

Regular article

Does following research-derived practice guidelines improve opiate-dependent patients' outcomes under everyday practice conditions? Results of the Multisite Opiate Substitution Treatment study

Keith Humphreys, (Ph.D.)^{a,b,*}, Jodie A. Trafton, (Ph.D.)^{a,b}, Elizabeth M. Oliva, (B.S.)^{a,c}

^aCenter for Health Care Evaluation, Veterans Affairs and Stanford University Medical Centers, Palo Alto, CA 94304, USA

^bDepartment of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305-5717, USA

^cDepartment of Psychology, University of Minnesota, Minneapolis, MN 55455-0344, USA

Received 30 October 2006; received in revised form 23 February 2007; accepted 4 March 2007

Abstract

The Multisite Opiate Substitution Treatment study evaluated whether adhering to clinical-trial-derived practice guidelines improves treatment outcomes of unselected opiate-dependent patients seen in everyday practice. Clinics that were relatively concordant ($n = 4$) or nonconcordant ($n = 4$) with guidelines concerning medication dose levels and psychosocial service provision were identified. Staff interviewed 256 patients at intake and 6-month follow-up regarding past month heroin use, criminal activities, and mental health. To represent real-world practice conditions, clinics provided care in accordance with their usual approach, and no patient exclusion criteria were employed. Patients in each type of clinic were similar at baseline, but by follow-up, heroin use and mental health outcomes were significantly better in guideline-concordant clinics than in guideline-discordant clinics. Notably, 60.6% of patients in concordant clinics had urinalysis-confirmed heroin abstinence versus only 40.0% in nonconcordant clinics. Following research-derived practice guidelines seems to increase opiate substitution treatment effectiveness for opiate-dependent patients in the real world. Published by Elsevier Inc.

Keywords: Opioid substitution; Evidence-based medicine; Clinical practice guidelines; Heroin dependence; Effectiveness research

1. Introduction

Frontline clinicians and academic scientists often disagree as to what constitutes “good evidence,” posing a major challenge to progress in evidence-based medicine. In the eyes of many scientists, clinical practice should be guided by the findings of well-controlled clinical trials with standardized treatments administered to homogenous, selected patient samples. Yet, many clinicians believe that everyday practice is too variable and that real-world patients are too diverse for practice to be based on efficacy studies conducted under ideal conditions (Kernick, 1998). Opiate substitution

treatment (OST) for heroin dependence exemplifies the dilemma. Multiple rigorous randomized trials have yielded important evidence about OST provision, but many frontline clinicians do not accept these trials as definitive and do not follow the practice guidelines they inform (D'Aunno & Vaughn, 1992). Scientists and clinicians have productively engaged the “efficacy in trials vs. effectiveness in the real world” debate at a conceptual level (Wells, 1999); the Multisite Opiate Substitution Treatment (MOST) study offers an empirical perspective on these issues. Specifically, we evaluate whether two OST clinical practices found efficacious in clinical trials and incorporated into practice guidelines (dosing in the recommended range and providing psychosocial services) improve the outcomes of typical opiate-dependent patients seen in everyday practice.

OST has probably been the subject of more randomized trials than any other type of treatment for illicit drug

* Corresponding author. Department of Psychiatry, Stanford University, Stanford, CA 94305-5717, USA. Tel.: +1 650 617 2746; fax: +1 650 617 2736.

E-mail address: knh@stanford.edu (K. Humphreys).

dependence. An early double-blind clinical trial of 100 mg/day methadone maintenance versus an initial 60-mg/day dose decreased by 1 mg/day found that retention was over seven times higher in the high and stable dose condition and that heroin use decreased only in this condition (Newman & Whitehill, 1979). In addition to supporting the efficacy of a high dose, this trial also helped establish the value of OST by showing that the placebo effect did not explain methadone's efficacy. Other trials evaluating dosing practices have found that retention and heroin use outcomes are significantly better in higher dose conditions (e.g., 80 mg of L- α -acetylmethadol [LAAM], 80–100 mg of methadone) than in conditions providing doses lower than 60 mg/methadone equivalent/day (Ling, Charuvastra, Kaim, & Klett, 1976; Strain, Bigelow, Liebson, & Stitzer, 1999). These studies have directly informed practice guidelines (American Psychiatric Association, 1994; Department of Veterans Affairs/Department of Defense, 2004) recommending that OST patients receive medication doses of at least 60 mg of daily methadone and at least 70 mg of 72-hour LAAM (e.g., over a weekend when the clinic is closed).

