

Published 2014. This article is a U.S. Government work and is in the public domain in the USA
DOI: 10.1111/1475-6773.12247
BEST OF THE 2014 ACADEMYHEALTH ANNUAL RESEARCH MEETING

Population-Level Cost-Effectiveness of Implementing Evidence-Based Practices into Routine Care

John C. Fortney, Jeffrey M. Pyne, and James F. Burgess, Jr.

Objective. The objective of this research was to apply a new methodology (population-level cost-effectiveness analysis) to determine the value of implementing an evidence-based practice in routine care.

Data Sources/Study Setting. Data are from sequentially conducted studies: a randomized controlled trial and an implementation trial of collaborative care for depression. Both trials were conducted in the same practice setting and population (primary care patients prescribed antidepressants).

Study Design. The study combined results from a randomized controlled trial and a pre-post-quasi-experimental implementation trial.

Data Collection/Extraction Methods. The randomized controlled trial collected quality-adjusted life years (QALYs) from survey and medication possession ratios (MPRs) from administrative data. The implementation trial collected MPRs and intervention costs from administrative data and implementation costs from survey.

Principal Findings. In the randomized controlled trial, MPRs were significantly correlated with QALYs ($p = .03$). In the implementation trial, patients at implementation sites had significantly higher MPRs ($p = .01$) than patients at control sites, and by extrapolation higher QALYs (0.00188). Total costs (implementation, intervention) were nonsignificantly higher (\$63.76) at implementation sites. The incremental population-level cost-effectiveness ratio was \$33,905.92/QALY (bootstrap interquartile range -\$45,343.10/QALY to \$99,260.90/QALY).

Conclusions. The methodology was feasible to operationalize and gave reasonable estimates of implementation value.

Key Words. Cost-effectiveness, implementation, depression

Determining the value of deploying evidence-based practices (EBPs) into routine clinical care is critical to policy makers, but cost-effectiveness is rarely assessed in implementation trials. An intervention demonstrated to be cost-effective in a randomized controlled trial (RCT) may not be cost-effective

when deployed into routine care if implementation costs are high or if population-level effectiveness is low. Population-level effectiveness in implementation trials will be lower than the efficacy/effectiveness observed in the RCTs that established the evidence base if the EBP does not reach a substantial portion of the targeted patient population or if poor fidelity to inclusion/exclusion criteria and treatment protocols compromise clinical effectiveness during roll-out (Zatzick, Koepsell, and Rivara 2009; Koepsell, Zatzick, and Rivara 2011).

Traditional cost-effectiveness analysis (CEA) is not well suited for implementation trials. In RCTs, cost-effectiveness represents the ratio of differences in costs and effectiveness between two groups of patients who have been sampled, recruited, deemed eligible, and then consented. This enrollment process artificially inflates reach into the targeted patient population by excluding individuals who are unavailable, unwilling, or uninterested in participating. In addition, both reach and fidelity are controlled during the RCT in artificial and nonreplicable ways. In *completers* data analyses, reach and fidelity are further inflated by excluding patients from the analytical sample who did not initiate and complete the treatment protocol. As a result, RCT findings often lack external validity and, for this reason, implementation trials are needed to inform policy makers about the value of deploying EBPs in routine care.

In contrast to RCTs, implementation trials allocate providers or practices to either a control condition or an implementation strategy designed to increase the likelihood that patients receive an EBP with high fidelity. Reach and fidelity are specified as outcomes rather than controlled artificially (Glasgow, Vogt, and Boles 1999). Currently, there is no consensus about how to conduct economic evaluations for implementation trials (Smith and Barnett 2008). To date, four alternative cost evaluation methods have been proposed (see Table 1): (1) trial-based CEA, (2) policy CEA, (3) budget impact analysis, and (4) systems-level CEA. Most of these methods have not been widely applied in implementation trials and each has methodological weaknesses (Smith and Barnett 2008).

Address correspondence to John C. Fortney, Ph.D., Health Services Research and Development, Central Arkansas Veterans Healthcare System, (152/NLR), North Little Rock, AR 72114; e-mail: fortneyj@uw.edu. Jeffrey M. Pyne, M.D., is with the Health Services Research and Development, Central Arkansas Veterans Healthcare System, North Little Rock, AR. John C. Fortney, Ph.D., and Jeffrey M. Pyne, are with the South Central Mental Illness Research Education and Clinical Center, Central Arkansas Veterans Healthcare System, Little Rock, AR. Department of Psychiatry, University of Arkansas for Medical Sciences, North Little Rock, AR. James F. Burgess, Jr., Ph.D., is with the Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System; Department of Health Policy and Management, School of Public Health, Boston University, Boston, MA.

Trial-based CEA is used for implementation trials that use traditional clinical trial methods for patient enrollment (sampling, recruitment, eligibility assessment, consenting) and primary data collection. The analysis is conducted on a *sample* of the targeted population (Schoenbaum et al. 2001; Dijkstra et al. 2006; Chaney et al. 2011). Both the cost of the EBP and the cost of the implementation strategy used to promote the adoption of the EBP are measured. Despite the obvious strengths with regard to internal validity, there are two main weaknesses to this approach. First, these types of implementation trials are extremely expensive to conduct, especially if low levels of reach and fidelity reduce population-level effectiveness, thus necessitating the recruitment of large samples. Second, the process of recruiting and consenting is likely to increase reach by artificially excluding those patients who are unavailable, unwilling, or uninterested. Thus, this approach may generate biased population-level cost-effectiveness ratios (CERs).

