Review

Evaluating Teen Pregnancy Prevention Programs: Decades of Evolving Strategies and Practices

Susan Philliber

Philliber Research and Evaluation, Accord, NY 12404, USA;
E-Mail: sphilliber@philliberresearch.com

Abstract: This paper reviews the changing strategies for both process and outcome evaluations of teen pregnancy prevention programs over the past few decades. Implementation evaluations have emphasized discovery of what program attributes are most effective in reducing teen pregnancy and its antecedents. Outcome evaluations have moved from collecting data to measure knowledge, attitudes, and program satisfaction to measuring behavior change including postponement of sexual involvement, increased used of contraception, or reduction in teen pregnancy. High quality randomized control trials or quasi-experimental designs are being increasingly emphasized, as are sophisticated analysis techniques using multi-variate analyses, controls for cluster sampling, and other strategies designed to build a more solid knowledge base about how to prevent early pregnancy.

Keywords: evaluation; teen pregnancy prevention; outcome evaluation; implementation evaluation

1. Introduction

Over the past four decades one of the likely contributing factors to reduced rates of teen pregnancy in the United States has been the search for and discovery of programs that are effective in preventing this behavior. More and more programs with at least one credible evaluation have been found to prevent teen pregnancy or its sexuality-related antecedents. There has also been a search for the characteristics of effective programs. Evaluators have tried to learn which programs work best for various populations, and have documented the magnitude of program effects on early pregnancy or its

antecedents. While the 1970's were characterized by attention to programs for pregnant and parenting teens, by the 1980's, the search for effective prevention programs was being fully pursued [1].

In more recent years, organizations began to produce lists of effective programs, using various criteria. In 2004, the National Campaign to Prevent Teen Pregnancy (the National Campaign) published a guide intended to provide education about these lists and about how programs were selected for inclusion [2]. The publication stressed a movement away from weak or non-empirical evaluation criteria and the adoption of more rigorous standards:

> Credible lists were not based on process evaluation data (that is, they do not simply assess client or staff satisfaction with the program, whether the program was delivered as planned or attendance patterns); intuition about program effects; faith in a particular approach or method; political or religious inclination; or rhetoric about what should or might work. Criteria for program selection should be based on the rigor of the evaluation design and methods, as well as the strength of the findings [2] (p. 3).

Their sentiments reflected a movement away from program satisfaction data and the reliance instead on high quality research.

One of the most dramatic developments in teen pregnancy prevention programming and evaluation was made possible because the 2010 federal fiscal year budget included $110 million for an evidence-based Teen Pregnancy Prevention Program requested by Health and Human Services (HHS). This program was to be implemented by a newly created Office of Adolescent Health (OAH) within the Office of the Assistant Secretary for Health, and was to coordinate its efforts with the Administration on Children and Families (ACF) and the Centers for Disease Control and Prevention (CDC) [3].

OAH was directed to spend $75 million to replicate teenage pregnancy prevention programs proven effective through rigorous evaluation (Tier 1). For the first time a governmental office was required to identify such programs, and thus, to develop standards by which to make such a judgment. OAH was also tasked with spending $25 million through research and demonstration grants to develop, replicate, refine and test innovative models for preventing teen pregnancy (Tier 2).

In that same year, OAH and its contractor defined the standards to be used for calling a program Tier 1 and identified 28 programs shown through rigorous evaluation to have an impact on important sexual behavioral outcomes, such as delaying sex, using contraception or preventing teen pregnancy. By February 2015, OAH listed 37 such programs on this list. These now appear in a searchable data base so that potential implementers can select a program that has been tested with the intended target group and is a program for which they have the capacity and resources [4].

The purpose of this paper is to review the evolution of the evaluation of teen pregnancy programs from the late 1980's to the present, examining both process and outcome evaluations. The OAH standards for effective evaluations of teen pregnancy prevention programs are reviewed, as are current remaining evaluation challenges including recruitment, data collection, the use of randomized designs, and loss to follow-up in longitudinal studies.

2. Early Evaluations

As in the evaluation of other programs, teen pregnancy prevention programs have most commonly measured changes in knowledge, changes in attitudes, or changes in behavior among young people. Examples of these variables are knowledge about sexuality or physiology, changes in attitudes toward contraceptive use, or changes in sexuality behaviors such as age at first intercourse or use of contraception.

