Evaluating health care programs by combining cost with quality of life measures: a case study comparing capitation and Fee for Service (FFS)

Richard Grieve, Jasjeet S. Sekhon, Teh-wei Hu and Joan Bloom

Forthcoming, *Health Services Research*

Objective. To demonstrate cost-effectiveness analysis (CEA) for evaluating different reimbursement models.

Data Sources/Study Setting. The CEA used an observational study comparing Fee for Service (FFS) versus capitation for Medicaid cases with severe mental illness (n=522). Under capitation, services were either provided directly (direct capitation or DC) by not-for-profit community mental health centers (CMHC), or in a joint venture between CMHCs and a for-profit managed behavioral health organization (MBHO).

Study Design. A non-parametric matching method (Genetic Matching) was used to identify those cases that minimized baseline differences across the groups. Quality Adjusted Life years (QALYS) were reported for each group. Incremental QALYs were valued at different thresholds for a QALY gained, and combined with cost estimates to plot cost-effectiveness acceptability curves.

Principal findings. QALYS were similar across reimbursement models. Compared to FFS, the MBHO model had incremental costs of -$1991 and the probability this model was cost-effective exceeded 0.90. The DC model had incremental costs of $4694; the probability that this model was cost-effective compared to FFS was less than 0.10.

Conclusions. A capitation model with a for profit element was more cost-effective for Medicaid patients with severe mental illness than not-for-profit capitation or FFS models.

Key words: cost-effectiveness analysis, managed care organizations, QALY

BACKGROUND

Many European countries, Australia and Canada use cost-effectiveness analysis (CEA) to decide which health technologies to provide (Hutton and Maynard 2000). However, in the United States policy makers do not routinely use CEA to set health care priorities (Neumann 2004). Commentators have suggested that methodological flaws in published economic evaluations may impede their use in decision-making (Rennie and Luft 2000). One concern is that many published studies still use the cost-consequence approach and report costs and effectiveness separately (OHE 2005). These partial evaluations do not provide decision makers with information on any tradeoffs between costs and outcomes. Methods are available that provide cost-effectiveness estimates appropriate for use in policy-making (NICE 2004).

We illustrate how appropriate CEA techniques can be applied to evaluate a health service intervention using a case study comparing reimbursement models for mental health care. The paper uses a new technique, Genetic Matching, to adjust for baseline differences in patient mix across the intervention groups (Sekhon 2007). Genetic Matching is more appropriate than alternatives such as model-based adjustment, as it does not rely on parametric assumptions that are implausible in this context. The paper reports outcomes using QALYs, as these can recognize the effect of the reimbursement model on both length and quality of life. The paper uses recommended methods for dealing with statistical uncertainty in CEA by presenting results using cost-effectiveness acceptability curves (CEACs). The paper demonstrates using these methods how CEA can evaluate a 'health system' intervention.

Evaluations of 'system level interventions' for mental health services (Hu and Jerrell 1991; Alegria et al 1999) and in particular, different reimbursement models (Bloom et al. 2002; Christiansen et al. 1995; Cuffel et al. 2002; Dickey 1997; Manning et al. 1984; Manning et al. 1999; Ray et al. 2003; Wells et al. 1990) have used the cost-consequence model. Some of these studies reported that reimbursement by capitation was associated with lower costs compared to FFS (Bloom et al. 2002; Christiansen et al. 1995; Dickey 1997; Manning et al. 1984) and no statistically significant differences in outcomes (Cuffel et al. 2002; Wells et al. 1990); other studies found that capitation was associated with worse quality of care (Manning et al. 1999; Ray et al. 2003). None of these studies combined costs and effectiveness in a formal CEA (Gold et al 1997, Drummond et al 2005)[1]. None of these studies used a choice-based outcome measure such as the QALY, or appropriate methods for reporting the statistical uncertainty surrounding the cost-effectiveness results.

METHODS

Study design in CEA

CEA *compares* two of more health care programs, and assesses the *incremental* cost-effectiveness for the decision context of interest. Usually each individual is only observed receiving *one* of the interventions in question. To address this causal inference problem CEA may be conducted alongside a RCT. However, for CEA of area level interventions RCT data may be unavailable and the only data may be from a non-randomized study. As selection is non-random, the cases in each group are not drawn from the same population and so cost-effectiveness estimates may reflect pre-existing differences between the groups rather than the effect of the intervention itself. Methods are therefore required that minimize differences across the groups, so it is as if the samples are drawn from the same population**.** We describe principles followed at both the design and analysis stages for minimizing differences across the groups.