Policy CEA combines the results of a previous RCT with regard to intervention cost and effectiveness with the results of an implementation trial with regard to the cost and effectiveness of the implementation strategy (Mason et al. 2001). Using algebraic formulas, intervention costs and implementation costs are summed, and intervention effectiveness (clinical) and implementation effectiveness (adoption/reach) are combined. This approach has the advantage that intervention costs and clinical effectiveness do not need to be collected during the implementation trial. The resulting weakness is that it assumes that the cost and effectiveness of the intervention (observed in the RCT) will remain unchanged when implemented as an EBP in routine care (Smith and Barnett 2008). This assumption will not be valid when the EBP is implemented with poor fidelity and/or is delivered to patients who would not have been eligible for the RCT. Thus, this approach may generate biased population-level CERs.

Budget impact analysis and systems-level CEA use decision analytic modeling to estimate how the adoption of an EBP would impact the allocation of treatments to patients in routine care. Budget impact analysis predicts how costs would be distributed across various stakeholders (Mauskopf et al. 2007; Fortney et al. 2011; Humphreys, Wagner, and Gage 2011; Anaya et al. 2012; Gidwani et al. 2012). It estimates the size of the population reached by the EBP, and the effect of implementation on costs, but not clinical outcomes. Systems-level CEA predicts both costs and clinical outcomes by having an expert panel estimate the decline in the clinical effectiveness of the intervention when it is delivered as an EBP in routine care (Frank et al. 1999; Alegria,

Table 1: Description of Cost-Effectiveness Methodologies for Implementation Research

<i>C EA Type</i>	<i>Enrollment</i>	<i>Generalizable to Population</i>	<i>Reach</i>	<i>Effectiveness</i>	<i>Intervention Costs</i>	<i>Implementation Costs</i>	<i>Weaknesses</i>
RCTCEA	Sampled Recruited Consented Assessed for eligibility	Depends on eligibility criteria	Controlled	Evaluated using primary data collection	Evaluated using primary data collection	Not evaluated	Reach is not evaluated Implementation costs are not evaluated Sample results may not generalize to population
TBCEA	Sampled Recruited Consented Assessed for eligibility	Depends on eligibility criteria	Evaluated using primary data collection	Evaluated using primary data collection	Evaluated using primary data collection	Evaluated using primary data collection	Expensive Sample results may not generalize to population
PCEA	Population identified from archival data	Yes	Evaluated using archival data	Modeled from literature	Modeled from literature	Evaluated using primary data collection	Intervention cost and effectiveness observed in RCT may not be the same in implementation trial Effectiveness is not evaluated
BIA	Population identified from archival data	Yes	Modeled from literature	Not evaluated	Modeled from literature	Modeled from literature	Implementation costs are not evaluated
SILCEA	Population identified from archival data	Yes	Modeled from literature	Estimated by expert opinion	Modeled from literature	Not evaluated	Expert opinion about effectiveness may not be accurate
PLCEA	Population identified from archival data	Yes	Evaluated using archival data	Evaluated using archival data using a proxy measure of clinical outcomes modeled from literature	Evaluated using archival data	Evaluated using primary data collection	Proxy measure of clinical outcomes may not be accurate

BIA, budget impact analysis; PCEA, policy cost-effectiveness; PLCEA, population-level cost-effectiveness analysis; RCTCEA, traditional randomized controlled trial cost-effectiveness analysis; SILCEA, systems-level cost-effectiveness analysis; TBCEA, trial-based cost-effectiveness analysis.

Frank, and McGuire 2005). While this approach is probably more accurate than the policy CEA assumption that effectiveness will not be diminished during implementation, a weakness is that validity depends on the accuracy of the expert panel.

Building on the strengths of these previous approaches, while attempting to avoid their limitations, we developed an alternative approach to conducting economic evaluations for implementation trials. Population-level cost-effectiveness analysis (PLCEA) is defined as *the ratio of incremental population-level costs and incremental population-level effectiveness*. *Incremental population-level costs* represent the differences in intervention and implementation costs between the patient population targeted by an EBP implementation strategy and an equivalent untargeted population. *Incremental population-level effectiveness* represents the differences in clinical outcomes between the targeted patient population and an equivalent untargeted population. The targeted population and the equivalent untargeted population represent those patients that would benefit clinically from receiving the EBP. However, PLCEA is designed for implementation trials that do not sample, recruit, and consent patients, and consequently, it will not be feasible to apply strict eligibility criteria based on structured clinical assessments. As a result, the analytical cohort may have to be broadly and imprecisely defined. This is similar to CEAs in public health where it is often difficult to define the population exposed to programs that have fuzzy temporal/geographic boundaries such as community-based health promotion (Gold et al. 2007). In addition, PLCEA is designed for implementation trials that minimize the use of primary data collection methods. As a consequence, it is necessary to estimate reach, clinical effectiveness, and intervention costs for all targeted (and equivalent untargeted) patients using archival/administrative data. To measure clinical effectiveness, it often will be necessary to identify a surrogate measure that is available in archival/administrative data (e.g., claims, electronic health records) and significantly correlated with clinical outcomes (Mason et al. 2005). An example of such a “proxy” is the correlation between HbA1c levels (available via laboratory records) and diabetic complications (Dijkstra et al. 2006). Ideally, the proxy measure should be correlated with quality-adjusted life years (QALYs), the recommended unit of measure for CEAs focused on chronic illnesses (Gold 1996).

To illustrate the use of PLCEA for an implementation trial, we applied this method to the rollout of telemedicine-based collaborative care for depression in VA community-based outpatient clinics. This EBP focuses on improving antidepressant adherence and was shown to be effective in an RCT (Fortney et al. 2007). We subsequently conducted an implementation trial to

test an implementation strategy which was designed to promote the deployment of this EBP into routine care (Fortney et al. 2012). Using the significant correlation between antidepressant medication possession ratio (MPR) and treatment response (50 percent reduction in severity) observed in the RCT (Fortney et al. 2010), we previously have described the population-level effectiveness of telemedicine-based collaborative care for all patients prescribed antidepressants (Fortney et al. 2013). For the current analyses, we use the RCT data to estimate the correlation between antidepressant MPRs and QALYs and use this proxy, along with data collected during the implementation trial, to conduct the PLCEA.