In the earliest years of teen pregnancy program evaluation, it was common for evaluators to use data collection instruments that elicited adolescent perceptions of the program (an attitude). Examples of such items are:

- I like the program.
- I have learned a great deal that is helpful.
- I would recommend this program to my friends [5].

Perhaps positive answers to these attitudinal questions provide the basis for other positive outcomes among young people, but these are measures of program satisfaction or youth attitudes, rather than measures of youth outcomes. When outcome measures were used in early evaluations, changes in knowledge were more often measured than changes in behaviors. Data on program implementation were largely confined to student attendance information.

3. Advances in Documenting Program Implementation

Over the years, several evaluators have tried to delineate the attributes of effective programs. Describing the components of programs helps to define exactly what is being evaluated. Then, if these attributes are found to be related to positive program outcomes, evaluators have empirical evidence that these are important program characteristics and this information can then guide future program development.

Table 1 illustrates some general criteria thought to influence the effectiveness of prevention programs. This list, produced in 1989, comes from a review of four types of programs affecting youth: Child abuse and neglect, poor school performance and school failure, teenage pregnancy, and teenage substance abuse programs [1].

Table 1. Factors Influencing the Effectiveness of Prevention Programs.

(1)	The capacity to identify a population at risk for the problem to be prevented.
(2)	The ability to reach an at-risk population with the program.
(3)	The appropriateness of the timing of the preventive intervention.
(4)	The duration and intensity of the program.
(5)	How broadly or narrowly the program is focused.
(6)	Experiential learning techniques in educational programs.
(7)	Parental involvement in programs focusing on children or adolescents.
(8)	Skill and training level of prevention program staff.
(9)	Program structure and integration/collaboration with the other community services.
(10)	Simplicity/complexity of prevention messages.
(11)	Avoiding negative effects of prevention programs.

Each of the factors is general rather than particular to teen pregnancy prevention, having come from several youth development literatures. While some of these factors may seem like common sense, when this list appeared, even seemingly easy criteria to meet—like being able to reach an at-risk population—was challenging since research on who was getting pregnant was still sparse and data on the optimal timing, duration, and intensity of programs were all but absent.

In 1994, ETR Associates produced a "Consumer's Guide" to sexuality education curricula [6]. While this guide did not cover other kinds of teen pregnancy prevention programs, sexuality education was the most common intervention at the time. The authors chose curricula to review that were school-based, published since 1985, available for review, and focused on more than one sexuality issue (such as sexuality transmitted disease, sexual abuse or other issues). They used the guidelines of the Sex Information and Education Council of the United States (SIECUS) and writing from American School Health Education as their criteria for evaluation of curricula [7,8]. Moreover, they developed guidelines for four stages of youth development: ages 5–8, ages 9–12, ages 12–15, and ages 15–18. As shown in Table 2, the Guide examined content, philosophy, skill building strategies, and the teaching methods used by each program reviewed.

Table 2. Categories Rated for Sexuality Education Curricula by ETR in 1994 [6].

Content	
Puberty	Sexual expression
Body image	STD transmission
Gender roles	Abstinence
Reproduce anatomy/physiology	Pregnancy prevention
Conception and birth	STD prevention
Sexual identity and orientation	HIV prevention
Relationships	Sexual exploitation
Parenting	Reproductive Health
Philosophy	
Healthy sexuality	Abstinence
Responsibility for decisions	Using protection if sexually active
	Philosophy not clear
Skill-Building Strategies	
Personal values	Community resources
Self-awareness/self-esteem	General communication skills
Influences on decisions	Assertiveness skills
Consequences of decisions	Refusal skills
Peer norms	Conflict-management skills
Perceived pregnancy risk	Decision-making skills
Perceived STD/HIV risk	Planning/goal setting skills
Teaching Strategies	
Ground rules	Cooperative learning/small groups
Anonymous question box	Case studies/scenarios
Teacher lecture	Skills practice and rehearsal
Large-group discussion	Audiovisual materials
Student worksheets	Community speakers/involvement
Journals/story writing	Peer helper component
	Parent/guardian involvement

Also rated were the following:

- Comprehensiveness (breadth and depth)
- Content accuracy and currency
- Skills building variety (breadth and depth)
- Methods variety
- Developmental appropriateness
- Cultural sensitivity
- Ease of implementation
- Evaluation
- Appearance/production quality, and
- Overall quality.