Another important clinical trial demonstrated that OST patients assigned to a methadone-only condition had worse drug use, job-related, and criminal behavior outcomes than did individuals assigned to receive additional psychosocial services, such as psychiatric and employment counseling (McLellan, Arndt, Metzger, Woody, & O'Brien, 1993). This finding is also reflected in many clinical practice guidelines, which emphasize the importance of psychosocial service provision in OST treatment (American Psychiatric Association, 1994).

Although providing high doses of methadone/LAAM and extensive psychosocial services seems well supported, many OST clinics do not follow these practices (D'Aunno & Vaughn, 1992). Sometimes, this is due to poor training and resource constraints, but another factor is the skepticism of some clinicians that trials can and should drive clinical practice, an attitude encapsulated in an editorial in *The Lancet*, entitled "Lies, damned lies, and evidence-based medicine" (Kernick, 1998). Two common reservations among clinicians about the usefulness of trials and the practice guidelines they support concern the way trials construct treatment conditions and the way they exclude some patients from participation. We discuss each of these concerns in turn.

First, in the experimental condition of a randomized trial, providers are usually well trained and closely monitored and may also be treatment manual guided. These features may result in more homogenous and higher quality clinical practice than is attainable in the real world. Relatedly, the control conditions in trials, such as those of McLellan et al. (1993), may also be unlike everyday practice because they provide care that is uniformly poor; that is, the control condition by definition ensures that all patients received the same guideline-discordant care (whereas, in the real world,

even a relatively poor clinic may provide some good care). By studying extreme and homogeneous treatment conditions, which may form endpoints outside of the usual range of real-world practice, controlled trials may overstate how much of a difference following clinical practice guidelines can make in everyday practice. An overstated analogy may help illustrate this point: Although it is important for science to know that exercising 4 hours a day leads to more weight loss than does watching television over the same period, this same information may not be of much use to most overweight people, who are making decisions more along the lines of whether or not to replace a few hours of television viewing per week with exercise.

Second, most addiction treatment trials have eligibility criteria that can prevent enrollment of the most severely troubled patients (Humphreys & Weisner, 2000; Humphreys, Weingardt, Horst, Joshi, & Finney, 2005). Frontline clinicians may therefore be reticent (reasonably or not) to rely on the results of clinical trials when treating patients who, for example, are homeless, are antisocial, or have serious medical and psychiatric comorbidities.

As mentioned, the different perspectives of clinicians and researchers about how much scientific knowledge can be generalized to practice have been much debated at a conceptual level. However, when one discusses specific practices and treatments, the size of any perceived efficacy—effectiveness gap becomes an empirical question, which is how it was addressed in the MOST study. Specifically, we conducted a quasi-experimental study under real-world practice conditions (Wells, 1999) to determine whether adhering to clinical practice guidelines on OST dosing and psychosocial service provision enhances the outcomes of the full array of patients with multiple problems who seek this treatment.

2. Materials and methods

2.1. Ethics committee approval

All MOST study procedures were approved by the Institutional Review Board of Stanford University and by each participating U.S. Department of Veterans Affairs (VA) medical center.

2.2. Selection of clinics and creation of study conditions

This study was conducted in the VA health system, a federal, publicly funded network of clinics and hospitals structured not unlike many European health care systems (e.g., the UK National Health Service). All 34 VA OST clinics were mailed a one-page screener to assess their concordance with practice guidelines regarding dosing patients above 60 mg/methadone per day and providing extensive psychosocial services; 31 clinics (91.1%) responded. These 31 clinics were ranked on the proportion

of their patients dosed in the recommend range and on the number of clinical staff available per enrolled patient. These two rankings were averaged to produce a single number ranging from 1 to 31 for each clinic, reflecting its overall concordance with practice guidelines.

Four clinics that were above the median rank (i.e., ranked 1st to 15th out of 31) on guideline concordance and four clinics that were below the median rank on guideline concordance were invited to participate. Within each sampling pool (i.e., the top half and bottom half ranked clinics), individual clinics were selected based on their being large enough to provide a steady stream of new potential research participants and on their being in different regions of the country. The four participating concordant clinics averaged a rank of 12.0 ($SD = 3.5$) on guideline concordance, and the four participating nonconcordant clinics averaged a rank of 21.9 ($SD = 3.6$) among the 31 clinics that completed the screener. This resulted in two study conditions (relatively concordant and nonconcordant OST care), each with four urban OST clinics throughout the country. Both conditions contained one clinic each in New England, the Southeast, the Midwest to Southwest, and the Far West.

