

Crime & Delinquency

<http://cad.sagepub.com/>

A Squandered Opportunity? A Review of SAMHSA'S National Registry of Evidence-Based Programs and Practices for Offenders

Benjamin J. Wright, Sheldon X. Zhang and David Farabee

Crime & Delinquency published online 26 August 2010

DOI: 10.1177/0011128710376302

The online version of this article can be found at:

<http://cad.sagepub.com/content/early/2010/08/09/0011128710376302>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Crime & Delinquency* can be found at:

Email Alerts: <http://cad.sagepub.com/cgi/alerts>

Subscriptions: <http://cad.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Aug 26, 2010

[What is This?](#)

A Squandered Opportunity?: A Review of SAMHSA'S National Registry of Evidence-Based Programs and Practices for Offenders

Crime & Delinquency
XX(X) 1-19
© 2010 SAGE Publications
DOI: 10.1177/0011128710376302
<http://cad.sagepub.com>


Benjamin J. Wright,¹ Sheldon X. Zhang,¹
and David Farabee²

Abstract

In the past decade, the push for evidence-based programs has taken on unprecedented prominence in the fields of substance abuse and correctional treatment as a key determinant for intervention funding. The National Registry of Evidence-based Programs and Practices (NREPP), managed and funded by the Substance Abuse and Mental Health Services Administration, was established in 1997 to aid community agencies in adopting intervention models for their particular clientele. Although well intentioned, the NREPP has also created opportunities that invite conflicts of interests and promulgate programs with questionable efficacy. After an exhaustive review of the literature that purports to have provided the “empirical evidence” for the NREPP registered programs, the authors found numerous irregularities in the studies with findings often based on small sample sizes. A more troubling finding is that much of the supporting literature is produced by

¹San Diego State University, San Diego, CA, USA

²University of California, Los Angeles, Los Angeles, CA, USA

Corresponding Author:

Benjamin Wright, Department of Sociology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4423, USA

Email: bjwright@kent.edu

the program developers themselves. There is a general lack of independent verification of the claimed treatment effects. If the NREPP is to fulfill its intended function, a tighter vetting process is needed for programs to be registered so that community agencies and treatment practitioners can consult with confidence.

Keywords

evidence-based programs, parolee reentry, program evaluation, NREPP, community corrections

Background

In 1974, Robert Martinson published a now very well-cited article that discussed the effectiveness of 231 correctional programs that had been evaluated over the previous 30 years (Martinson, 1974). Unsurprisingly, this article caused quite a stir among both fellow researchers and the media, leading many to conclude—either with disappointment and shock or unflinching certainty—that when it comes to rehabilitating offenders, “nothing works.” But, as has been acknowledged by several subsequent researchers, including Martinson himself, the problem with evaluations of correctional programming may not necessarily be that nothing works but that interventions are not being effectively implemented and/or evaluations are not being conducted in a methodologically sound and rigorous manner (Cullen & Gendreau, 2000; Farabee, 2002; Martinson, 1974). To this regard, it must be asked what the research community expects in terms of quality of research and who should get to produce the evidence in support of a given program. On one hand, the scope of the present study is fairly narrow, assessing the process by which evaluations of programs geared toward or applicable for criminal justice populations are reviewed for qualification on the Substance Abuse and Mental Health Services Administration’s (SAMHSA’s) National Registry of Evidence-based Programs and Practices (NREPP). On the other hand, this article deals with a much broader issue, albeit within a specific context: assessing the issue of how “evidence” is conceptualized.

The greatest contribution made by Martinson was not his pessimistic assessment of correctional programming but his attempt to establish higher benchmarks for what qualifies as sound, methodologically rigorous research. In their critique of the work by Martinson and his colleagues, Cullen and Gendreau (2000) note that Martinson acknowledged two alternative explanations for the nothing works conclusion other than the simple explanation that offenders are incapable of being rehabilitated. If it was not the case that the

programs Martinson and his colleagues analyzed were not working, it could have potentially been the case, they reasoned, that the programs were not being properly implemented and/or the evaluations of these programs were not being conducted in methodologically rigorous and effective ways to demonstrate treatment effects. The call, therefore, has not been simply for research that produces “evidence” for the sake of evidence but for sound, quality research capable of determining whether or not certain correctional programs are, in fact, living up to their expectations.

Farabee (2002), in a critique of Martinson’s 1974 article and of the research carried out by Lipton, Martinson, and Wilks (1975), points out that Martinson focused on two factors in these evaluations: “(a) weaknesses in program implementation and (b) bias in how these programs were evaluated,” focusing largely on methodological flaws with the evaluations themselves (Farabee, 2002, p. 190). Farabee (2002) adds:

A dangerous ecology can develop between treatment providers who may, in fact, be offering ineffective services and evaluators who set out to “prove” their effectiveness. Researchers that fail to find positive results will have difficulty publishing their findings. This can lead to an unspoken—even an unconscious—collusion between provider and researcher that undermines the independence of the evaluation. (pp. 190-191)

He adds that by 1979 Martinson had essentially retracted his previous, pessimistic position, not based on “new and more compelling evidence” but rather on inclusion of studies that were excluded from the original analysis due to methodological concerns, suggesting that Martinson had perhaps lowered his earlier standards for what evaluation research should entail (2002). Farabee (2002) concludes that while Martinson may have focused too extensively on the shortcomings of treatment efforts, his critics may be just as guilty of exaggerating program effectiveness.

Whether the findings of Martinson and his colleagues were overly harsh and pessimistic, or whether some of his critics tended to employ certain methodological techniques to dramatize their findings, the need for quality, rigorous evaluations was widely acknowledged. But a clear definition of what this entailed was much less clear. Emerging from this debate was an ever-growing concern centered on the need for evidence-based programming to ensure that programs that do not actually work do not continue to give false hopes and waste resources.

Over the years the debate intensified, and largely in response to concerns that programs may not be working effectively for the populations they serve, the phrase “evidence-based” began to gain much prominence in correctional literature as the key element in assessing, describing, or advocating intervention programs—especially since the late-1990s—so much so that the terminology has now become a standard insert in program development documents. Program administrators rarely engage in any practices these days without attributing them evidence-based status. But the problems of what qualifies as “evidence” and who gets to produce it were never effectively resolved, despite some great strides made within the research community.