Each of these criteria had sub-dimensions to be rated. For example, to get the highest rating for cultural sensitivity, the curriculum in question had to have no stereotypic references about gender, race/ethnicity, family types, sexual orientation, or age, and had to have a variety of social groups and lifestyles depicted, as well as taking into account the cultural and ethnic values, customs and practices of the community.

While each of the criteria was chosen from a review of available literature at the time, most of this literature was descriptive. These guidelines then, were not based on high quality studies showing that curricula meeting these criteria had better outcomes than curriculum absent these attributes. Still this was an attempt to get closer to understanding what program features were most likely to actually reduce teen pregnancy or its antecedents. Over time, program implementation studies then, have focused more specifically on the core components needed for teen pregnancy prevention and thus become more relevant to those who design or select such programs for use in their own communities.

In 1997 the National Campaign continued work intended to help would-be program implementers to choose programs that had promise to reduce teen pregnancy. They published a report entitled *No Easy Answers* [9], emphasizing high quality outcome evaluations using experimental or quasi-experimental designs. The report focused chiefly on high quality outcome studies but included some discussion of program content or delivery styles. Based on a descriptive review of these programs, the review concluded:

> both the studies of antecedents and the evaluations of programs suggest that there are no simple approaches that will markedly reduce adolescent pregnancy. Instead, if pregnancy prevention initiatives are to reduce pregnancy markedly, they must have multiple effective components that address both the more proximal sexual antecedents of adolescent sexual behavior as well as the more distal antecedents involving one or more aspects of poverty, lack of opportunity and family dysfunction, as well as social disorganization more generally [9] (p. 46).

By 2007, in the second update of this review (the first published in 2001), Kirby and the National Campaign published under the somewhat more encouraging title *Emerging Answers* and included an entire chapter on the characteristics of effective curriculum-based programs [10]. While the numbers of programs included in the review increased substantially from the first review in 1997, the characteristics of effective programs were offered for only the curriculum-based sex and STD/HIV

education programs—A group of eight programs with strong evidence of positive impact on sexual behavior, pregnancy, or STD rates. Kirby divided their desirable characteristics into three groups (see Table 3): The process for developing the curriculum, its contents, activities and teaching methodologies, and the process of implementation.

Table 3. Characteristics of Effective Curriculum-Based Programs, Kirby in 2007 [10].

Process of Developing the Curriculum	
• Involved multiple people with experience in theory, research, sex and STD/HIV education	• Activities consistent with community values and available resources
• Assessed relevant needs and assets of target group	• Program was pilot-tested
• Used a logic model specifying health goals, behavior affecting these, risk and protective factors and activities to change these	
Contents of curriculum and activities or teaching methodologies	
• Clear health goals	• Safe social environment for young people
• Focused narrowly on specific behaviors leading to these	• Multiple activities to change risk and protective factors
• Addressed sexual psychosocial risk and protective factors	• Instructionally sound teaching methods
	• Age, culturally, and sexual experience appropriate
	• Covered topics in logical sequence
Process of Implementing the Curriculum	
• At least minimal support from appropriate authorities such as school districts or health departments	• When necessary, activities to recruit and retain teens and overcome barriers to their participation
• Used educators with desired characteristics who were trained and provided monitoring, supervision and support	• Implemented virtually all activities with reasonable fidelity

Note that in this 2007 review, content and activities are grouped and are not specific. This framework calls for "clear health goals" for example, rather than specifying specific topics such as puberty or conception, as the ETR guide had suggested. The 2007 list of characteristics of effective curricula also included the processes by which the program was developed as an important factor in its likely success—a relatively new consideration in the literature. While there is yet much research to do, the search to define what elements create programs that successfully reduce teen pregnancy has become more and more empirically based and has focused on the reduction of early pregnancy and its antecedents more specifically.

4. A Demand for Fidelity

Perhaps most importantly, *Emerging Answers* 2007 specifically listed implementation fidelity as important. The field had begun to realize that development of effective teen pregnancy prevention programs would be for naught if these programs were not implemented as intended. In fact, in recognition that lack of fidelity to a well-researched program was common, the National Campaign, along with other agencies funded by the Centers for Disease Control and Prevention (CDC), began an

Initiative to learn about and try to lessen barriers to faithful implementation of programs. The National Campaign called its effort *Putting What Works to Work*. As part of this Initiative, a descriptive survey was completed with 614 program implementers, local and state teen pregnancy coalition members, funders, and state officials who funded teen pregnancy prevention programs, asking whether they were implementing programs found to be effective with fidelity and if they were not, why not [11].