METHODS

Overview

Conducted from the payer's perspective, the PLCEA combined data from an RCT (Fortney et al. 2007) and an implementation trial (Fortney et al. 2012). Both were conducted in the same primary care practice setting (VA Community-Based Outpatient Clinics) and patient population (veterans with depression prescribed antidepressants). The RCT used an experimental design and demonstrated that the EBP is clinically effective and cost-effective (Fortney et al. 2007; Pyne et al. 2010). The implementation trial used a quasi-experimental design and demonstrated that the implementation strategy promoted adoption, reach, and fidelity (Fortney et al. 2012).

The PLCEA was conducted in eight steps. First, we used pharmacy data and patient survey data from the RCT to create a proxy measure for QALYs using MPRs. Second, we used pharmacy data from the implementation trial to estimate group differences in MPRs between implementation and control sites. Third, using the proxy from step one, we converted group differences in MPRs observed in the implementation trial into group differences in QALYs. Fourth, we estimated the implementation costs using survey data collected from the providers and administrators who were part of the implementation team. Fifth, we estimated the intervention costs using administrative data. Sixth, we summed the intervention costs and the implementation costs. Seventh, we calculated the incremental CER (ICER) by dividing group differences in total cost (intervention and implementation) by group differences in QALYs. Eighth, we conducted a bootstrap analysis to assess ICER variation and conducted sensitivity analyses.

Randomized Controlled Trial. The RCT was conducted in VA community-based outpatient clinics, which are satellite clinics of VA Medical Centers. Seven clinics were matched by VA Medical Center and staffing (lack of on-site psychiatrists), and one clinic within each group was randomized to the intervention. We screened patients for depression and enrolled 395 patients who screened positive. We excluded patients with serious mental illness (e.g., schizophrenia, bipolar disorder) and those receiving specialty mental health services (e.g., psychiatrist and psychologist visits). The intervention was designed to optimize antidepressant treatment. Primary care providers located at clinics were supported by an off-site depression care team (nurse care manager, pharmacist, and consulting tele-psychiatrist). Care manager activities included symptom monitoring, medication adherence monitoring/promotion, and side-effects monitoring/management. The tele-psychiatrist made prescribing recommendations to the primary care provider for patients failing antidepressant trials.

At baseline, the following case mix factors were collected: sociodemographics, health beliefs, depression history, Short Form 12V (a measure of physical and mental health functioning adapted for veterans), and psychiatric comorbidity. In addition, ZIP code was used to categorize patients as rural or urban. At baseline and 6-month follow-up ($n = 360$), we administered the Hopkins Symptom Checklist (Derogatis et al. 1974) to assess depression severity (range 0–4), the primary effectiveness measure of the RCT. In CEA, effectiveness is measured using QALYs, which are a non-disease-specific outcome measure that facilitates comparing the cost-effectiveness of interventions that target different disorders. Measuring QALYs assumes that a year of life lived in perfect health is worth 1.0 QALY, death is worth 0.0, and a year of life lived in a state of less than perfect health is between 0.0 and 1.0. Following validated methods, baseline and follow-up depression severity scores were converted into depression-free days (DFDs) using scores <0.5 representing all DFDs and scores >1.7 representing zero DFDs (Pyne et al. 2007). DFDs were then converted into QALYs using validated methods (Pyne et al. 2007). This conversion was based on the assumption that a nondepressed person has a utility score of 1.0 and a severely depressed person has a utility score of 0.6. To calculate QALYs for the main reference case analysis, we multiplied DFDs by 0.4 (representing an improvement from severely depressed to nondepressed) and divided by 182.5 (6 months) and added 0.6.

Using pharmacy data obtained from VA administrative data, the prorated days supply of nonconcurrent antidepressant prescription fills were summed to calculate the total cumulative days' supply during the 6 months.

For multiple concurrent prescriptions of the same antidepressant, the days' supply of each prescription fill was summed to calculate cumulative days' supply. For multiple concurrent prescriptions of different antidepressants, the days' supply of the earlier prescription was truncated (to account for possible switching). MPR was calculated by dividing the cumulative days' supply by 182.5 days. Patients were classified as adherent if their $MPR \geq 0.9$.

Statistical Analysis. We specified QALYs as the dependent variable in a linear regression, with $MPR \geq 0.9$ as the explanatory variable and baseline case mix variables as covariates. Only those variables found to significantly predict the dependent variable at the $p < .2$ level in bivariate analyses were included in the multivariate analyses. Intervention status was included to account for possible correlation between MPR and other potential mechanisms of action of the intervention. The intraclass coefficients were close to zero at the clinic level (0.015) and the Medical Center level (0.004) with respect to changes in depression severity, so we did not adjust standard errors for clustering. Independent variables with missing values were imputed using the MI and MIANALYZE procedures in SAS.

Implementation Trial

Implementation sites included 11 community-based outpatient clinics lacking on-site psychiatrists associated with three VA Medical Centers. The control group included 11 matched (on region, staffing, and preperiod MPR) clinics associated with seven other Medical Centers. The EBP that was implemented was the same as the intervention tested in the RCT (i.e., support and consultation from an off-site depression care team). However, eligibility for the EBP was determined by the primary care provider (via referral) rather than by a structured clinical assessment of depression severity as was done in the RCT.

Because the EBP focused on optimizing antidepressant treatment, we chose to define the analytical cohort as all patients diagnosed with depression and prescribed an antidepressant medication. This definition is likely to be an overly broad approximation of the true target population because many patients prescribed antidepressants are stable (i.e., asymptomatic) and would not have been good candidates for referral to the EBP. However, symptom severity was not available in VA administrative data and could not be used to further narrow the definition of the target population. Patient characteristics (age, gender, race, marital status, percent service connection, and zipcode)

and clinical data were extracted from VA administrative datasets. Zipcode was used to determine rurality and per capita income (per the US Census). MPR ≥ 0.9 was calculated using the same methods as the RCT.