A key issue in CEA is to combine costs and outcomes in a way that can appropriately recognize the statistical uncertainty surrounding estimates of cost-effectiveness (Willan and Briggs 2006). In this paper we estimate incremental net benefits (INB) by valuing differences in outcomes across the health care programs by λ, the willingness to pay for a QALY gained. CEACs are then derived by re-estimating the INB varying λ between £0 and $200000 per QALY gained and plotting the probability that each program is cost-effective at each value of λ (Fenwick et al. 2004).

Overview of case study

To illustrate how CEA can be applied in health services research, the techniques described were used to compare the cost-effectiveness of different reimbursement mechanisms. The scope of the study was limited to those cases who were already Medicaid beneficiaries, had used a mental health service before and were diagnosed as having a severe mental illness. This CEA compares FFS with two different capitation models using a study conducted alongside the Colorado Medicaid Capitation Pilot Program (Bloom et al. 2002). In the first capitation model services were provided directly (direct capitation or DC) by not-for-profit community mental health centers (CMHC). In the second capitation model services were provided in a joint venture between CMHCs and a for-profit managed behavioral health organization (MBHO). The MBHO was an organization that covered several states, and had previous experience in implementing managed care, but not in the context of Medicaid services. In the remaining regions services continued to be provided by FFS.

Selection of areas for each reimbursement model

The Colorado Pilot program was implemented in selected parts of the state in August and September 1995 and required all Medicaid beneficiaries to participate, and providers were not allowed to select cases. Hence, unless cases moved or enrolled into Medicaid programs as a consequence of the pilot scheme, selection by patients or providers did not arise. The selection of areas for each reimbursement method was by the state who invited bids for capitation contracts to any entity that had the capacity to receive, manage and execute a capitation contract for Medicaid mental health services. During the bidding process existing mental health service providers

(community mental health centers of CMHCs) grouped together to form Mental Health Service Agencies (MHSAs**).** The state assessed how ready each entity would be to deliver capitated mental health services for the Medicaid program. In four areas the state perceived that existing CMHCs were ready for capitation and awarded them direct contracts with the state (DC model). In three areas the state judged that existing CMHCs were 'not ready' for capitation and awarded the contract to a for-profit managed behavioral firm. The state encouraged this firm to form an alliance with existing CMHCs, which the study termed the Managed Behavioral Organization Model (MBHO). In three areas the state judged inappropriate for capitation, reimbursement by FFS was maintained (Bloom et al 1998). The key concern for the CEA was that the selection of areas for capitation was non-random and according to the perceived readiness of the organizations concerned for capitation.. The state assessed 'readiness for capitation' according to criteria such as: whether was an appropriate management information system, whether there was a suitable financial system in place for costing services, and whether there were appropriate strategies for utilization review (see Bloom et al 2000). The DC areas scored highest on these readiness criteria. The for-profit managed behavioral firm had no previous experience of administering capitation services for Medicaid. As the DC group was perceived to be 'most ready' it was anticipated that the non-random selection would exaggerate any cost reductions observed in this group.

Sampling strategy

The purpose of the study was to compare the relative cost-effectiveness across three different reimbursement models; hence it was important to minimize differences in area and patient characteristics across all three groups. The study used a matched

group design which aimed to include similar areas across the three groups. The study used 1990 US census data on the proportion of the population in each area in poverty, the degree of rurality and the industrial base as it was anticipated that these variables could be associated with costs and outcomes (Bloom et al. 1998). The study then selected those counties that had similar area level characteristics (Appendix 1).

From those areas included, the study took a random sample of those cases who were already Medicaid beneficiaries, had used a mental health service before and were diagnosed as having a severe mental illness (diagnoses of schizophrenia, bipolar affective disorder, or at least one 24-hour inpatient stay with a primary mental health (DSM-IV) diagnosis). A total of 522 cases were available for the CEA.

Measurement of cost and utilization

Cost and outcome data were collected for one year pre capitation when all regions were reimbursed by FFS and two years post capitation. In the period immediately following capitation the first three months were viewed as an implementation period, and were excluded from the CEA as were the corresponding periods in the second period post capitation and prior to capitation. This gave cost and outcome data for three nine months period (one pre, and two post capitation).

The cost measurement took a Medicaid perspective and excluded costs borne by other payers. Costs included were those in the capitation rates that covered all Medicaid-eligible individuals for psychiatric inpatient care, specialty mental health outpatient services, and mental health services in nursing homes, but excluded the cost of pharmaceuticals. Costs for all three groups prior to capitation and for the FFS group

for all three time periods (1995-1998) were taken from Medicaid claims databases. Cost data were not available from the Medicaid claims database for the capitation group following capitation; these data were recorded from the state's shadow billing system. The shadow billing system required the capitated providers to report identical cost information to claims data.