When asked to cite specific barriers to program implementation, a large number of respondents referred to: (1) the political climate at the time advocating strongly for "abstinence only" programs; (2) a greater focus placed on other issues affecting youth including AIDS, education and/or poverty; and (3) a distrust of science-based findings. This survey was during a time of substantial conflict in the nation over whether young people should receive "comprehensive" sexuality education or "abstinence only" education. Some of those surveyed believed that at least some evaluations had political agendas and should be viewed with skepticism.

Still, a majority of all groups, except funders, felt that the rigorous evaluation and proven effectiveness of new approaches was very important in choosing a teen pregnancy prevention program. Interestingly, among funders only 45% believed rigorous evaluation was very important. No one believed that these factors were completely unimportant [11].

These projects on "implementation science" began to focus on how to get those adopting programs to use new and effective approaches to preventing teen pregnancy. The questions asked had some similarities with the literature on the adoption of innovations. The chief question was—"If we know what works, why don't we implement it?" The study described above revealed several barriers, including:

- Lack of resources to purchase or receive training in successful programs;
- Local barriers to full implementation such as a school board forbidding a field trip to a contraceptive clinic;
- Programs seen as out of date or inappropriate for a given population;
- Modification of programs to fit the time available, the setting, or the population of a given community, and a quite frequent reason,
- "I just wanted to make the program my own [11]."

Thus, programs that had been theory-based, pilot-tested, designed for given populations, and focused on known risk and protective factors relative to teen pregnancy were refashioned. Sometimes, program implementers would take a chapter of one curriculum, two chapters from another, and then add their own favorite activities—still calling their program by its original title. As might be imagined, this began to alarm program developers as these "edited" programs might have had only a slight resemblance to the original theory and content. Some evaluations did not document these program alterations and thus, many such changes are likely to have been undetected.

This led program owners and founders to begin to establish "certification standards" for those who wanted to use their programs [12]. Some developed "minimum standards" that had to be met to call a program by its original name. By 2010, when OAH issued its first round of Funding Opportunity Announcements to fund replications of programs previously found effective, these announcements included language about the monitoring and maintenance of fidelity [13]. OAH defined a set of measures that all grantees were required to collect and report: Participant attendance, sessions implemented, facilitator fidelity logs including information on activities implemented as intended or

whether adaptations were made and observations of at least 10% of the actual program sessions by independent, outside evaluators [14].

This new-found emphasis on faithful replication of a program also pressured program designers and owners to specify exactly what constituted fidelity to their programs. The field began to ask program developers to name "the essential elements" of their programs [15]. Could the program be offered in another language? Was every activity necessary? Could some chapters of the curriculum be skipped? Very few programs could provide evidence to answer these questions because they did not have multiple evaluations of their programs offered with and without certain of their components. Most programs on existing lists of effective programs had been evaluated only once, often with the program developers keeping a close eye on implementation fidelity.

Still the past decades have seen an increased consciousness about, and new strategies to implement, evidence-based programs while being true to their original philosophy, content, intensity, and delivery styles [16,17]. And, from the newest round of program replications funded by OAH, new evidence should emerge about how successful evidence based programs are when they are and are not implemented with fidelity.

5. Upgrading Outcome Studies

The early years of the struggle to reduce rates of teen pregnancy in the United States were marked by the recognition of the poor quality of available research and evaluation data and multiple recommendations for increasing both the quality and quantity of data on potential success strategies (e.g., [18]).

In a 1986 brief to Senator John Chafee, who requested available information on what strategies might be effective in reducing teen pregnancy, the U.S. General Accounting Office (GAO) wrote:

> "The information on the effectiveness of preventing pregnancy is limited. …School-based teenage health clinics that include family planning services are frequently associated with reduced teenage birthrates but have not provided conclusive evidence that the programs were responsible for such declines. The information on the effectiveness of comprehensive service programs is limited" [19] (pp. 20–21).

Three years later another summary of evaluations of teen pregnancy prevention programs lamented:

> "While there are excellent examples of prevention program effectiveness studies...the number of such studies is small. Such studies are expensive and difficult to carry out. Consequently the evaluation components of many prevention programs…have been weak or poorly designed. …Often the program outcome indicators measured…do not include the central problem the program is attempting to prevent (e.g., teenage pregnancy…)" [1] (p. 17).