The start date for each implementation site was the date the first patient enrolled in the EBP. The start date for control sites was the average of the implementation site start dates. During the 6 months after the start date at each site, we identified patients' first primary care encounters with a primary or secondary depression diagnosis. Patients were excluded from the analytical cohort if they had a specialty mental health encounter or a diagnosis of serious mental illness during the 6 months prior to the index visit or if they were not prescribed an antidepressant during the 6-month follow-up period.

The intervention and other medical costs were calculated using VA administrative data. The intervention cost was determined by summing the cost of all care manager encounters (identified by their verified provider IDs). Other medical costs were measured by summing all other VA outpatient encounters (including primary care, specialty mental health, specialty medical, ancillary) and VA dispensed medications. Depression-specific costs were defined as the cost of encounters with a primary or secondary depression diagnosis and antidepressant medications. Implementation costs represent the personnel resources associated with the implementation strategy and were calculated by multiplying the hourly wage of each of the 16 implementation team members by the number of meetings they attended plus their estimated hours per month (both pre- and poststart date) devoted to between-meeting implementation activities. The dependent variable for the incremental cost analysis was total cost, defined as the sum of implementation costs, intervention costs, and other medical costs.

Incremental Cost Analysis. Because of their nonnormal distributions, Kruskal–Wallis nonparametric tests were used to compare intervention and other medical costs for patients at the implementation and control sites across specific categories (e.g., primary care, pharmacy). The total cost regression included all the patient characteristics available in the administrative data as covariates: age, gender, marital status, service connection, per capita income, and rurality. In addition, for the 6-month preperiod, we used the methods described above to identify a cohort of patients diagnosed with depression and prescribed antidepressants at implementation and control sites, and created an additional covariate representing the average medical costs at each clinic in the preperiod. To identify the best distribution and link function for the regression

analysis, we specified generalized linear models (PROC GLIMMIX SAS Enterprise Guide 5.1, SAS Inc., Cary, NC, USA) with distributions (normal, gamma, or inverse Gaussian) and link functions (identity, log, or square root) to identify which distribution and link function combination had the smallest Akaike Information Criterion and Bayesian Information Criterion values. The combination of the inverse Gaussian distribution with the square root link had the smallest criterion values. To control for the clustering of patients within sites, we used PROC GENMOD (SAS Enterprise Guide 5.1) with the inverse Gaussian distribution and the square root link for the final total cost regression analysis. We then used the results of this regression analysis to generate the marginal effect for total costs. To generate the marginal effect, we used the parameter estimates from the regression to calculate two cost predictions for each patient. The first prediction was based on the assumption that the patient was treated at an implementation site, and the second prediction was based on the assumption that the patient was treated at a control site. We then averaged the difference between the two predicted values for each patient across all patients to generate an overall marginal effect representing incremental cost.

Incremental QALY Analysis. As previously reported, a higher percentage of patients at implementation sites had MPR ≥ 0.9 compared to patients at control sites (42.5 percent vs. 30.9 percent), which was significant after controlling for differences in patient case mix and preperiod MPR (OR = 1.38, CI₉₅ = [1.07, 1.78], $p = .01$) (Fortney et al. 2013). As previously reported, the marginal effect of being treated at implementation sites compared to control sites represents a 7.35 percent point increase in MPR ≥ 0.9 (Fortney et al. 2013). For this analysis, we converted this previously reported marginal effect for MPR ≥ 0.9 estimated from the implementation trial data into a marginal effect for QALY by multiplying the 7.35 percent by the MPR \rightarrow QALY proxy (i.e., MPR ≥ 0.9 parameter coefficient taken from the regression equation predicting QALYs estimated from RCT data).

Incremental Cost-Effectiveness Ratio. The ICER was calculated by dividing the marginal effect for total cost by the marginal effect for QALYs. To assess the variation in the ICER, we used nonparametric bootstrapping with replacement and 1,000 replications to estimate the joint distributions of incremental total costs and QALYs. Because the incremental effectiveness (denominator) was close to zero for many of the bootstrapped samples, there were some

extreme positive and negative ratios and the ICER distribution was characterized by high kurtosis (441.63). Therefore, the median ICER rather than the mean is reported.

Sensitivity Analyses. Two sensitivity analyses were conducted. The first sensitivity analysis used an alternative assumption about the improvement from moderately depressed to nondepressed for the DFD-to-QALY conversion. Specifically, we multiplied DFDs by 0.2 (instead of 0.4) and divided by 182.5 and added 0.8 (instead of 0.6) to calculate QALYs. The second probabilistic sensitivity analysis (Briggs 2005; Claxton et al. 2005) was conducted to assess the added variation in QALYs resulting from the prediction error associated with the MPR \rightarrow QALY proxy. For each of the 1,000 bootstrapped samples, we sampled 1,000 random numbers from the normal distribution $(N(0,1))$, multiplied the random number by the standard error of the parameter estimate ($se[\beta]$) and added the parameter estimate (β) (i.e., $\beta + se[\beta] * N(0,1)$). We chose to use one standard error because it is conceptually similar to the interquartile range we used for the median ICER.

RESULTS

There were 360 patients enrolled in the RCT sample and 1,558 patients in the implementation trial cohort. In the RCT, there were no significant differences in case mix between intervention and control patients (Fortney et al. 2007). Compared to control sites in the implementation trial, patients at implementation sites were significantly ($\chi^2 = 15.0, p < .01$) and substantially less rural (27.8 percent vs. 37.7 percent) and were statistically ($t = -2.09, p = .04$), but not substantially older (1.7 years) (Fortney et al. 2013). The patients in the RCT sample and implementation trial cohort were similar with respect to variables measured in both trials with the exception of rurality (51.4 percent in the RCT and 30.7 percent in the implementation trial).