Importantly, this method of forming study conditions provides some protection against patient self-selection bias. The VA operates OST clinics in only 34 cities throughout the United States; thus, whether opiate-dependent VA patients receive guideline-discordant or guideline-concordant OST care is primarily a function of where they live rather than what type of OST they choose (i.e., in practical terms, it would be difficult to travel several hundred miles per day to enroll in the next nearest VA clinic because one preferred its approach to clinical practice). Second, the study conditions were formed based on the typical practice patterns of each clinic across its caseload, not by the specific services each individual participant in the study received. This approach to forming the study conditions helps surmount the self-selection bias inherent in a patient-level analysis, in which, for example, the same patient variables (e.g., being well organized, motivated, and educated, as well as having less severe problems) probably predict both services received and outcomes.

In the four concordant clinics, an average of 79.2% ($SD = 12.3\%$) of patients were dosed in the guideline-recommended range, compared with an average of 46.7% ($SD = 23.3\%$) in the four nonconcordant clinics. Assessing concordance with guidelines for psychosocial services is inherently less precise because the guidelines themselves are not highly specific, recommending, for example, that medication be provided “in combination with appropriate counseling” (American Psychiatric Association, 1994) and that treatment should be a “supportive recovery environment” that offers “regular counseling” (Department of Veterans Affairs/Department of Defense, 2004). In the MOST study, the concordant clinics had substantially more

ability to comply with these general guidelines because they had more counselors and better staff-to-patient ratios. Specifically, although the two groups of clinics had similar patient censuses (concordant mean = 170, $SD = 87$; nonconcordant mean = 154, $SD = 64$), the concordant clinics averaged 0.77 of a standard deviation more full-time equivalent clinical staff than the nonconcordant clinics (concordant mean = 8.67 full-time equivalents, $SD = 3.3$; nonconcordant mean = 6.35, $SD = 2.6$; pooled sample $SD = 3.0$).

Thus, the concordant clinics were relatively more concordant to practice guidelines for both medication dose level and psychosocial service provision, but the study conditions were not as distinct as would be the case in a typical randomized trial. Rather, consistent with our goal of maximizing external validity, the difference between treatment conditions was a direct function of the range of practice variation in real-world treatment of opiate-dependent patients.

2.3. Patient recruitment

To represent typical practice conditions, new OST patients were asked to participate in the study regardless of their demographic or problem characteristics (i.e., there were no participant eligibility criteria). Clinic staff explained the study to incoming patients and obtained informed consent from those who were interested in participation (95.7% of patients expressed interest in participation). Participants' signed consent form and contact information were faxed to the research team, who telephoned the patient to administer a research interview and to gather contact information to facilitate a reinterview 6 months later. Of the 267 patients who consented to participate during the recruitment period, 256 (95.9%) completed the intake interview. The follow-up rate at 6 months was extremely high (232 of 256 patients [90.6%]). Participants were compensated US\$25 for each interview they completed.

2.4. Measures

Five core outcome measures were chosen based on their coverage of domains typically assessed in OST clinical trials and on their being of immediate, clear relevance to frontline clinicians. Days of heroin use, days of employment, and days of illegal activities other than drug use in the past 30 days were assessed by single, self-report items at intake and follow-up. Global mental health was assessed by the mental health scale of the Short Form-36 for Veterans (SF-36V), a widely used measure with excellent psychometric characteristics (Kazis et al., 1999). Finally, so as not to rely solely on self-report, data regarding urinalysis-confirmed abstinence from illicit opiates in the month prior to follow-up were drawn from the pharmacy and toxicology database of each clinic's medical center for all patients still in treatment.

3. Results

3.1. Participants

Almost all of the 232 participants who completed intake and follow-up interviews were men (97.8%), and just more than half (53.9%) were African American. A total of 21.1% were married, 54.3% were separated or divorced, 3.9% were widowed, and 20.7% had never married. The most common religions were Protestant (56.9%) and Catholic (24.1%). Participants had a mean age of 49.3 ($SD = 7.1$) years and had, on average, 12.6 ($SD = 1.9$) years of education. Supporting the generalizability of the sample, these demographic characteristics are almost identical to those of national samples of thousands of heroin-dependent patients who participated in the VA's national substance abuse treatment monitoring program (Moos, Federman, & Finney, 1999).