This criticism is still somewhat true of teen pregnancy prevention studies. Of the 37 programs currently on the HHS list of effective programs assessed with high quality designs, only four have measured and found an actual difference in pregnancy or birth rates between their program and comparison or control groups. Other programs have made the list by finding outcomes such as "had sex in the past 3 months" or "reductions in number of sexual partners"—both related to teen pregnancy but not actually measures of this outcome.

Even as late as 1995, a Child Trends summary of recent research reported relative to the data on determinants of teenage contraception use:

> "…much of this work is of poor quality. …studies are often based on tiny and non-representative samples… Studies are often cross-sectional, when prospective analyses are needed to identify determinants of contraceptive use and non-use. …bivariate analyses are often presented, although multivariate controls are needed…" [20] (pp. 60–61).

At the conclusion of the *No Easy Answers* review in 1997, Kirby made similar observations about the extant evaluation research:

> "…studies conducted to date are simply too few to evaluate each of the different approaches, let alone the various combinations of approaches. …Far too often studies have not used experimental designs; have had sample sizes that were too small…have used exploratory analytic techniques instead of confirmatory techniques,…have failed to control for clustering of youth in schools or agencies…or have failed to report and publish negative results" [9] (p. 45).

In this report, Kirby also complained about the failure to replicate single evaluations of programs thus limiting knowledge about to whom and under what conditions these programs might or might not produce positive outcomes. In his 2007 review, Kirby cited many of the same problems with existing research on how to prevent teen pregnancy [10].

Still, there has been progress. As noted above, the most dramatic step to improve the search for programs that are effective in preventing teen pregnancy was taken by HHS through OAH. In 2010, they received over 1000 applications to either replicate the evaluations of existing evidence based programs or to test new and promising strategies to prevent teen pregnancy. They funded 75 organizations in 32 states for replication work [14]. These studies were intended to expand the populations on which such programs were tested and to see, particularly for some of the older ones, whether they still seemed effective. They also tested programs that had not been rigorously evaluated previously, but which had promising early results. Funding was provided to evaluate these programs using randomized control trials or quasi-experimental designs so as to increase the numbers and quality of teen pregnancy program evaluations.

OAH emphasized fidelity and put into place a variety of mechanisms to help promote strict delivery of the program as intended, including site visits, observations of program sessions, training for replicators, and adherence to a set of standards for the implementation sites.

OAH designed and required a set of performance measures both for programs and participants. The grantee-level measures included reporting of both informal and formal partners working with grantees, training provided to facilitators, dissemination of manuscripts or presentations, and program delivery measures such as number of participants reached, the dosage of the program they received and fidelity in delivery of the program [21].

In addition HHS designed and monitored—through its subcontractor Mathematica Policy Research—a set of evaluation and analysis standards designed to improve many of the past research and evaluation practices. These standards were developed to provide transparency about how

effectiveness of programs was being determined and to improve standards generally. A brief description of these standards appears in Table 4 [22].

Table 4. The Office of Adolescent Health Evaluation Research Standards.

Criteria Category	High Study Rating	Moderate Study Rating	Low Study Rating
Study design	Random or functionally random assignment	Quasi-experimental design with a comparison group; random assignment design with high attrition or reassignment	Does not meet criteria for high or moderate rating
Attrition	What Works Clearinghouse standards for overall and differential attrition	No requirement	Does not meet criteria for high or moderate rating
Baseline equivalence	Must control for statistically significant baseline differences	Must establish baseline equivalence of research groups and control for baseline outcome measures	Does not meet criteria for high or moderate rating
Reassignment	Analysis must be based on original assignment to research groups	No requirement	Does not meet criteria for high or moderate rating
Confounding factors	Must have at least two subjects or groups in each research group and no systematic differences in data collection methods	Must have at least two subjects or groups in each research group and no systematic differences in data collection methods	Does not meet criteria for high or moderate rating

The standards call for random assignment or at least quasi-experimental designs, low attrition from the sample at follow-up intervals as well as little differential between the follow-up rates for treatment and control groups, controls for baseline equivalence of samples, low rates of group reassignment, and similarity of data collection methods in both treatment and control groups. Thus, the designs receiving the highest ratings are randomized control trials meeting all of these standards, whereas those with moderate ratings are quasi-experimental designs or randomized control trials that do not meet these additional criteria. Those with moderate ratings also do not have to meet the standards for attrition or reassignment because their weaker designs already include group differences that might bias the impact estimates and they are thus, not eligible for the highest ratings.