Randomized Controlled Trial Findings

Table 2 provides the results of the regression with QALYs specified as the dependent variable and $MPR \geq 0.9$ as the explanatory variable. As hypothesized, $MPR \geq 0.9$ was significantly and positively correlated with QALYs defined using the 0.4 multiplier ($\beta = 0.0256, CI_{95} = [0.0029-0.0483]$,

Table 2: QALYs Regression Results from the RCT ($N = 360$ Patients)*

Variable	Parameter Estimate	Parameter Estimate 95% Confidence Interval	<i>p</i> -value
MPR ≥ 0.9	0.0256	[0.0029–0.0483]	.0274
Intervention	0.0097	[–0.0090 to 0.0285]	.3088
Socioeconomic			
Age	–0.0006	[–0.0015 to 0.0004]	.2303
Income category	0.0027	[–0.0031 to 0.0084]	.3625
Married	0.0080	[–0.0133 to 0.0293]	.4639
Social support (0–1)	0.0208	[–0.0261 to 0.0686]	.3848
Perceived need for treatment (0–6)	–0.0092	[–0.0170 to –0.0014]	.0208
Perceived treatment effectiveness (0–2)	–0.0087	[–0.0220 to 0.0047]	.2019
Clinical			
PCS (physical component score)	0.0004	[–0.0005 to 0.0013]	.3865
MCS (mental component score)	0.0024	[0.0013–0.0035]	<.0001
Family history of depression	–0.0008	[–0.0207 to 0.0190]	.9358
Age of onset	0.0146	[0.0038 to 0.0253]	.0082
Prior depression episodes	–0.0025	[–0.0089 to 0.0039]	.4468
Prior depression treatment	–0.024	[–0.0497 to 0.0012]	.0621
Current depression treatment	0.0133	[–0.0118 to 0.0383]	.2990
Antidepressants acceptable	–0.0067	[–0.0197 to 0.0062]	.3066
Current major depressive disorder	–0.0455	[–0.0764 to –0.0147]	.0038
Current dysthymia	–0.0192	[–0.0727 to 0.0342]	.4805
Current panic disorder	0.1274	[–0.2042 to 0.4589]	.3797
Current generalized anxiety disorder	–0.0410	[–0.0638 to –0.0183]	.0005
Current posttraumatic stress disorder	–0.0211	[–0.0447 to 0.0026]	.0804

*The RCT is described in more detail in Fortney et al. (2007).

$p = .03$), indicating that those with a MPR ≥ 0.9 have about 0.03 greater QALYs. For the first sensitivity analysis, MPR ≥ 0.9 was significantly and positively correlated with QALYs defined using the 0.2 multiplier ($\beta = 0.0128$, $CI_{95} = [0.0014–0.0242]$, $p = .03$). Other significant predictors included baseline depression severity, comorbid generalized anxiety disorder, and age of depression onset.

Implementation Trial Findings

QALY Results. As previously reported, only 10.3 percent ($n = 116$) of the cohort at implementation sites had an encounter with a care manager documented in the administrative data in the 6 months after the index visit (Fortney et al. 2013). Multiplying the previously reported marginal effect for MPR (0.0735) obtained from the implementation trial (Fortney et al. 2013) by the MPR \rightarrow QALY proxy obtained from the RCT (0.0256) reported above

yielded an incremental QALY of 0.00188 (CI₉₅ = 0.0002–0.0306). For the sensitivity analysis, multiplying the marginal effect for MPR (0.0735) by the MPR-QALY proxy (0.0128) yielded an incremental QALY of 0.00094 (CI₉₅ = 0.0001–0.0018).

Implementation Costs Results. The total cost of attending implementation meetings was \$8,270.90 and total between-meeting cost was \$84,483.45. The total implementation cost was \$92,753.79. Dividing total implementation cost by the 1,132 patients with a depression diagnosis and prescribed an antidepressant at implementation sites yielded an implementation cost per targeted patient of \$81.93.

Intervention Cost Results. Table 3 displays the intervention costs. Kruskal–Wallis nonparametric tests indicated that intervention costs were significantly ($p < .0001$) higher at implementation sites ($\mu = \$107$, $SD = 459$) compared to

Table 3: Implementation, Intervention, and Other Medical Costs in Implementation Trial ($N = 1,558$ patients)*

Dependent variable	Control Clinics Mean (SD) $N = 1,132$	Implementation Clinics Mean (SD) $N = 456$	p -value [†]
Total costs	3,884.81 (3,332.48)	3,925.36 (4,357.41)	.4864
Implementation cost			
Implementation cost per patient	0.00 (0.00)	83.91 (0.00)	NA
Direct intervention costs			
Care management	0.00 (0.00)	106.67 (459.32)	<.0001
Other medical costs			
Pharmacy	1,212.08 (1,302.17)	1,269.18 (1,449.65)	.3688
Total outpatient	2,672.72 (2,587.14)	2,574.25 (3,635.35)	.0251
Primary care	552.83 (669.16)	427.66 (317.72)	<.0001
Medical specialty	541.64 (1,161.88)	416.73 (1,341.67)	.0242
Mental health excluding care management	287.97 (611.24)	228.28 (612.76)	.0163
All other outpatient and ancillary	519.98 (956.40)	488.06 (1,347.36)	.0652
Other depression-specific costs			
Pharmacy	252.18 (508.31)	224.81 (307.37)	.4452
Primary care	342.41 (549.04)	305.21 (192.28)	.2918
Mental health excluding care management	157.74 (460.96)	135.24 (375.99)	.3905

*The Implementation Trial is described in more detail in Fortney et al. (2013).