All participants were opiate dependent; participants had used heroin, on average, on 22.0 ($SD = 11.8$) of the past 30 days. In the 30 days prior to intake, participants were paid for working an average of 7.2 ($SD = 10.5$) days and engaged in criminal activity other than drug use on 3.6 ($SD = 8.7$) days. Mental health functioning at baseline averaged 54.5 ($SD = 25.2$) on the 100-point scale of the SF-36V. Published population norms (Kazis et al., 1999) on this scale for men aged 45–54 are a mean score of 76.4 and a 25th percentile score of 68.0. Thus, as would be expected from the MOST study's lack of exclusion criteria, the sample had a very high level of psychiatric impairment.

3.2. Baseline comparison

Large differences in patient characteristics at intake across the two conditions would compromise the ability to draw inferences about the effects of treatment. Hence, individuals receiving treatment in concordant ($n = 150$) versus nonconcordant ($n = 82$) clinics were compared at baseline using chi-square and t tests on demographic and problem variables. The two groups did not differ significantly on gender (concordant = 97.3% male, nonconcordant = 98.8% male), race (concordant = 57.3% African American, nonconcordant = 47.6% African American), marital status (concordant = 20.0% married, 53.3% separated/divorced; nonconcordant = 23.2% married, 56.1% separated/divorced), or religion (concordant = 58.7% Protestant, nonconcordant = 53.7% Protestant), but patients at nonconcordant sites were about 2 years older than those at compliant sites (mean = 51.1 years vs. 48.3 years, $t = 2.85$, $p < .01$). There were no baseline differences in the four self-reported outcomes. Rate of follow-up was almost identical across conditions (concordant = 90.9%, nonconcordant = 90.1%).

3.3. Outcome analysis

Mixed-effect linear models were used to evaluate the impact of guideline concordance on the four continuous

Table 1

Descriptive statistics and effects of guideline-concordant care for heroin, employment, crime, and mental health outcomes at 6-month follow-up

Condition	Intake, <i>M</i> (<i>SD</i>)	6-Month follow-up, <i>M</i> (<i>SD</i>)	Effect (<i>SE</i>) ^a	95% CI
Days of heroin use				
Concordant	23.0 (11.0)	2.8 (6.9)	-2.9 (1.0)	-5.0, -0.9
Nonconcordant	20.2 (13.1)	5.0 (8.5)		
Days of employment				
Concordant	7.9 (10.9)	7.9 (10.6)	-0.2 (1.3)	-2.8, 2.4
Nonconcordant	5.7 (9.8)	6.6 (10.8)		
Days of illegal activities				
Concordant	3.5 (8.7)	0.9 (4.8)	0.0 (0.8)	-2.0, 2.0
Nonconcordant	3.7 (8.8)	0.9 (4.0)		
SF-36V mental health scale				
Concordant	54.5 (24.9)	63.5 (25.6)	7.7 (2.9)	2.1, 13.4
Nonconcordant	55.1 (24.9)	56.0 (24.7)		
	%	%		
Urine-confirmed abstinence from illicit opiates				
Concordant	0.0	60.6	2.7 (1.0)	1.7, 4.5
Nonconcordant	0.0	40.0		

^a For the effect values in boldface, $p < .01$. Effects are the difference in the group means (concordant – nonconcordant) estimated in a mixed-effect linear model with the fixed effects of treatment condition, centered baseline levels of outcome and age, effect of region, and clinics as a grouping variable (random effect).

outcome variables. Each outcome was predicted by the fixed effect of clinic type (concordant vs. nonconcordant) in a model that controlled for geographic region, patient age, and the intercorrelated nature of the patients within clinics (random-effect grouping variable). Predictors were centered prior to analyses, and maximum likelihood estimation was used. Analyses were conducted using the NLME (Pinheiro, Bates, DebRoy, & Sarkar, 2000), packages within the R software system (R Development Core Team, 2005).

As shown in Table 1, patients in concordant clinics reduced their days of heroin use (from 23.0 to 2.8 days per month [88%]) more than did individuals in guideline-discordant clinics (from 20.2 to 5.0 days per month [75%]). The mixed-effect model estimated the effect size as 2.9 ($SE = 1.0$) fewer days of heroin use per month in concordant than in nonconcordant clinics ($p < .01$). No significant effects of clinical guideline concordance were found in the models for days of employment or days of illegal activities.