For our purposes, there are at least two ways to conceptualize evidence. One may refer to the mere fact that an evaluation has been conducted showing some outcomes (negative or positive). In this loose definition, practically anything can qualify as evidence as long as the data are not fabricated. Martinson, before retracting his earlier position, seemed to hold a rather high standard for what qualified as evidence in support of a program, perhaps aligning more with an understanding that we call a “layman’s trust” in the integrity of scientific work. Under this conceptualization, any claimed treatment efficacy is produced through rigorous procedures and repeated tests, and those in charge of gathering the evidence do not stand to benefit financially from their own work (or at least declare their intent to profit). This is a matter of criteria-setting akin to the one commonly practiced in the field of medical research.

The emphasis on evidence-based practices and programs in corrections reflects years of concerted efforts by researchers and policy makers—including Martinson as well as his colleagues and critics—who call for verifiable and measurable outcomes as justifications for implementing and expanding correctional programs rather than relying on personal experience, political convictions, or common sense as grounds for new interventions (Cullen & Sundt, 2003; Gendreau, Goggin, Cullen, & Paparozzi, 2002). Many correctional programs at one time or another received much fanfare but were later found ineffective. Referred to as “correctional quackery” by Latessa, Cullen, and Gendreau (2002), some of these programs had gained much publicity in their heydays, such as boot camps and Scared Straight programs. Cullen and Sundt (2003) lamented that for too long criminal justice policy makers have been implementing programs with little regard to empirical evidence.

Although the virtues of evidence-based programs and the use of rigorous study designs in their evaluation have long been advocated by the research community, the gap between scientific evidence and its acceptance by practitioners remains vast. Sherman et al. (1997) suggested that with available

research of a reputable quality, programs could be evaluated to better determine “what works, what doesn’t and what’s promising,” a point repeatedly emphasized by the research community (Cullen & Gendreau, 2000; MacKenzie, 2000; Sherman et al., 1997).

Despite the development of the Maryland Scale for ranking the methodological rigor of studies by Sherman et al. (1997)—which certainly provided a clearer picture of what methodologically sound research should entail—and subsequent contributions by other researchers, the lack of a clear definition of what “evidence-based” implies and the pervasive reluctance among correctional and justice agencies to embrace randomized clinical trials remain major barriers in the move toward quality evidence-based studies of correctional programs (Sherman, 2000; Weisburd, 2000, 2003). The distance between scientific evidence and its acceptance by practitioners remains vast and bridging this gap is no easy task, particularly among the rank-and-file practitioners. Sherman (2009) suggests a top-down approach, counting on enlightened high-level government administrators or law makers to recognize the importance of randomized clinical trials. Such changes appear to be happening as a growing number of public and private funding agencies make clear their intent to support programs that can demonstrate empirical evidence of effectiveness or that are evidence-based. For instance, it has become a common practice that federal agencies such as the National Institute of Health, National Institute of Justice, and the Office of Management and Budget require that intervention studies employ either randomized assignment or quasi-experimental design with sophisticated procedures to control for subject variations. The need for demonstrated outcomes as a prerequisite for funding is now becoming the norm (Chemers & Reed, 2005).

The main trouble, however, which has been repeatedly emphasized up to this point, and that is the central concern of this article, is that the evidence-based concept can be quite slippery. Empirical evidence comes from multiple sources, depending on who is in charge of producing it. Gorman (2005) notes that when program developers evaluate their own interventions, there is a clear desire to demonstrate treatment effects or to downplay findings that fail to support treatment efficacy. Lack of a standard threshold for what qualifies as quality research is problematic and may lead some to conclude that, based on the quality of evaluations, nothing works, whereas it may lead others to dramatize their findings to increase the chance of publishing or give the impression of bias.

What we are concerned with herein is in reviewing the mechanisms by which the SAMHSA, an influential branch of the Department of Health and Human Services, reviews program evaluations that are listed on its NREPP,

and more so with the quality of those evaluations. In particular, we are concerned with programs that are intended for or could be disseminated for use among criminal justice populations. Herein, we are concerned principally with the question of: What qualifies as “evidence” and who gets to produce it?

SAMHSA and the NREPP

Traditionally, treatment providers interested in adopting evidence-based treatment strategies have relied either on time-consuming literature reviews or expert opinions. Recognizing the need for a more efficient and expedient way to access and/or adopt effective treatment interventions, the SAMHSA launched the NREPP in 1997 “to assist the public in identifying approaches to preventing and treating mental and/or substance use disorders that have been scientifically tested and that can be readily disseminated to the field” (NREPP, 2009). In creating the web-based registry, NREPP (2009) wanted to reduce the “time lag between the creation of scientific knowledge and its practical application in the field.” Since its inauguration, NREPP has undergone a number of modifications, including the adoption of the current searchable registry in 2007.

The NREPP was adopted in 1997 as part of the Model Programs Initiative of SAMHSA’s Center for Substance Abuse Prevention, and the current registry of “model,” “effective,” or “promising” programs is the result of more than 1,100 program reviews, indicating that just more than 10% of programs reviewed actually make it into the registry (NREPP, 2009). NREPP requires that programs submitted for review have at least one study conducted using an experimental or quasi-experimental design, indicate at least one positive outcome ($p < .05$), that the findings have been published in a “peer-reviewed journal or other publication or documented in a comprehensive evaluation report,” and that the program is ready to be disseminated for public use. Yet our analysis suggests that many programs are riddled with (real or alleged) conflict of interest and oftentimes inconsistencies in terms of quality ratings and readiness for dissemination ratings. The NREPP can certainly be a useful tool in assisting various governmental and community agencies in adopting programs, particularly those dealing with substance abuse and mental health, but it is imperative that it ensures that only quality programs make it into this registry.