In addition, to be selected for inclusion on the list of evidence-based programs, each program must have a behavioral impact on pregnancy, STIs, or sexual risk behaviors such as sexual activity, contraceptive use or number of sexual partners. Evaluations measuring only knowledge or attitude change are not included. Clearly these criteria for outcomes are vastly different from measuring whether young people liked the program.

As this work proceeded, a series of briefs was produced by OAH providing guidance on such topics as how to meet the highest research standards, how to avoid sample attrition, how to analyze data when attrition is present, and how to control for cluster sampling [23]. They discuss theory-driven interventions that focus on risk and protective factors associated with teen pregnancy and recommend sound analysis practices to declare that an intervention is effective. And these standards are likely to become even more sophisticated and demanding in the future.

6. Some Remaining Barriers to High Quality Teen Pregnancy Prevention Evaluation

Because such interventions have been most common, much of the writing cited above has focused primarily on curriculum-based approaches. But these are not the only kinds of teen pregnancy prevention interventions. There are parent-child communication programs, school-based clinics, comprehensive youth development programs, community based programs, use of mass media, and other approaches, each of which may or may not include a curriculum. The evaluations of these programs face some common obstacles, perhaps worth mentioning here as the work of our future.

Recruitment or Targeting—As noted by the very earliest attempts to delineate the characteristics of effective programs to prevent teen pregnancy, these programs must be able to reach the population at risk of early conceptions and births. Effective recruitment depends on identifying who these young people are, knowing where to find them, devising effective strategies to recruit them, and then engaging them in an effective intervention.

Decades of research have now made it clear that teen pregnancy is not equally common in all communities. Table 5 cites an abbreviated list of the factors among young people that are related to early pregnancy [10] (p. 52):

Table 5. Factors predictive of early pregnancy.

Communities (e.g., exposure to violence and substance abuse)	Families (e.g., single parent families, poor relationships with parents, parents who do not model responsible values about sex and contraception, low level of parent education)
Friends and peers (e.g, poor performance in school, drug use, permissive and unprotected sex)	Romantic partners (e.g., an older boyfriend)

In the real world of program implementation however, programs might miss the most "at-risk" youth if they work with the schools that are most cooperative or choose youth programs that are not attracting these young people. Even if these young people are at the sites chosen for programs, they may have after-school jobs to contribute to the family income, they may provide after school child care for their younger siblings, or have other interests and priorities that make them not only hard to recruit but hard to hold. Young people in "alternative schools" may be there only briefly or be sporadic in their attendance, resulting in their receiving low doses of the intended program.

Parent-child communication programs are particularly likely to have recruitment and engagement challenges. While a frequent mantra among some in the U.S. is that parents should be the first and most important sex educators of their children, these programs are often difficult to deliver, since parents may not want to attend multiple sessions at night or on the weekends or they cannot take part

due to child care or work demands. Thus, their evaluations may suffer from small sample sizes or sparse attendance.

Data Collection—Assuming that we reach the population most in need of some teen pregnancy prevention intervention, care must be taken in data collection. Many programs want to operate in schools, where there are assembled groups of youth. But sexuality education interventions often face resistance from principals, teachers, and superintendents who are fearful of parental backlash. Such programs often have to secure active parental consent for their children to participate, and certainly for their children to be part of an evaluation collecting data on sexual behaviors. The students too, need to have such sensitive data collected with assurances of confidentiality. And questions asked of these young people must be on their reading level and take account of their cultural backgrounds and language proficiency. Failure to collect data with protocols considering these challenges can lead to false or incomplete information.

Randomization—The evaluators of teen pregnancy programs are well aware that a randomized control trial is the "gold standard" or the most respected research design to show that a program was the likely cause of any differences between young people who received the program and young people who did not. Pre-post designs without such controls may show change but do not persuade us that the program is the likely cause of that change.

Randomization also enables control over factors unmeasured. For example, if a group of young people who all tried to get into a program are randomized so that some get the program and others serve as controls, we are at least comparing a group of young people who all had the motivation to join the program.

Consider the challenges, however. Random assignment is widely disliked by program staff since they have to deny services to some young people, while appearing to favor others. Because they are often acquainted with the students being randomized, it is usually best to let an outside evaluator carry out this assignment so that the choice is actually random and not personal. Another strategy is to randomly assign units such as schools, school classes, or youth recreation centers to either receive or not receive the program. All of these randomization strategies tax resources since students receiving little or no intervention must be followed for data collection over time.