However, on the SF-36V global mental health scale, patients in concordant clinics improved over a third of a standard deviation (54.5 at intake, 63.5 at follow-up) compared with negligible improvement among patients in nonconcordant clinics (55.1 at intake, 55.9 at follow-up). The mixed-effect model estimated the effect size of guideline concordance as 7.7 points ($SE = 2.9$), which was statistically significant ($p < .01$).

Self-report data could be supplemented for patients still in treatment with urinalysis results for the month prior to follow-up. A total of 164 patients provided one or more urine tests during this period. The rate of test completion

was 63.0% for concordant clinics and 65.9% for non-concordant clinics. Patients providing versus not providing urine tests did not differ significantly ($p > .05$ on independent-samples t tests) on study condition, demographic variables, or baseline values of the outcome variables. A generalized estimating equation (a preferred method when the outcome is dichotomous and within-site nesting of observations exists) was used to estimate the odds of a patient having all illicit-opiate-free urinalyses in the sixth month, after controlling like the prior mixed-effect models for age and region. An exchangeable covariance structure and binomial link function were specified. In the sixth month, the percentage of patients having all of their urine tests free of illicit opioids was significantly higher in concordant clinics (60.6% of patients) than in nonconcordant clinics (40.0% of patients; odds ratio = 2.7; $SE = 1.0$; 95% confidence interval [CI] = 1.7, 4.5; $p < .01$). A sensitivity analysis assuming that all patients who did not provide urinalysis data would have screened positive yielded similar results (odds ratio = 2.2; $SE = 0.8$; 95% CI = 1.35, 3.60; $p < .01$). In short, the urine tests echoed the self-report data in showing greater decreases in illicit opiate use in the concordant clinics than in the non-concordant clinics.

This article focused on the five core outcomes of the MOST study. In supplemental analyses, a range of casemix-adjusted secondary outcomes were examined, including alcohol and cocaine use, social functioning, physical pain and other medical problems, high-risk injection practices, and treatment satisfaction. The pattern of secondary results (available from the authors) resonated with those found on the core outcomes, with the concordant clinics producing better casemix-adjusted 6-month outcomes than nonconcordant clinics on about 60% of outcomes assessed and comparable outcomes on the remaining 40% of outcomes measures. As with the core outcomes presented here, in no case were the secondary outcomes of the nonconcordant clinics superior to those generated by the clinics that more closely followed clinical practice guidelines.

4. Discussion

The MOST study evaluated whether two clinical practices found efficacious in controlled trials and recommended in practice guidelines improved OST outcomes under everyday practice conditions. The results demonstrate the clinical utility of trial-supported guidelines: Within the range of real-world practice variation and with unselected, severely impaired patients, OST clinics that dose more patients in the clinically recommended range and provide more psychosocial services have significantly better drug use and mental health outcomes. Because controlled trial conditions are so different from typical practice conditions, this result was not necessarily expected but is certainly welcome, as it bolsters modern medicine's fundamental faith

in the ability of rigorous science to inform frontline clinical practice. It also suggests that efforts to increase adherence to clinical practice guidelines for OST through policy changes, performance incentives, training programs, and so forth will have a positive impact on patients.

How "good" were the outcomes achieved in the relatively guideline-concordant clinics relative to those observed in controlled trials? McLellan et al.'s (1993) well-known trial was also conducted in the VA and also reported the proportion of patients providing opiate-free urines in the 4 weeks prior to the 6-month follow-up. Abstinence rates were in the 65–70% range for patients in the guideline-concordant conditions (a similar rate was reported in the trial of Newman & Whitehill, 1979). The 60.6% abstinence rate observed here is, if not equal, fairly close to the standard set in studies conducted under ideal conditions with closely monitored staff and selected samples of patients. This finding is encouraging because it shows that, at its best, everyday clinical practice in OST programs can be almost as effective as that occurring under the ideal conditions of an academic research project.