As of February 2009 (the end of data collection for this study), the registry listed 119 intervention programs that reportedly have demonstrated efficacy and have been reviewed by independent researchers. Although SAMHSA favors no programs in particular, its willingness to host the information online

Table 1. Types of Interventions Included in This Study

Program Type	Target Population	Frequency	Percentage
Substance Abuse	Adults with or at risk for substance abuse disorders	17	55
Mental Health	Adults with mental health disorders	8	26
Hybrid	Adults with any combination of the above	6	19
Total		31	100

Note. The category “Hybrid” includes programs that provide a combination of mental health and substance abuse services. All percentages are rounded.

appears to have transformed the registry into an implicit stamp of approval, an unprecedented move in a field lacking a formal FDA-type review mechanism. Websites for many of these registered programs prominently display the NREPP logo, insinuating proof through official recognition of their evidence-based status.¹ It was not clear from our review exactly how many programs are reviewed annually, but according to the NREPP website, between 40 and 50 new programs are added to the registry each year. The volume of submissions for NREPP reviews soon seemed to overwhelm reviewers and the staff at SAMHSA, and at the time of this study, it was announced that the agency would no longer accept new submissions until October 1, 2009.

Scope of Review

In this article, we examined these NREPP registered programs and evaluated their supporting studies in terms of research design and readiness for dissemination and explored the relationship between the identified authors and the program development. This review focused on programs applicable to criminal justice populations. Of the 119 registered interventions on the NREPP, we found a total of 31 programs that were either targeting specifically or were considered suitable for criminal justice populations. Table 1 provides a breakdown of these programs based on their treatment populations. The NREPP classifies programs in three broad categories: those dealing with mental health, those dealing with substance abuse, and those dealing with co-occurring disorders, which we classify as “hybrid programs” in the present review.

Because the current NREPP does not offer a repository where empirical papers on the registered programs can be retrieved and reviewed, we conducted

an exhaustive search to locate all studies submitted as “empirical evidence” in support of these 31 programs. Data sources employed in this review include university libraries and online search engines, such as Academic Search Premier, EBSCOhost, Proquest, Elsevier, Springer Link, and Google Scholar. With the exception of eight studies (6.5%), we were able to locate full text files for 115 of the 123 submitted studies. Of those papers that we were unable to locate, two were unpublished papers, one of which was presented at a professional conference, one was an unpublished master’s thesis, and one was a manuscript that was still in preparation at the time the program was listed on NREPP. It appeared rather unusual that program developers would submit unpublished manuscripts and conference presentations as “empirical evidence,” giving the impression that just about any written material could be put forth for NREPP registration purposes.

Description of Interventions

Substance abuse programs. Seventeen of the 31 programs we reviewed were specific to substance abuse. The substance abuse programs on the NREPP included a wide range of treatment strategies dealing with a variety of drugs, including alcohol, marijuana, cocaine, and methamphetamine. Treatment strategies included self-help, professional help, and computer-based treatments.

Mental health programs. Eight of the programs included in our review were specific to mental health. As with the substance abuse treatment programs, discussed above, there were also a wide range of programs included in the assessment of mental health programs. However, most interventions were intended for those with schizophrenia, depression, anxiety, or trauma-related problems.

Hybrid programs. Six programs were classified as hybrid programs because they were comprised of a combination of treatment strategies that did not neatly fit into either of the other two categories and their target client populations were not specific to any particular disorder. These were mostly behavior modification programs designed to alter the cognitive-behavioral processes of participants with various disorders. The intended outcomes of these programs range from increasing moral judgment to improving stress reduction techniques.

Many of the NREPP registered programs employ well-known interventions such as Motivational Interviewing or Moral Enhancement Therapy. In this review, we avoid discussing any specific programs for anonymity reasons. Only broad categories of treatment orientations—such as those listed

above—are used. Our intent is not to draw attention to or single out any specific program, but to critique the qualifying process for programs to be listed on a national registry. Our goal is to present our assessment of the current NREPP registration practices and suggest how it may be improved, to reduce the possibility that programs with questionable methodologies or potential conflict of interest may find their way into the registry.

Findings and Analysis

Quality of Research as Rated by SAMHSA

The NREPP had a brief description for each of the registered programs and posted a few studies for review that are representative of selected outcomes. In addition, the registry provided information about study designs, program settings, and program replication. The NREPP employed a two-prong scheme to rate the registered program: (a) quality of research and (b) readiness for dissemination. For quality of research, NREPP quantified its assessment on the strength of evidence with a scale that ranges from 0 to 4, with 4 being the strongest. Quality of research was evaluated along the following components: reliability and validity of measures, intervention fidelity, missing data and attrition, confounding variables, and analytical strategies. Each of these components was rated as follows: “0” indicated a lack of evidence, inappropriate analysis of data, or other methodological flaws that made the reported outcomes questionable; “2” indicated that adequate controls were used despite potentially significant methodological problems; and a score of “4” indicated a strong study design or that measurement errors, attrition, and confounding variables were appropriately addressed.

Similarly, readiness for dissemination was rated on a 5-point scale (0-4), with 4 being the strongest. The scale factored in the program’s implementation materials, training and support, and quality assurance. A score of “0” indicated a lack of program materials, support, and resources. A score of “1” indicated that there was limited program literature or support, and the available materials did not directly assist in implementation, training, coaching, and/or quality assurance. A score of “2” indicated that there were limited materials, but they may be used to directly assist in the implementation of the described intervention. A score of “3” indicated that there were “adequate” materials, resources, and other tools to assist in the direct implementation, training, coaching, and/or quality assurance of the particular intervention. A “4” indicated the program implementation materials were “adequate” and of “high quality and appropriate for the intended audience(s)” (NREPP, 2009).²

Table 2. Range of NREPP Quality Ratings and Dissemination Readiness Scores

	NREPP Rating Scale (0-4)						
NREPP rating	<1.0	1.0-1.4	1.5-1.9	2.0-2.4	2.5-2.9	3.0-3.4	3.5-4.0
Quality (frequency)	<i>n</i> = 0	<i>n</i> = 0	<i>n</i> = 1	<i>n</i> = 5	<i>n</i> = 11	<i>n</i> = 9	<i>n</i> = 5
Dissemination readiness (frequency)	<i>n</i> = 1	<i>n</i> = 2	<i>n</i> = 6	<i>n</i> = 2	<i>n</i> = 4	<i>n</i> = 6	<i>n</i> = 10

Note. Total *N* = 31.