Loss to Follow-Up—In any study of teen pregnancy prevention, following young people for longer periods of time allows measurement of how long any discernible program effects might last or measurement of how long it takes for program effects to appear. But the most at risk students are often mobile, particularly in poorer, high risk neighborhoods. When a study begins with 100% of those assigned to the program and control groups or with only 90% to 80% of the intended sample since all of the parents did not consent to their children's participation, and then over subsequent years more and more of them are lost, the study loses its quality (see OAH standards above). Particularly if the loss to follow-up is higher in the program or control groups or if the loss is particularly common among one type of student—say the boys, for example—what began as two well-matched and comparable groups can degenerate into unmatched, small, and thus, non-comparable samples.

The evaluation community is beginning to develop new techniques for stemming this loss. Use of multiple data collection strategies (in-person, telephone, on-line, or in-home surveys), multiple attempts to contact students to obtain data, use of incentives for teachers or youth workers and students, all enhance sample retention.

High mobility in schools where programs are being offered can also complicate the evaluations of school based clinic programs or school based curricula. In some schools serving the students at highest risk for teenage pregnancy, the mobility rate of these students during a given year can be 40% or more. Thus, those who received the intervention are gone by the time of the post-survey and new students may not have yet had any intervention or services from these programs. Schools rarely have a foolproof mechanism for collecting accurate pregnancy data both because of student turnover and because pregnant students often drop out of school without disclosing their reasons for doing so. These are some of the evaluation struggles that yet plague programs, on and off school grounds.

7. Conclusions

Evaluation of teen pregnancy prevention programs has come a long way in the past few decades. This is challenging research in the real world, not manipulations in a bell jar or vacuum. It requires care for human subjects (most often the youth to whom we are offering programs) and because of the importance of the results, it demands our best strategies and most rigorous methods. Teen pregnancy prevention programs are often offered to the students most at risk in a community—the poorest, often discriminated against young people—who do not need to be wasting time in a program simply because we "think" it works or even because they enjoy it.

We are now way past "testimonial" evaluation, in both the program and funding communities. Those who pay for such programs are likely to ask us about our analysis strategies, our loss to follow-up, and the fidelity with which the program was implemented. But in a discipline where a randomized control trial (RCT) was an infrequent event and where evaluators have often complained about the scarcity and poor quality of the available evaluations, things are improving. This may mean that some programs that were selected as having strong evaluations with a positive behavioral impact on a teen pregnancy related behavior, will come off the lists of evidence based programs if further replications show them not to be replicable or to no longer be effective with the young people who are currently most at risk of teen pregnancy.

Going forward there is yet much to do. Our emphasis on high quality research designs that meet sound scientific standards should continue. While the past decades have seen authors try to create lists of the necessary elements for programs to successfully prevent teen pregnancy, we still have few actual tests of these hypotheses. More should be invested in using research to discover the particular strategies and content that will make these programs successful. And we yet know very little about which programs are best for which young people. It is unlikely that the same program will be equally resonant with all ethnic and cultural groups, all ages of youth, young people in foster care, lesbian, gay and transgendered youth, and other subgroups.

Over the past several decades studies of programs to prevent teen pregnancy and its antecedents have improved in their strategies, sampling, statistical techniques, and thus, the reliability of their conclusions. These improvements have not come cheaply—they will continue to tax our resources, energy, and commitment. Given the dire consequences of early pregnancies for our nation's youth, the quest to find what will prevent this event seems a worthy one in which to make future investments.

Conflicts of Interest

The author declares no conflicts of interest.