The study measured the total costs of each episode of care for each user including: inpatient stays (state and local hospitals) and outpatient care (individual or group therapy, case management and day treatment programs). These total costs per episode were used to derive measures of unit cost and utilization such as the proportion of cases using inpatient or outpatient services during each period.

Measurement and valuation of health outcomes

The CEA reported health outcomes using QALY, which required that the vital status of each case was noted, and for the decedents, information on the date of death was obtained from death certificates, to record survival duration. To estimate Health-related quality of life (HRQOL), trained investigators administered the SF-36 health survey at six monthly intervals throughout the study. The algorithm developed by Brazier et al (2002) was chosen to value the health states described by the SF-6D, a subsample of the SF-36 health states. For each case, HRQOL at each time point was multiplied by the corresponding survival time to give QALYs for each nine-month period.

Matching at the analysis stage

Randomizing a sufficiently large number of cases to each reimbursement model would ensure that there were no baseline differences in patient or center characteristics across the intervention groups. This non-randomized study recorded patient characteristics prior to the introduction of capitation, and despite the attempts to match areas with similar characteristics at the design stage, there were differences between the patient groups at baseline (see Table 1). For example, mean costs prior to capitation were significantly higher in the MBHO ($6,822) than the FFS group ($4,820) (t-test p=.02). These differences in baseline costs partly reflect differences in patient-mix, for example the mean costs for men were higher than for women, and the MBHO group had the highest proportion of men. However, the MBHO model clearly had higher baseline costs even after allowing for differences in patient factors. Hence, it is important to match on baseline cost as well as case-mix variables. By adjusting the samples according to baseline cost, the analysis recognizes differences in baseline cost that arise according to the areas concerned.

Where there are large imbalances in baseline covariates as in this case study, using a parametric model to adjust for differences is problematic: the results are generally, sensitive to the choice of model specification (Rubin 2006). The previous cost analysis of the same data used a parametric model, the two part model, to try and adjust for baseline differences between the groups (Bloom et al 2002). A problem with this approach is that it only allows for ***mean*** differences across the groups, and therefore ignores differences elsewhere in the distribution.

To allow causal inferences to be made when parametric adjustment is problematic, matching methods are recommended (Rubin 2006). This study employs a non-parametric matching method, Genetic Matching which is a generalization of propensity score and Mahalanobis distance matching (Morgan and Harding 2006, Raessler and Rubin 2005). The method has been shown to outperform more commonly used matching methods (such as propensity scores) and has been applied in a wide range of areas (see for example Heron and Wand 2007; Morgan and Harding 2006; Raessler and Rubin 2005). The method does not require the analyst to make parametric assumptions, which is important in this context given that cost data generally have highly irregular distributions. The method uses a genetic algorithm (Sekhon and Mebane 1998) to identify those matches which achieve the best possible covariate balance (Diamond and Sekhon 2006; Sekhon 2007).

In this case study Genetic Matching was used to identify cases in each capitation group to match to cases in the FFS group. The matching algorithm used the same covariates as the previous parametric model which were baseline measures for: demography (age, sex, ethnicity), diagnosis (schizophrenia, bipolar affective disorder, other), pre-capitation utilization, QALYs, and cost. The algorithm selected cases using the results of t-tests and non-parametric Kolmogorov-Smirnov (KS) tests that compared the distribution of these covariates across the groups. The KS test is a non-parametric test of the equality of two empirical cumulative distributions. This test is distribution free so it does not rely on the assumption of normality which is important given the highly skewed and kurtotic distribution of cost data. When the KS test is bootstrapped, it is consistent even when variables do not have a continuous

distribution (Abadie 2002). For example, in this dataset the distribution of the cost variable has a point mass at zero and it is certainly not normally distributed.

After applying the matching algorithms no significant differences remained between the groups (Table 1). All subsequent analyses were conducted using the matched dataset.

Cost and cost-effectiveness analysis

Costs and QALYs were reported for each patient for each observation period (nine months pre capitation, and two nine month periods post capitation). Costs and QALYs in the second follow-up period were discounted at the recommended rate of 3% (Gold et al. 1996). Total costs and QALYs were calculated by summing costs and QALYs across the two follow-up periods. Given the skewed nature of the cost data, the analysis did not assume that the data were drawn from a normal distribution, and instead used the bootstrap KS test, and the non-parametric bootstrap (bias corrected) to report 95% CIs around incremental costs and QALYs (Thompson and Barber 2000). CEACs were derived by using the bootstrap replications to plot the probability that each capitation model was cost-effective at different values for λ.