As shown in Table 2, of the 31 programs examined in this study, many clearly were either of questionable quality or not ready to be disseminated. On the quality of research, more than half of the examined programs (17 out of 31) obtained a rating less than “3,” which indicates some methodological problems in the submitted studies or simply a lack of empirical evidence. On readiness for dissemination, nearly half (15 out of 31) of the programs in this study scored less than “3,” which indicates inadequate curriculum materials, technical support, or quality assurance protocols for program adoption.

For the sake of discussion, we decided on “3” as the threshold of acceptable score for both quality of research and readiness for dissemination. Considering the fact that these clinical interventions were designed to exert real impact on participants’ lives, perhaps one should accept only a rating of “4,” in which case the majority of the registered programs would be rejected outright. Ideally, a program needs to attain adequate ratings on both scales (i.e., a minimum of “3”) to be considered suitable for adoption by practitioners. Although 14 programs scored a minimum of “3” on quality of research and 16 programs on dissemination readiness, only 7 of the 31 programs (23%) reviewed in this study scored a minimum of “3” on both scales. The rest (77% of the programs) made it into the registry nevertheless.

Because SAMHSA requires only one experimental or quasi-experimental study to qualify for NREPP registration purposes, we found a wide variation in the number of studies submitted as empirical evidence for the registered programs. For instance, 5 of the 31 programs demonstrated their treatment efficacy with only one study. Other programs received more studies. But lack of replication seems problematic.

We compiled scores for quality of research and readiness for dissemination for each category of programs to gain an idea of how these programs compared with one another. Tremendous variations were found in these registered

Table 3. Average Quality Ratings and Dissemination Readiness Scores (by Program Type)

Program Type	Outcome Quality			Dissemination Readiness		
	Mean	SD	Range	Mean	SD	Range
Substance Abuse (<i>n</i> = 17)	2.84	0.52	1.7-3.8	2.64	0.87	0.8-4.0
Mental Health (<i>n</i> = 8)	3.0	0.37	2.2-3.5	2.7	0.92	1.3-4.0
Hybrid (<i>n</i> = 6)	2.98	0.49	2.1-3.5	3.17	0.82	1.5-3.9
Total (<i>N</i> = 31)	2.91	0.49	1.7-3.8	2.76	0.96	0.8-4.0

programs. As shown in Table 3, the quality of research for substance abuse programs (*n* = 17) varied from the high of 3.8 to a low of 1.7. The average quality rating was 2.8. Dissemination readiness scores showed an even wider variation. Although the average dissemination readiness score for substance abuse programs was 2.6, the range extended from 0.8 to a perfect score of 4.

Mental health programs (*n* = 8) varied from the high of 3.5 to a low of 2.2 on quality of research and from 1.3 to 4.0 on readiness for dissemination. On average, mental health programs scored a 3 on quality of research and a 2.7 on the readiness for dissemination scale.

For the hybrid category, the mean quality rating for these programs was 2.98, ranging from a low of 2.1 to a high of 3.5. Dissemination readiness scores were fairly high, ranging from 2.9 to 3.9, with the exception of one program that dragged down the mean with a score of 1.5, with an average dissemination readiness rating of 3.17. However, there was a wide range in the number of evaluation studies for these six programs, ranging from a high of 12 to a low of 1. On average, these hybrid programs received a greater number of evaluations (5.33 per program) than programs in either of the other two categories.

Sample Size Problems

Because all NREPP programs are clinical interventions, one would assume any evaluation studies must justify sample size based on sound power analysis. In this review, we searched for three standard signs of sample size calculation: (a) the specification of null hypothesis; (b) how power analysis is related to the anticipated effects, allowing for a reasonable margin of error; and (c) the sample size required to detect a reasonable treatment effect

Table 4. Sample Size and Follow-Up Length

Number of Studies by Program Type	Sample Size			Sample Size Summary			Power Analysis			Follow-Up Period (Months)		
	(1), <100	(2), 100-399	(3), ≥400	Mean	Range	SD	(1) Yes	(2) No	(3) NA	Mean	Range	SD
Substance Abuse (n = 64)	28	25	7	231	5-2,729	397.56	21	42	1	10.11	0-36	7.0
Mental Health (n = 24)	11	12	0	140	40-342	76.63	6	18	3	15.55	3-48	10.69
Hybrid (n = 27)	16	4	7	316	23-2,041	466.17	15	12	5	12.1	1.5-48	9.21
Total (N = 115)	55	41	14	232	5-2,729	379.05	42	72	9	11.83	0-48	8.79

Note. The information in this table was based on 115 studies (out of the original 123) for which information was available. The sample size summary is based on 110 studies that disclosed this information. The follow-up information was based on 99 studies for which information was available.

Table 5. Evaluation of NREPP Registered Programs

Program Type	Number of Evaluation Studies Submitted			Developer as Evaluator
	<i>N</i>	Mean	Range	
Substance Abuse	64	3.88	1-10	Yes = 37 (58%)
Mental Health	27	3.38	2-6	Yes = 15 (56%)
Hybrid	32	5.33	1-12	Yes = 12 (38%)
Total	123	3.97	1-12	Yes = 64 (52%)

(i.e., departure from the null hypothesis). For instance, assuming an effect size of .30 (a relatively strong effect), with power of .80, and $\alpha = .05$, one would propose a sample of 360 subjects.

The range of sample sizes among the reviewed studies varied substantially, with most having fewer than 100 subjects. One study had only five subjects. Although large sample sizes are not always necessary, researchers should ideally demonstrate how they arrived at the sample size used in their studies. For the majority of the studies in this review, the number of subjects seemed to be based either on available funding or convenience instead of on sound methodologies.

Table 4 provides a summary of the sample sizes found in these studies and whether any power analysis was included in the justification of their decisions on sampling procedures. Although 42 of the 123 studies (34%) included justification of their sample sizes, 27 of these had sample sizes greater than 100, and 15 studies had relatively small sample sizes (fewer than 100 subjects). The majority of studies had relatively small sample sizes without any justification or power analysis for the sampling procedures.

Program Developer as Evaluator

Aside from examining the quality of research and readiness for dissemination, we took a look at those who produced the evidence that enabled the programs to be registered. Again, the SAMHSA sets no threshold as to who may submit evaluation studies in support of the registered programs. Nor does the SAMHSA stipulate that evaluations for NREPP registry purposes be carried out by independent researchers.