[†]Kruskal–Wallis test.

control sites that had no care manager encounters. Total outpatient costs were significantly ($p = .03$) lower at implementation sites ($\mu = \$2,574$, $SD = 3,635$) than control sites ($\mu = \$2,673$, $SD = 2,587$). There were no significant between-group differences in pharmacy or depression-specific costs. Table 4 reports the results of the multivariate total cost regression. There was not a significant difference in total costs between implementation and control sites, despite the higher implementation cost and intervention costs at implementation sites. The estimated marginal effect for total costs was \$63.76. Age and rurality were significant negative predictors of total cost, while percent service connection >50 percent and average medical costs at each site in the preperiod were significant positive predictors.

Incremental CER Results. The incremental cost (\$63.76) divided by the incremental effectiveness in the main analysis (0.00188 QALYs) yields a population-level ICER of \$33,905.92/QALY. The median ICER from the 1,000 bootstrapped samples was \$29,297.00/QALY (interquartile range -\$45,343.10/QALY to \$99,260.90/QALY). Figure 1 displays the cost-effectiveness plane. The acceptability curve indicated that there is a 0.392 probability that the ICER represents cost savings and a 0.575 probability that the ratio is below \$50,000/QALY. For the sensitivity analysis that used the alternative DFD-to-QALY conversion formula, the ICER is \$67,809.19/QALY for the

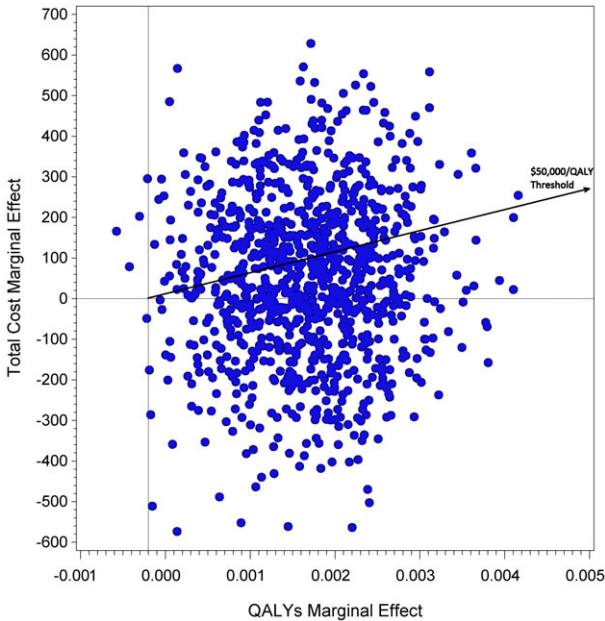
Table 4: Total Cost Regression Results with Inverse Gaussian Distribution and the Square Root Link from the Implementation Trial* ($N = 1,558$ patients)

Variable	Parameter Estimate	Parameter Estimate 95% Confidence Interval	p-value
Implementation site	0.5118	[-2.7778, 3.8013]	.7604
Preperiod total cost at site	0.0031	[0.0013, 0.0048]	.0007
Age	-0.1286	[-0.2328, -0.0244]	.0155
Male	5.8835	[-0.3469, 12.1138]	.0642
Married	-2.8296	[-6.7176, 1.0584]	.1537
Rural	-3.6195	[-6.1199, -1.1192]	.0045
Income	0.0899	[-0.1348, 0.3145]	.4330
Service connection 1%–49% [†]	-1.7729	[-4.6527, 1.1070]	.2276
Service connection ≥50% [†]	9.7574	[6.3844, 13.1304]	.0045

*The Implementation Trial is described in more detail in Fortney et al. (2013).

[†]Percent service connection represents the degree to which health problems caused or exacerbated by military service prevent the veteran from working.

Figure 1: Cost-Effectiveness Plane from Bootstrapped Sample



original sample and the median ICER from the bootstrapped samples was \$58,591.70/QALY. For the probabilistic sensitivity analysis of the MPR → QALY proxy, the median ICER from the 1,000 bootstrapped samples and the 1,000 sampled proxy values was \$27,879.00/QALY (interquartile range −\$47,010.70/QALY to \$111,603.00/QALY), and the acceptability curve indicated that there is a 0.395 probability that the ICER represented cost savings.

DISCUSSION

In the implementation trial, the other medical costs (\$3,885) were similar to the other medical costs in the RCT (\$4,250) (Fortney et al. 2011). Likewise, patients in the two trials were similar with regard to characteristics available in the administrative data. The similarity of the patients in the two trials supports our use of the MPR → QALY proxy derived from the RCT in the estimation of QALYs in the implementation trial.

When distributed across the entire targeted patient population, the per-patient implementation and intervention costs in the implementation trial

were relatively low. Overall, the difference in total costs across implementation and control sites was small and not statistically significant. Similar to other studies of population-level effectiveness (Finkelstein et al. 2005), the impact of EBP on the targeted patient population was small (0.00188 QALYs). An important factor to consider when interpreting these population-level results is that the analytical cohort was an overly broad approximation of the patient population actually targeted by the EBP. This likely contributed to the finding that only 10 percent of patients in the analytical cohort were reached by EBP. Moreover, it is likely that the clinical effectiveness of EBP was higher for the subset of patients actually targeted by the EBP compared to the entire analytical cohort, which may have included some asymptomatic patients. While this diluted the population-level effectiveness estimate by some unknown proportion, it also diluted the population-level cost estimate by the *same* unknown proportion, and thus the ICER was not biased by this “dilution effect.”

While both the incremental costs and incremental effectiveness were close to zero, the ICER of \$33,906 is below the \$50,000 threshold commonly used to determine whether an intervention is cost-effective in a RCT. While the high degree of variation in the ICER necessitates cautious interpretation, the policy implication of this finding is that EBP is cost-effective when implemented in routine care. Interestingly, the population-level ICER observed in the implementation trial was lower than the ICER observed in the RCT (\$90,320, adjusted to 2007 dollars) despite the additional costs associated with implementation in the implementation trial. This may be because primary care providers in the implementation trial were more likely to refer patients that benefitted from EBP than the enrollment procedures (i.e., structured clinical assessment) used in the RCT.