Importantly, even those clinics with relatively guideline-discordant practices produced improved outcomes in absolute terms and adhered to guidelines more than would be the case than in a control condition of a clinical trial. The world of practice is more heterogeneous and individualized than the world of trials, which, by design, typically evaluate the impact of uniformly good care versus uniformly poor care. Because even a relatively guideline-discordant clinic may generate major health improvements in some patients (as was clearly the case in the MOST study), practitioners in such clinics understandably react negatively to implications that their clinics will transform from ugly worms into beautiful butterflies if "they just follow the scientist's guidelines." For their part, treatment providers who are dismissive of trial-supported practice guidelines should weigh carefully the present findings showing that, for OST at least, increasing adherence to practice guidelines will probably translate into increased effectiveness, even if some patients are benefiting under the current practice regime. The MOST study's results suggest that even modest increases in an OST clinic's guideline adherence could be lifesaving for some patients: After all, every injection of illicit opiates carries risk of overdose and HIV/hepatitis C infection, and individuals with extremely poor mental health are at high risk for suicide.

The central conclusion of the MOST study rests on the assumption that the observed outcomes are attributable to differences in the degree of guideline concordance between the two clinics. However, other factors may have contributed to the obtained results. In any naturalistic study, one such alternative explanation that must be considered is patient self-selection into treatment condition. Patients in each condition did not differ at intake on any outcome variable. They were also similar on all but one demographic variable. Patients in the guideline-discordant condition were

about 2 years older than those in the guideline-concordant condition. This difference seems unlikely to have produced the observed results because greater age tends to predict better rather than worse OST outcomes (McLellan, 1983); hence, any selection bias due to age would have understated rather than overstated the main conclusion. All that said, one can only test for baseline differences in patient groups on measured variables. In a randomized trial, there is a theoretical case for assuming no selection bias from unmeasured variables, but in a nonrandomized study, any such differences cannot be assumed unimportant. It is hard to think of an unmeasured patient variable that would have produced the MOST study's outcomes but not evinced itself in baseline differences, but this possibility cannot be ruled out. That is the sacrifice in internal validity we made by conducting an observational study instead of a randomized experiment.

Could unmeasured differences in the OST clinics explain the results? In a laboratory experiment, one factor can be manipulated independently from other factors; in health care organizations, this is almost never true. For example, multiple studies have documented that how mental health staff interact with patients has reciprocal relationships with how staff interact with each other (Kyrouz & Humphreys, 1997). Our team's informal, unstandardized observations during the conduct of the MOST study were that clinics that implemented practice guidelines in an orderly and efficient manner completed most work tasks also in an orderly and efficient manner (cf. Moos, 1994). In contrast, in their interactions with the MOST project team, the nonconcordant clinics were generally more likely to be slow in returning telephone calls, to misplace forms and paperwork associated with the project, or both. As clinics, they also seemed less efficient at maintaining a steady flow of new patients into treatment. None of this is to argue that these aspects of work culture are a "confound" in the experimental sense of being the true and only explanation for the results (which would constitute asserting that medication level and amount of services have no real effect on patients). However, it does suggest that efforts to promote clinical practice guideline concordance may inherently involve changes in other aspects of a health care organization's work culture. This is perhaps disappointing from the point of view of "clean" organizational theories of technology transfer, but in practical terms, it may be less so because disseminating practice guidelines—which give a rationale, order, and shared mission in treatment—may, in itself, be a valuable intervention for less organized and efficient work settings.

What the MOST study lost in internal validity (at least relative to a randomized trial), it gained in external validity. Many addiction treatment studies exclude patients who are homeless, are socially isolated, are interpersonally aggressive, or have serious psychiatric and medical comorbidities (Humphreys & Weisner, 2000). None of such challenging patients were excluded from the MOST study. The study's

external validity was further enhanced by its examination of treatment practices attainable in everyday public sector practice. The study's central inference on the value of OST practice guidelines is, thus, generalizable to frontline clinical work, rather than only to situations where clinicians are highly trained, well resourced, closely monitored, and/or manually guided as they often are in controlled trials.

We believe that the research design employed here makes particular sense when put in the context of the maturity of the OST literature. OST has the strongest efficacy evidence of any treatment for opiate dependence, based on trials conducted over the past 30 years (Newman & Whitehill, 1979). We could have added another controlled trial to the literature but consciously chose not to do so because most of what can be gained in knowledge—and, just as importantly, in clinical credibility—from randomized trials of OST has already been achieved. When a research area develops to this point, the next stage in our view should be hybrid designs that assess whether trial findings replicate under typical practice conditions. We concur with other scientists in the field (Concato, Shah, & Horwitz, 2000; Wells, 1999) that such effectiveness evaluation designs should be more broadly employed, particularly in those mental health treatment research areas where the evidence base consists primarily of studies conducted under ideal treatment conditions with poorer prognosis patients excluded.