For the 31 programs in this review, a total of 123 evaluation studies were submitted as proof of “empirical evidence,” averaging slightly less than four

studies per program. We triangulated the authors of these studies against program descriptions, identified program developers, and accessed program websites. In 8 of the 31 programs (26%), we were unable to identify the program developer. NREPP does not consistently identify and disclose the program developer's identity on its website. In these cases, we assumed—on the side of caution—that the studies for these eight programs were carried out by independent researchers.

As shown in Table 5, we found that in 64 out of the 123 studies submitted, program developers were among the authors. In more than half of the 64 studies (52%), these developers were the lead authors. In some studies ($n = 20$), multiple program developers were listed as authors.

The number of evaluation studies also varied tremendously. On the low end, two of the registered programs had only one evaluation each, both involving the program developer. In one case, the program developer was listed as a lead author, whereas in the other case the developer was listed as a second author. On the high end, one substance abuse program was evaluated in 10 studies—the greatest number of studies reviewed for any particular substance abuse program—but at least one program developer was involved in each study. In total, program developers were involved in the evaluation of 52% of all studies, when a program developer could be identified.

Of the mental health programs, where we could identify the developers, all were listed as evaluators of their own programs in the submitted studies ($n = 15$), constituting 56% of the reviewed studies. In the substance abuse programs, program developers acted as their own evaluators in 58% of the submitted studies (37 out of 64). Among the hybrid programs, program developers authored 38% of their own evaluations, or 12 out of 32. Compared with the other types of interventions, these hybrid programs were less likely to have evaluation studies conducted by the program developer.

Although program developers are understandably concerned about the efficacy of their own treatment activities, these self-evaluations perhaps should be identified in the NREPP so the readers or potential adopters would not attribute equal weight as research conducted by independent evaluators because of the potential for conflict of interest. Moreover, in this review, we were unable to find any disclosures or cautionary statements on the NREPP site to warn readers of these possible conflicts of interest. Our finding was certainly not unique.

In a review of evaluation studies of “model” school-based drug and violence prevention programs, Gorman, Conde, and Huber (2007) found that program developers were listed as authors in 78% of studies. In another study, Gorman and Conde (2007), citing Tobin (2003), define conflict of

interest as “a set of conditions in which professional judgment concerning a primary interest, such as validity of research, might be influenced by a secondary interest, such as financial gain” (Tobin, 2003, p. 1161; as cited in Gorman & Conde, 2007, p. 423). Tobin juxtaposes conflict of interest with bias, the latter of which he asserts occurs whenever one’s secondary interest clearly influences their primary interest and the first of which occurs regardless of whether or not one’s primary interest was influenced by a competing or secondary interest. In other words, the mere presence of a secondary interest is enough to suggest the existence or potential existence of a conflict of interest (Gorman & Conde, 2007, p. 423). Although we did not address the differences in effect sizes between evaluations by the program developer(s) and those by independent researchers, several studies have noted that evaluations involving program developers, in general, report much larger effect sizes than evaluations carried out by independent researchers (Gorman, 2005; Gorman & Conde, 2007; Lipsey, 1995; Petrosino & Soydan, 2005).

When program developers act as their own evaluators, the “empirical evidence” becomes vulnerable to challenge. In one extreme example, one behavioral-oriented program received “validation” in 10 studies, all of which involved the program developer as either the lead or second author.³ Another program received a dozen evaluation studies, half of which were coauthored by the program developer. For the purposes of aiding community agencies in adopting interventions, it appears that the clear separation between program developers and program evaluators will at least serve to dispel any ethical complications.

Discussion

The purpose of this study was to examine the extent to which the NREPP fulfills its intended purpose, to “assist the public in identifying approaches to preventing and treating mental and/or substance use disorders that have been scientifically tested and that can be readily disseminated to the field” (NREPP, 2009). The extent to which this purpose is being carried out is important to ensure that only research of the highest quality make its way onto a registry that often serves as an implicit stamp of approval, and which community agencies consult with confidence to adopt interventions that are expected to make a significant difference in the lives of their clientele. Evidence-based studies are important, but the quality of research and the ease of dissemination must be given serious consideration.

One consistent theme that runs through this review is the tremendous variation in all aspects of the assessment among the 31 NREPP registered programs

that we reviewed. It remains unclear as to why many of these studies, even determined by SAMHSA reviewers to be methodologically flawed or not ready for dissemination, were allowed to gain entry into the NREPP. The inclusion of studies that scored less than “2” on either scale indicated a clear lack of quality threshold for what qualified as evidence-based programs.

Similarly, SAMHSA sets no criteria as to what a program needs to score on both scales to be registered. Ideally an intervention needs to meet some threshold criteria (e.g., a minimum of “3”) on both rating scales to be considered clinically useful for adoption. In other words, a program needs to demonstrate not only clinical effectiveness but also a delivery mechanism that can be used by community agencies. Currently, no such threshold seems to exist in the NREPP for qualifying purposes. Once on the list, program developers are able to market their interventions as “NREPP-registered,” which the authors of this article have witnessed on numerous occasions.

Despite the federal government’s emphasis on the importance of evidence-based research in determining funding levels for interventions, the treatment community has paid relatively little attention to the quality of evaluation research. This brings us back to the concerns addressed by Martinson and those who have built on his concerns with program implementation and the methodological quality of evaluation research, discussed at the beginning of this article. What is troubling, aside from the lack of consistency in the quality of research and readiness for implementation, is the large number of program developers who were involved in the evaluations of their own programs. In this review, 52% of the submitted studies were conducted in part by the program developer (68% if those studies were excluded in which a program developer could not be identified). Although there is nothing intrinsically wrong with developers evaluating their own programs, one would expect that for public adoptions some independent evaluation or “second opinion” should be sought to remove any slightest hint of conflict of interest. At the least, a full disclosure of the relationship between the developers and the authors should be made up front. The NREPP currently does not have a mechanism that consistently reveals the program developer’s identity on its website.