An important methodological question is whether CEA thresholds used for RCTs are directly applicable to PLCEA findings from implementation trials. One factor to consider is how reach is related to economies of scale (i.e., per-patient costs may decrease as patient volume increases). If there are substantial economies of scale, lower PLCEA thresholds should be used for implementation trials with high levels of reach into the target population. On the other hand, incremental costs in PLCEA include both the intervention and implementation costs. In this study, implementation costs per patient (\$84) were nearly as large as the intervention cost per patient (\$107), suggesting that higher thresholds should be used for PLCEA. Clearly, more thought is needed to establish meaningful PLCEA thresholds for different implementation contexts.

Another methodological issue worth noting is that because many of the incremental effectiveness estimates from the bootstrapped samples were very close to zero, these small denominators yielded some very extreme positive and negative ICERs. As a result, the distribution of the population-level ICERs from the bootstrapped samples was characterized by high kurtosis. Consequently, the acceptability curve was more linear than what is typically observed in traditional CEA. We anticipate that this finding will be replicated in future applications of the PLCEA methodology and that it is not an artifact of this particular dataset. It also is worth noting that the probabilistic sensitivity analysis accounting for prediction error in the proxy did not substantially increase the interquartile range of the ICER.

A limitation of this research is that implementation costs were not measured as rigorously as they have been in some previous studies, such as Liu et al. (2009). Another limitation may be the external/internal validity of the MPR \rightarrow QALY proxy. Despite the similarities of the practice setting and patient population, there may have been differences in sample characteristics that limited the generalizability of the RCT findings to the implementation trial (i.e., external validity). While all patients in the RCT sample were symptomatic at baseline, the depression severity of patients in the implementation trial cohort was not observed, and some could have been asymptomatic. Future applications of PLCEA methodology will need to consider the generalizability of the sample used to estimate the proxy to the population in which the proxy is applied. A potential threat to internal validity is that the MPR \rightarrow QALY proxy was estimated from observational data collected during the RCT and therefore subject to endogeneity bias. In addition, the relationship between MPRs and QALYs may not be linear. Another limitation is that if the EBP had other mechanisms of action besides improving MPR, the application of the proxy may have underestimated the effect of MPR on QALYs.

Compared to alternative methods for conducting economic evaluations of implementation trials, the proposed new PLCEA methodology has several important strengths. First, PLCEA empirically estimates the impact of implementation on QALYs rather than relying on expert opinion or the assumption that RCT efficacy/effectiveness results will be replicated in routine care. With recent advancements in comparative effectiveness research, patient-centered outcomes may become increasingly available in archival/administrative datasets, further facilitating QALY estimation. Second, reach, effectiveness, and intervention costs of the EBP are calculated for the entire target population using archival/administrative data. This maximizes external validity (compared to the sampling, recruiting, and consenting approach of trial-based

CEA) and minimizes research expenses. On the basis of these strengths, we believe that PLCEA represents an efficient method for generating relevant findings for policy makers.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was supported by VA Quality Enhancement Research Initiative (QUERI) IMV 04-360 grant to Dr. Fortney, VA IIR 00-078-3 grant to Dr. Fortney, VA NPI-01-006-1 grant to Dr. Pyne, the VA HSR&D Center for Mental Healthcare and Outcomes Research, and the VA South Central Mental Illness Research Education and Clinical Center. The clinical trials numbers are NCT00105690 and NCT00317018.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Alegria, M., R. Frank, and T. McGuire. 2005. "Managed Care and Systems Cost-Effectiveness: Treatment for Depression." *Medical Care* 43 (12): 1225–33.
- Anaya, H. D., K. Chan, U. Karmarkar, S. M. Asch, and G. M. Bidwell. 2012. "Budget Impact Analysis of HIV Testing in the VA Healthcare System." *Value Health* 15 (8): 1022–8.
- Briggs, A. 2005. "Probabilistic Analysis of Cost-Effectiveness Models: Statistical Representation of Parameter Uncertainty." *Value in Health* 8 (1): 1–2.
- Chaney, E. F., L. V. Rubenstein, C. F. Liu, E. M. Yano, C. Bolkan, M. Lee, B. Simon, A. Lanto, B. Felker, and J. Uman. 2011. "Implementing Collaborative Care for Depression Treatment in Primary Care: A Cluster Randomized Evaluation of a Quality Improvement Practice Redesign." *Implementation Science* 6: 121.
- Claxton, K., M. Sculpher, C. McCabe, A. Briggs, R. Akehurst, M. Buxton, J. Brazier, and T. O'Hagan. 2005. "Probabilistic Sensitivity Analysis for NICE Technology Assessment: Not an Optional Extra." *Health Economics* 14: 339–47.
- Derogatis, L. R., R. S. Lipman, K. Rickels, E. H. Uhlenhuth, and L. Covi. 1974. "The Hopkins Symptom Checklist (HSCL): A Measure of Primary Symptom Dimensions." *Pharmacopsychiatry* 7: 79–110.
- Dijkstra, R. F., L. W. Niessen, J. C. Braspenning, E. Adang, and R. T. Grol. 2006. "Patient-Centered and Professional-Directed Implementation Strategies for Diabetes Guidelines: A Cluster-Randomized Trial-Based Cost-Effectiveness Analysis." *Diabetic Medicine* 23 (2): 164–70.
- Finkelstein, J. A., P. Lozano, A. L. Fuhlbrigge, V. J. Carey, T. S. Inui, S. B. Soumerai, S. D. Sullivan, E. H. Wagner, E. H. Weiss, S. T. Weiss, and K. B. Weiss. 2005.

