

Thoemmes F (2011). An SPSS R menu for propensity score matching. https://sourceforge.net/projects/psmspss/files/?.

U.S. General Accounting Office (2005). Higher Education: Federal Science, Technology, Engineering, and Mathematics Programs and Related Trends, Washington, DC.

Wang X (2013). Why students choose STEM majors: motivation, high school learning, and postsecondary context of support. Am Educ Res J *50*, 1081–1121.

Wei CA, Woodin T (2011). Undergraduate research experiences in biology: alternatives to the apprenticeship model. CBE Life Sci Educ *10*, 123–131.

Wenger E (1999). Communities of Practice: Learning, Meaning, and Identity, Cambridge, UK: Cambridge University Press.

West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, Szapocznik J, Fishbein M, Rapkin B, Clatts M, *et al.* (2008). Alternatives to the randomized controlled trial. Am J Public Health *98*, 1359–1366.