This takes us back to the question asked at the beginning of this article: What qualifies as “evidence” and who gets to produce it? Is an evidence-based study one in which an evaluation has been conducted demonstrating some outcomes (either negative or positive) or should it be more rigorous, carried out by independent researchers with no vested interest in the outcome of their studies? Developing a clear understanding of what qualifies as evidence and ensuring that evaluations are methodologically sound and ready for dissemination into the field can prevent the possibility of concluding later on that a

particular intervention—which may initially have received much fanfare—is not working. Many NREPP consumers may lack the time and training to critically evaluate the veracity of the claims found in the reports and, instead, will likely believe that these registered programs can be trusted.

It should be noted that the NREPP programs selected for this study were considered suitable for criminal justice populations. Other programs not included in this review may have much higher scores on both quality of research and readiness for dissemination. Therefore, suggestions and comments in this article regarding the NREPP must be interpreted with caution and limited to those programs that target, or could potentially be applied to, criminal justice populations, the latter of which admittedly carries a degree of subjectivity. Additionally, because of the relative small number of NREPP programs included in this study, we were unable to carry out a more rigorous meta-analysis of these descriptive findings. Despite these limitations, we still feel that some important conclusions can be drawn.

Although the NREPP can be a beneficial tool to aid community service programs in adopting interventions to better assist the populations they service, the review process can be strengthened in several ways. These should include the following: (a) the requirement of power analysis to justify sampling procedures for detecting intended treatment effects, (b) increasing the minimum number of studies that must be submitted to substantiate treatment effects, and (c) a mandatory disclosure of whether the researchers are affiliated with the programs under evaluation. Ensuring that these criteria are met will help strengthen the overall credibility and utility of the NREPP, and of evaluation-based research generally speaking, and allow community agencies to best choose interventions that can benefit their clients.

Acknowledgement

The authors would like to thank the anonymous reviewers for their comments and suggestions. Points of view expressed in this article are solely those of the authors.

Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

Notes

1. For reasons of anonymity, this study chose not to list any examples of the programs that displayed the "NREPP" logo in their program materials. However, such information is readily available should one explore the NREPP programs and search their official websites.
2. Unlike the ratings for Quality of Research, ratings on Readiness for Dissemination are more specific, with each of the five points (0-4) operationalized. More details on the scale can be found at: <http://www.nrepp.samhsa.gov/review-readiness.asp>.
3. For anonymity reasons, we chose not to reveal the name of the program or the citations of the studies.

References

- Chemers, B., & Reed, W. (2005). Increasing evidence-based programs in criminal and juvenile justice: A report from the front lines. *European Journal on Criminal Policy and Research, 11*, 259-274.
- Cullen, F. T., & Gendreau, P. (2000). Assessing correctional rehabilitation: Policy, practice and prospects. In J. Horney (Ed.), *Criminal justice 2000: Volume 3. Policies, processes, and decisions of the criminal justice system* (pp. 109-175). Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Cullen, F. T., & Sundt, J. (2003). Reaffirming evidence-based corrections. *Criminology & Public Policy, 2*, 353-358.
- Farabee, D. (2002). Reexamining Martinson's critique: A cautionary note for evaluators. *Crime & Delinquency, 48*, 189-192.
- Gendreau, P., Goggin, C., Cullen, F. T., & Paparozzi, M. (2002). The common sense revolution and correctional policy. In J. Maguire (Ed.), *Offender rehabilitation and treatment: Effective programs and policies to reduce re-offending* (pp. 359-386). Chichester, England: Wiley.
- Gorman, D. M. (2005). Drug and violence prevention: Rediscovering the critical rational dimension of evaluation research. *Journal of Experimental Criminology, 1*, 39-62.
- Gorman, D. M., & Conde, E. (2007). Conflict of interest in the evaluation and dissemination of "model" school-based drug and violence prevention programs. *Evaluation and Program Planning, 30*, 422-429.
- Gorman, D. M., Conde, E., & Huber, J. C., Jr. (2007). The creation of evidence in "evidence-based" drug prevention: A critique of the Strengthening Families Program Plus Life Skills Training evaluation. *Drug and Alcohol Review, 26*, 585-593.
- Latessa, E. J., Cullen, F. T., & Gendreau, P. (2002). Beyond correctional quackery: Professionalism and the possibility of effective treatment. *Federal Probation, 66*, 43-49.
- Lipton, D. S., Martinson, R., & Wilks, J. (1975). *The effectiveness of correctional treatment: A survey of treatment evaluation studies*. New York, NY: Praeger.

- Lipsey, M. W. (1995). What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents? In J. McGuire (Ed.), *What works? Reducing reoffending* (pp. 63-78). New York, NY: Wiley.
- MacKenzie, D. L. (2000). Evidence-based corrections: Identifying what works. *Crime & Delinquency*, *46*, 457-471.
- Martinson, R. (1974). What works?—Questions and answers about prison reform. *Public Interest*, *35*, 22-54.
- National Registry of Evidence-based Programs and Practices. (2009). Retrieved from <http://www.nrepp.samhsa.gov/about.asp>
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, *1*, 435-450.
- Sherman, L. W. (2000). Reducing incarceration rates: The promise of experimental criminology. *Crime & Delinquency*, *46*, 299-314.
- Sherman, L. W. (2009). Evidence and liberty: The promise of experimental criminology. *Criminology and Criminal Justice*, *9*, 5-28.
- Sherman, L. W., Gottfredson, D., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising*. Washington DC: National Institute of Justice.
- Weisburd, D. (2000). Randomized experiments in criminal justice policy: Prospects and problems. *Crime & Delinquency*, *46*, 181-193.
- Weisburd, D. (2003). Ethical Practice and evaluation of interventions in crime and justice: The moral imperative for randomized trials. *Evaluation Review*, *27*, 336-354.

Bios

Benjamin J. Wright is currently a researcher with the San Diego State University Research Foundation. His current research involves prisoner reentry/program evaluation, immigration, and social class and media.

Sheldon X. Zhang is a professor of sociology at San Diego State University. His main research interests are community-based correctional efforts, program evaluation, and transnational criminal organizations. He has authored three books and his articles have appeared in *Criminology*, *British Journal of Criminology*, *Research in Crime and Delinquency*, *Criminology and Public Policy*, and *Crime & Delinquency*.

David Farabee is professor in residence at the Semel Institute for Neuroscience and Human Behavior Institute for Neuroscience and Human Behavior and director of the Juvenile Justice Research Group at the Integrated Substance Abuse Programs at the University of California, Los Angeles.