The CEA was repeated for different patient subgroups, for example those cases with schizophrenia as opposed to bipolar affective disorder. Sensitivity analysis applied parametric models to adjust for remaining differences in patient and area-level characteristics across the groups. This analysis used a two-part model to estimate incremental costs (Mullahy 1998), and a multiple linear regression model to estimate incremental effectiveness.

RESULTS

For these previous users of mental health services, service utilization fell in all three groups over the study's observation periods. For inpatient services the reduction in service use was similar across the groups (Table 2). These overall changes may reflect reversion to the mean; however the key finding is that there were differences in the reduction in outpatient services according to reimbursement model. The reduction in outpatient utilization was largest in the MBHO group; where there was a 22% reduction by the end of the second follow-up period (post 2^{nd}) compared to a 7% reduction in the FFS group (p=0.04). The corresponding reduction in outpatient utilization in the DC group (12%) was not significantly different to the FFS group (p=0.29) (Table 2). The mean cost for service users was lower post capitation in the MBHO group but higher in the DC group, compared to FFS. The net effect of these changes in utilization and cost were that post capitation, the mean costs per case were higher for the DC model than the FFS model, whereas the MBHO model had lower mean costs per case (Table 2).

A total of 373 (82%) cases completed SF-36 surveys at each time point; the mean HRQOL was 0.63 for each group at baseline (Table 2). The mean HRQOL was higher in the MBHO group at follow-up, and so this group had higher mean QALYs.

Compared to FFS, the MBHO model had negative incremental costs (-$1991). Although the bootstrapped 95% CIs around this estimate of incremental costs included zero (Table 3), the p-value for the bootstrapped KS-test was 0.01. This non-parametric KS test is more appropriate given the highly non-normal distribution of the cost data. The DC model had positive incremental costs of $4694 compared to FFS

(95% CI from 302 to 10170; KS-test p=0.08). The incremental costs of the MBHO model compared to DC were -$6685 (95% CI from -$11242 to -1658). Aside from the significant difference in mean costs, the MBHO model had significantly lower costs as determined by the non-parametric KS-test (p=0.002). Indeed, the MBHO model had lower costs across the entire distribution of costs (empirical QQ-plots available upon request).

The MBHO model had positive incremental QALYs compared to FFS or DC, although the confidence intervals around the central estimates were wide and included zero. The CEA found that the DC model was not cost-effective compared to the FFS or MBHO models, across a range of values for the cost-effectiveness threshold, λ. For example, when λ was valued at $50000 per QALY, the INB for the DC model compared to FFS was -$5477 (Table 3). The MBHO model was cost-effective compared to either the FFS or DC models. For example, at $50000 per QALY the mean INB of MBHO compared to DC was $8428 (95% CI from $3338 to $13297) (Table 3).

The CEACs plot the probability that either capitation model is cost-effective for different levels of λ (Figure 1). The intersection with the y axis, shows the probability that 'the intervention' is cost-effective when only cost differences are considered. As the value for λ increases, relatively more weight is given to the incremental effectiveness. At all realistic levels of the cost-effectiveness threshold, the probability that the MBHO model is cost-effective compared to either FFS or DC, exceeds 0.90. For example at $50,000/QALY the probability that the MBHO model is cost-effective compared to FFS is 0.91. The CEAC for the MBHO versus FFS comparison does not

asymptote to 1 as it is not certain that the MBHO model is more effective than FFS; although the mean incremental QALYs are positive, the CIs surrounding this estimate include zero. The CEACs also show that the probability that the DC model is cost-effective compared to FFS is less than 0.1 at all values of λ.

The sensitivity analysis found that the MBHO model remained the most cost-effective after applying parametric and semi-parametric models to allow for any remaining differences across the groups post-matching**.** As the data were well matched, the results were not sensitive to the choice of model. The sub group analysis showed that for patients with schizophrenia (72% of cases) the MBHO model was most cost-effective. However, for patients with bipolar affective disorder, both capitation models were associated with increased costs and no gain in QALYs compared to FFS.