Acknowledgments

This study was funded by Grants SUS99-026 and RCS04-141-01 to the first author from the Department of Veterans Affairs Health Services Research and Development (VA HSR&D) Service. Some of the results of this study were presented in a plenary session at the 2005 VA HSR&D Annual Meeting in Washington, DC. We thank Doyanne Horst and Jared Minkel for assisting with participant interviews, Alex H. S. Harris and Yesenia Luna for statistical consultation, our scientific advisory board for helpful guidance throughout the study, and the clinic staff members and patients for facilitating the evaluation.

References

- American Psychiatric Association. (1994). Position statement on methadone maintenance treatment. *American Journal of Psychiatry*, *151*, 792–794.
- Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*, *342*, 1887–1892.
- Department of Veterans Affairs/Department of Defense. (2004). Clinical practice guideline for the management of substance use disorders. Retrieved February 17, 2007, from http://www.oqp.med.va.gov/cpg/SUD/SUD_Base.htm.
- D'Aunno, T., & Vaughn, T. E. (1992). Variations in methadone treatment practices: Results from a national study. *Journal of the American Medical Association*, *267*, 253–282.

- Humphreys, K., Weingardt, K. R., Horst, D., Joshi, A. A., & Finney, J. W. (2005). Prevalence and predictors of research participant eligibility criteria in alcohol treatment outcome studies, 1970–1998. *Addiction, 100*, 1249–1257.
- Humphreys, K., & Weisner, C. (2000). Use of exclusion criteria in selecting research subjects and its effect on the generalizability of alcohol treatment outcome studies. *American Journal of Psychiatry, 157*, 588–594.
- Kazis, L. E., Ren, X. S., Lee, A., Skinner, K., Rogers, W., Clark, J., et al. (1999). Health status in VA patients: Results from the Veterans Health Study. *American Journal of Medical Quality, 14*, 28–38.
- Kernick, D. P. (1998). Lies, damned lies, and evidence-based medicine. *Lancet, 351*, 1824.
- Kyrouz, E. M., & Humphreys, K. (1997). Do health care workplaces affect treatment environments? *Journal of Community and Applied Social Psychology, 7*, 105–118.
- Ling, W., Charuvastra, C., Kaim, S. C., & Klett, C. J. (1976). Methadyl acetate and methadone as maintenance treatments for heroin addicts. A Veterans Administration cooperative study. *Archives of General Psychiatry, 33*, 709–720.
- McLellan, A. T. (1983). Patient characteristics associated with outcome. In J. R. Cooper, F. Altman, B. S. Brown, & D. Czechowicz (Eds.), *Research on the treatment of narcotic addiction: State of the art* (pp. 500–540). Rockville, MD: National Institute on Drug Abuse.
- McLellan, A. T., Arndt, I. O., Metzger, D. S., Woody, G. E., & O'Brien, C. P. (1993). The effects of psychosocial services in substance abuse treatment. *Journal of the American Medical Association, 269*, 1953–1959.
- Moos, R. H. (1994). *Work environment scale manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Moos, R. H., Federman, B., & Finney, J. W. (1999). *Outcomes monitoring for patients with substance use disorders: III. Cohort 2 patients' characteristics and treatment at baseline*. Palo Alto, CA: Program Evaluation and Resource Center.
- Newman, R. G., & Whitehill, W. B. (1979). Double-blind comparison of methadone and placebo maintenance treatments of narcotic addicts in Hong Kong. *Lancet, 314*, 485–488.
- Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2000). *nlme: Linear and nonlinear mixed effects models. R package version 3* (pp. 1–56).
- R Development Core Team. (2005). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved June 15, 2005, from <http://www.R-project.org>.
- Strain, E. C., Bigelow, G. E., Liebson, I. A., & Stitzer, M. L. (1999). Moderate- versus high-dose methadone in the treatment of opioid dependence: A randomized trial. *Journal of the American Medical Association, 281*, 1000–1005.
- Wells, K. B. (1999). Treatment research at the crossroads: The scientific interface of clinical trials and effectiveness research. *American Journal of Psychiatry, 156*, 5–10.