DEVELOPING CRIMINAL JUSTICE HEALTH HOMES FOR CALIFORNIA

Outline

- Accountable Care Act Section 2703 Guidelines
- State Statute – AB 361 and DHCS Process
- Differences between Health Home vs. Patient Centered Medical Home
- Considerations for Health Home / Medical Home design

Who Is Eligible for an ACA Section 2703 Health Home?

1. Health Homes are for people with Medicaid who:
 - Have 2 or more chronic conditions
 - Have one chronic condition and are at risk for a second
 - Have one serious and persistent mental health condition
2. Chronic conditions listed in ACA Section 2703 include:
 - Mental health
 - Substance abuse
 - Asthma
 - Diabetes
 - Heart disease
 - Being overweight
 - Additional chronic conditions, such as HIV/AIDS, may be considered by CMS for approval
3. States may set acuity level eligibility requirements and geographic limit
4. States cannot exclude people with both Medicaid and Medicare from health home services

Section 2703 Health Home Services

Must include:

- Comprehensive care management
- Care coordination
- Health promotion
- Comprehensive transitional care/follow-up
- Patient & family support
- Referral to community & social support services

Federal guidance strongly encourages the use of health information technology as feasible and appropriate to link these services.

Section 2703 Health Home Providers

States have flexibility to determine eligible health home providers. Health home providers can be:

- A designated provider: May be a physician, clinical/group practice; rural health clinic, community health center, community mental health center, home health agency, pediatrician, OB/GYN, or other provider.
- A team of health professionals: May include physicians, nurse care coordinators, nutritionists, social workers, behavioral health professionals, and can be free-standing, virtual, hospital-based, or a community mental health center.
- A health team: Must include medical specialists, nurses, pharmacists, nutritionists, dieticians, social workers, behavioral health providers, chiropractics, licensed complementary and alternative practitioners.

Section 2703 Reporting Requirements

Health Home service providers must report quality measures to the state. States are also required to report utilization, expenditure and quality data for an interim survey and an independent evaluation.

Quality measures tracked by CMS are:

- Adult Body Mass Index (BMI) Assessment
- Ambulatory Care - Sensitive Condition Hospital Admission
- Care Transition – Transition Record Transmitted to Health care Professional
- Follow-up After Hospitalization for Mental Illness
- Plan- All Cause Readmission
- Screening for Clinical Depression and Follow-up Plan
- Initiation and Engagement of Alcohol and Other Drug Dependence Treatment
- Controlling High Blood Pressure

Section 2703 Health Home Financing

States have the flexibility in designing their payment methodologies and may propose alternatives. States receive a 90% enhanced Federal Medical Assistance Percentage (FMAP) for the specific health home services in Section 2703. The enhanced match doesn't apply to the underlying Medicaid services also provided to people enrolled in a health home.

The 90% enhanced FMAP is good for the first eight quarters the program is effective. A state can get more than one period of enhanced FMAP, but can only claim the enhanced FMAP for a total of eight quarters for one enrollee.

AB-361 Medi-Cal: Health Homes for Medi-Cal Enrollees and Section 1115 Waiver Demonstration Populations with Chronic and Complex Conditions

- Enacted and signed in 2013
- Authorizes DHCS to submit Medicaid State Plan Amendment and 1115 Waiver to implement
- Stakeholder process to begin in spring 2014
- DHCS lead is Brian Hansen

AB 361 Target Populations

Target populations include adult beneficiaries who meet both of the following criteria:

1. Have current diagnoses of chronic, physical health, mental health, or substance use disorders prevalent among frequent hospital users.
2. Have a level of severity in conditions established by DHCS, based on one or more of the following factors:
 - Frequent inpatient hospital admissions, including hospitalization for medical, psychiatric, or substance use related conditions.
 - Excessive use of crisis or emergency services.
 - Chronic homelessness.

AB 361 Health Home Provider Definition

“Health home” means a provider or team of providers designated by DHCS that satisfies all of the following:

1. Meets the criteria described in federal guidelines.
2. Offers a whole person approach, including, but not limited to, coordinating other available services that address needs affecting a participating individual’s health.
 - Offers services in a range of settings, as appropriate, to meet the needs of an individual eligible for health home services.
 - A lead provider may contract with Medi-Cal providers, including, but not limited to, a managed care health plan, a community clinic, a mental health plan, a hospital, physicians, a clinical practice or clinical group practice, a rural health clinic, a community health center, a community mental health center, substance use disorder treatment professionals, school-based health centers, community health workers, community-based service organizations, a home health agency, nurse practitioners, physician’s assistants, social workers, and other paraprofessionals, to the extent that contracting with these providers is allowed under federal Medicaid law. Health home providers shall also establish non-contractual relationships with, and provide linkages to, housing providers.
 - For purposes of serving the populations identified DHCS may require a lead provider to be a physician, a community clinic, a mental health plan, a community-based organization, a county health system, or a hospital.
 - DHCS may determine the model of health home it intends to create, including any entity, provider, or group of providers operating as a health team, as a team of health care professionals, or as a designated provider

AB 361 Health Home Provider Selection Criteria

1. DHCS health home providers or providers who plan to subcontract with health home team members with all of the following:
 - Demonstrated experience working with frequent hospital or emergency department users.
 - Demonstrated experience working with people who are chronically homeless.
 - The capacity and administrative infrastructure to participate in the Health Home Program, including the ability to meet requirements of federal guidelines.
 - A viable plan, with roles identified among providers of the health home, to do all of the following:
 - i. Reach out to and engage frequent hospital or emergency department users and chronically homeless eligible individuals.
 - ii. Link eligible individuals who are homeless or experiencing housing instability to permanent housing, such as supportive housing.
 - iii. Ensure coordination and linkages to services needed to access and maintain health stability, including medical, mental health, and substance use care, as well as social services and supports to address social determinants of health.

Section 2703 Health Homes vs. Patient Centered Medical Homes

Section 2703 Health Home	Patient Centered Medical Home
Enhanced Medicaid reimbursement for coordination of care for persons with chronic conditions	Team-based, integrated, whole-person care
May include primary care practices, community mental health organizations, addiction treatment providers, federally quality health centers, health home agencies, and other provider groups	A single primary care or behavioral health organization that coordinates the care for a patient
Currently a Medicaid-only construct	Not based on a single source of payment

What is a Patient Centered Medical Home?