References and Notes

1. Card, J.J. (Ed.) *Evaluating Programs Aimed at Preventing Teenage Pregnancies*; Sociometrics Corp.: Palo Alto, CA, USA, 1989.
2. Solomon, J.; Card, J.J. *Making the List: Understanding, Selecting and Replicating Effective Teen Pregnancy Prevention Programs*; The National Campaign to Prevent Teen Pregnancy: Washington, DC, USA, 2004.
3. Kappeler, E.M.; Farb, A.F. Historical context for the creation of the Office of Adolescent health and the Teen Pregnancy Prevention Program. *J. Adolesc. Health* **2014**, *54*, S3–S9.
4. Searchable Data Base. Available online: www.hhs.gov/ash/oah/oah initiatives/teen_pregnancy/db/tpp-searchable.html (accessed on 21 June 2015).
5. The Center for Dropout Prevention at the University of Miami. *What Works and Why: A Guide to Evaluating Teen Pregnancy and Parenting Programs*; Center for Dropout Prevention at University of Miami: Coral Gables, FL, USA, 1988.
6. Ogletree, R.J.; Fetro, J.V.; Drolet, J.C.; Rienzo, B.A. *Sexuality Education Curricula: The Consumer's Guide*; ETR Associates: Santa Cruz, CA, USA, 1994.
7. National Guidelines Task Force. Guidelines for Comprehensive Sexuality Education: Kindergarten-12th Grade. In Proceedings of the Sex Information and Education Council of the United States, New York, NY, USA, 5 April 1991.
8. Neutens, J.J.; Drolet, J.C.; Dushaw, M.L.; Jubb, W. *Sexuality Education with Comprehensive School Health Education*; American School Health Education: Kent, OH, USA, 1991.
9. Kirby, D. *No Easy Answers: Research Findings on Programs to Reduce Teen Pregnancy*; The National Campaign to Prevent Teen Pregnancy: Washington, DC, USA, 1997.
10. Kirby, D. *Emerging Answers, 2007: Research Findings on Programs to Reduce Teen Pregnancy and Sexually Transmitted Diseases*; The National Campaign to Prevent Teen Pregnancy: Washington, DC, USA, 2007.
11. Philliber Research Associates. *Putting What Works to Work: Overall Project Evaluation*; Unpublished Report; Philliber Research Associates: Accord, NY, USA, 2005.
12. The Teen Outreach Program, Owned by Wyman Center in St. Louis is an example of such a practice. They have created certified replication partners, committed to delivery of TOP with fidelity. Available online: www.wymancenter.org/nationalnetwork (accessed on 20 June 2015).
13. Office of Adolescent Health, Teen Pregnancy Initiatives, Evaluation. Available online: www.hhs.gov/ash/oah/oah-initiatives/tpp (accessed on 2 September 2015).
14. Margolis, A.L.; Roper, A.Y. Practical experience from the Office of Adolescent Health's large scale implementation of an evidence-based teen pregnancy prevention program. *J. Adolesc. Health* **2014**, *54*, S10–S14.
15. Summerville, G. *Copy That: Guidelines for Replicating Programs to Prevent Teen Pregnancy*; The National Campaign to Prevent Teen Pregnancy: Washington, DC, USA, 2006.

16. Stid, D.; Neuhoff, A.; Burkhauser, L.; Seeman, B. *What Does It Take to Implement Evidence-Based Practices? A Teen Pregnancy Prevention Program Shows the Way*; The Bridgespan Group: Boston, CA, USA, 2013.

17. Kerschner, S.; Flynn, S.; Prince, M.; Potter, S.; Craft, L.; Alton, F. Using Data to Improve Fidelity When Implementing Evidence-Based Programs. *J. Adolesc. Health* 2014, *54*, S29–S36.

18. Hayes, C. *Risking the Future: Adolescent Sexuality Pregnancy and Childbearing*; National Academy Press: Washington, DC, USA, 1987.

19. United States General Accounting Office. *Teenage Pregnancy: 500,000 Births a Year but Few Tested Programs: Briefing Report to the Honorable John H. Chafee, United States Senate*; United States General Accounting Office: Washington, DC, USA, 1986.

20. Moore, K.A.; Miller, B.C.; Glei, D.; Morrison, D.R. *Adolescent Sex, Contraception and Childbearing: A Review of Recent Research*; Child Trends: Washington, DC, USA, 1995.

21. Farb, A.F.; Burrus, B.; Wallace, I.F.; Wilson, E.K.; Peele, J.E. From Mission to Measures: Performance Measure Development for a Teen Pregnancy Prevention Program. *J. Adolesc. Health* **2014**, *54*, S15–S20.

22. Mathematica Policy Research and Child Trends. In *Identifying Programs that Impact Teen Pregnancy, Sexually Transmitted Infections and Associated Sexual Risk Behaviors: Review Protocol*; Available online: http://www.hhs.gov/ash/oah/oah-initiatives/teen_pregnancy/db/eb-programs-review-v2.pdf (accessed on 2 September 2015).

23. Briefs. Available online: www.hhs.gov/ash/oah/oah-initiatives/for-grantees/evaluation/ta.html (accessed on 2 September 2015).