- “Population-Level Effects of Interventions to Improve Asthma Care in Primary Care Settings: The Pediatric Asthma Care Patient Outcomes Research Team.” *Health Services Research* 40 (6, Part I): 1737–57.
- Fortney, J. C., J. M. Pyne, M. J. Edlund, D. K. Williams, D. E. Robinson, D. Mittal, and K. L. Henderson. 2007. “A Randomized Trial of Telemedicine-Based Collaborative Care for Depression.” *Journal of General Internal Medicine* 22 (8): 1086–93.
- Fortney, J. C., J. M. Pyne, M. J. Edlund, and D. Mittal. 2010. “Relationship between Antidepressant Medication Possession and Treatment Response.” *General Hospital Psychiatry* 32 (4): 377–9.
- Fortney, J. C., M. Maciejewski, S. Tripathi, T. L. Deen, and J. M. Pyne. 2011. “A Budget Impact Analysis of Telemedicine-Based Collaborative Care for Depression.” *Medical Care* 49 (9): 872–80.
- Fortney, J., M. Enderle, S. McDougall, J. Clothier, J. Otero, L. Altman, and G. Curran. 2012. “Implementation Outcomes of Evidence-Based Quality Improvement for Depression in VA Community-Based Outpatient Clinics.” *Implementation Science* 7: 30.
- Fortney, J. C., M. A. Enderle, J. L. Clothier, J. M. Otero, J. S. Williams, and J. M. Pyne. 2013. “Population Level Effectiveness of Implementing Collaborative Care Management for Depression.” *General Hospital Psychiatry* 35: 455–60.
- Frank, R. G., T. G. McGuire, S. T. Normand, and H. H. Goldman. 1999. “The Value of Mental Health Care at the System Level: The Case of Treating Depression.” *Health Affairs* 18 (5): 71–88.
- Gidwani, R., M. B. Goetz, G. Kominski, S. Asch, K. Mattocks, J. H. Samet, A. Justice, N. Gandhi, and J. Needleman. 2012. “A Budget Impact Analysis of Rapid Human Immunodeficiency Virus Screening in Veterans Administration Emergency Departments.” *The Journal of Emergency Medicine* 42 (6): 719–26.
- Glasgow, R. E., T. M. Vogt, and S. M. Boles. 1999. “Evaluating the Public Health Impact of Health Promotion Interventions: The RE-AIM Framework.” *American Journal of Public Health* 89 (9): 1322–7.
- Gold, M. 1996. *Cost-Effectiveness in Health and Medicine: Report of the Panel on Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.
- Gold, L., A. Shiell, P. Hawe, T. Riley, B. Rankin, and P. Smithers. 2007. “The Costs of a Community-Based Intervention to Promote Maternal Health.” *Health Education Research* 22 (5): 648–57.
- Humphreys, K., T. H. Wagner, and M. Gage. 2011. “If Substance Use Disorder Treatment More Than Offsets Its Costs, Why Don’t More Medical Centers Want to Provide It? A Budget Impact Analysis in the Veterans Health Administration.” *Journal of Substance Abuse Treatment* 41 (3): 243–51.
- Koepsell, T. D., D. F. Zatzick, and F. P. Rivara. 2011. “Estimating the Population Impact of Preventive Interventions from Randomized Trials.” *American Journal of Preventive Medicine* 40 (2): 191–8.

- Liu, C. F., L. V. Rubenstein, J. E. Kirchner, J. C. Fortney, M. W. Perkins, S. K. Ober, J. M. Pyne, and E. F. Chaney. 2009. "Organizational Cost of Quality Improvement for Depression Care." *Health Services Research* 44 (1): 225–44.
- Mason, J., N. Freemantle, I. Nazareth, M. Eccles, A. Haines, and M. Drummond. 2001. "When Is It Cost-Effective to Change the Behavior of Health Professionals?" *Journal of the American Medical Association* 286 (23): 2988–92.
- Mason, J. M., N. Freemantle, J. M. Gibson, and J. P. New. 2005. "Specialist Nurse-Led Clinics to Improve Control of Hypertension and Hyperlipidemia in Diabetes: Economic Analysis of the SPLINT Trial." *Diabetes Care* 28 (1): 40–6.
- Mauskopf, J. A., S. D. Sullivan, L. Annemans, J. Caro, C. D. Mullins, M. Nuijten, E. Orlewska, J. Watkins, and P. Trueman. 2007. "Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices-Budget Impact Analysis." *Value in Health* 10 (5): 336–47.
- Pyne, J. M., S. Tripathi, D. K. Williams, and J. Fortney. 2007. "Depression-Free Day to Utility-Weighted Score: Is It Valid?" *Medical Care* 45 (4): 357–62.
- Pyne, J. M., J. C. Fortney, S. Tripathi, M. L. Maciejewski, M. J. Edlund, and D. K. Williams. 2010. "Cost-Effectiveness Analysis of a Rural Telemedicine Collaborative Care Intervention for Depression." *Archives of General Psychiatry* 67 (8): 812–21.
- Schoenbaum, M., J. Unutzer, C. Sherbourne, N. Duan, L. V. Rubenstein, J. Miranda, L. S. Meredith, M. F. Carney, and K. Wells. 2001. "Cost-Effectiveness of Practice-Initiated Quality Improvement for Depression: Results of a Randomized Controlled Trial." *Journal of the American Medical Association* 286 (11): 1325–30.
- Smith, M. W., and P. G. Barnett. 2008. "The Role of Economics in the QUERI Program: QUERI Series." *Implementation Science* 3 (20).
- Zatzick, D. F., T. Koepsell, and F. P. Rivara. 2009. "Using Target Population Specification, Effect Size, and Reach to Estimate and Compare the Population Impact of Two PTSD Preventive Interventions." *Psychiatry* 72 (4): 346–59.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.