1. **Personal Physician:** Each patient has an ongoing relationship with a personal physician that provides continuous, comprehensive care.
2. **Physician-directed Care:** Personal physicians lead a team of individuals who take collective responsibility for a patient.
3. **Whole-person Orientation:** Personal physicians take responsibility for providing or arranging all of a patient's health needs.
4. **Coordination and/ or Integrated Care:** Care is coordinated or integrated across all elements of the health care system and a patient's community.
5. **Quality and Safety:** Practices advocate for patients to attain optimal, patient-centered outcomes resulting from evidence-based care and decision-making.
6. **Enhanced Access:** Care is available through open scheduling, expanded hours, and new communication options.
7. **Payment:** Reimbursement reflects the added value of medical home services.¹⁵

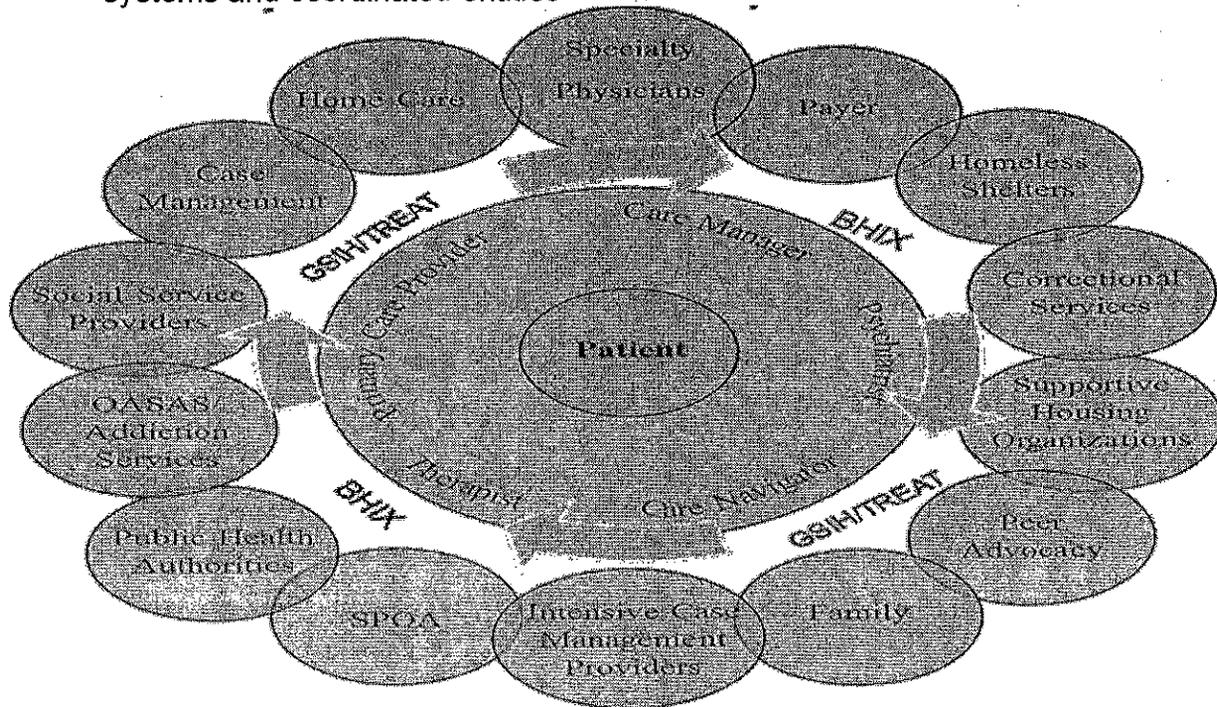
Source: Joint principles developed in 2007 by American Academy of Family Physicians, the American Academy of Pediatrics, the American College of Physicians, and the American Osteopathic Association.

Health Home Policy Considerations

- Section 2703 Health Homes are limited to Medi-Cal recipients with chronic conditions
- How will health care gaps for non-Medi-Cal patients and for those without chronic conditions be addressed?
- Should CDCR develop a separate State Plan Amendment?
- Links to AB 720 Medi-Cal Enrollment and to AB 109

NY State Health Home Planning Process

- Planning apparently began after the health homes were created
- No separate Medicaid state planning amendment for CJ population
- Involves Six Section 2703 Health Homes
- Model of care coordination showing patient, immediate care providers, health information systems and coordinated entities



Source: LAC_2013_CL_Brooklyn_NYS_Health_Homes_and_Criminal_Justice_Implementation.pdf, Retrieved 01.12.13.

California Medi-Cal Health Plans

- May see themselves as natural location for health home services
- All provide some degree of care coordination typically over the phone with the patient
- Already coordinating care for dual eligibles (must coordinate with carved out MH/SUD county agencies)
- Funding for health homes may be through additional capitation paid to health plans and passed on as a capitated payment to those providing health home services
- If health home services provided by another organization need to avoid duplication of health home services/cost

Importance of Care Integration

- Integration of Primary Care and Behavioral Health is Health Home Requirement
- SAMHSA Primary Care Behavioral Health Integration Grantees in CA as models
- CMS approval process for Health Home State Plan Amendment requires clearance with SAMHSA PBHCI Program

Importance of Care Coordination and Health Information Exchange

- Impacting the quality measures listed on page 2 will require close coordination of care with hospitals and specialty care providers
- Fostering of HIT systems to promote patient data exchange are especially critical for transitions of care
- Health Information Exchange is lagging the rest of the U.S. in most of California

Conclusions and Next Steps

- Section 2703 health homes could help a high proportion of released inmates because many are Medi-Cal eligible and have chronic conditions.
- If the problem is broader than those with chronic conditions, foster creation of patient centered health homes
- Develop a criminal justice specific 2703 state plan amendment to pilot test approaches to developing health homes for CJ
- Coordinate now with the public health plans in the two plan counties – e.g., LA Care, about health homes which would be good partners for health home development – but not to operate health homes
- Funding for health homes may be through additional capitation paid to health plans paid to those providing health home services.