DISCUSSION

This paper presented some key methodological features of CEA and illustrated these techniques with a case study. The CEA found that the capitation model with the for-profit component was the most cost-effective at all levels of willingness to pay for a QALY gained. The CEA incorporated any differences in both costs and outcomes across the reimbursement models, and therefore extended previous cost minimization analyses (CMA) that have focused on the relative costs of managed care compared to FFS (Bloom 2002; Christiansen et al. 1995; Dickey 1997; Manning et al. 1984). The CEA used appropriate techniques to measure and value outcomes, to deal with baseline imbalances across the groups (Morgan and Harding 2006) and to allow for the skewed distribution of the cost data (Adabie 2002). The techniques presented could be used more generally for evaluating different ways of financing and providing health services where there may be differential impacts on costs and outcomes and where RCT data are unavailable.

An earlier paper reporting cost results from the same study found that both the not-for-profit capitation model (DC) and the capitation model with a for profit element (MBHO model) were associated with cost reductions compared to FFS (Bloom et al 2002). Our paper finds that the DC model is associated with higher costs, and the MBHO model lower costs compared to FFS. Under the DC model the costs for service users were higher compared to FFS, whereas in the previous paper these costs were reported as similar in the DC and FFS groups.

The reason for the difference in the cost results across the papers is the approach taken to adjusting for baseline differences across the groups. The previous paper used

a parametric model, the two part model and only allowed for ***mean*** differences across the groups at baseline. This is a particular deficiency for a variable such as baseline cost which is highly skewed; using the mean differences at baseline ignores important differences elsewhere in the distribution. Instead, we used a non-parametric technique Genetic Matching, as recommended in the biostatistics literature (Rubin 2006). The two key advantages of Genetic Matching are that: firstly it did not rely on parametric assumptions such as assuming that the baseline costs were normally distributed; secondly, rather than just adjusting the samples based on mean characteristics, it allowed for baseline differences across the groups right across the distribution. When this method was applied excellent covariate balance was achieved. Our results are not sensitive to model-based parametric adjustment post-matching.

The study illustrated that CEA can provide clear information on the relative cost-effectiveness of alternative reimbursement methods. Methodological guidance for economic evaluation requires that authors place appropriate limits on the generalizability of their results (Drummond et al. 2005). It is therefore important to recognize that the finding that a capitation model with a for-profit element was more cost-effective than a not-for-profit capitation model, may not be transferable to other health care contexts. When capitation was introduced for Colorado Medicaid mental health services, the state took steps to try and maintain service quality. For example, the state specified the services to be delivered in the capitation contract; strict limits were imposed on profits and further investment in mental health services was encouraged. These features may have been important in ensuring that similar health outcomes were maintained across reimbursement models. In other contexts, if

capitation schemes are less carefully implemented, they can lead to poorer quality of care (Ray et al. 2003), and may be less cost-effective than FFS.

This study was restricted to previous users of mental health services. These patients were relatively costly (average cost of $7500 per year) and there may have been more scope for reductions in utilization for these users than for other groups, for example patients with less severe mental illness or newly identified patients. The sub group analysis found that while the for-profit model was most cost-effective for patients with schizophrenia, FFS was more cost-effective for patients with bipolar affective disorder who had lower average costs.

The cost-effectiveness results in the case study were driven by cost differences across the reimbursement models. A potentially important feature of the capitation models was that contracts were re-tendered every two years. In the for-profit areas the contracts moved between health care organizations, whereas in the DC areas the contracts remained with the same CMHCs. Faced with this greater risk coupled with the incentive to make profits, the MBHO group may have been more inclined to adopt processes that reduced costs while maintaining quality. For example, a qualitative investigation of care processes found that in the MBHO areas utilization review (UR) informed the management of each case (Bloom et al. 2000). By contrast in the DC areas, administrators only employed UR for outlying cases. For patients with severe mental illness, costs are notoriously difficult to predict and using UR for all cases would be more likely to identify those cases with scope for cost reduction. Furthermore, interviews with decision-makers in the DC areas, suggested that, faced with little incentive to reduce costs, there was more emphasis on expanding services

(Bloom et al. 2000). This strategy appeared to lead to higher costs without improvements in patient outcomes.

General concerns that capitation leads to 'cream skimming' are unlikely to apply in this study as health care providers were legally required to maintain access to care for the cases in the study who were all Medicaid enrollees**.** The state selected for the not-for-profit capitation model those CMHCs judged 'ready' for capitation those CMHCs judged 'not ready' were linked with a for-profit managed behavioral health organization (MBHO) (Bloom et al. 2000). It was anticipated this selection process would lead the CEA to overstate the cost-effectiveness of the not-for-profit capitation model. As the study found the for-profit capitation model was relatively cost-effective, the findings are robust to bias in the selection of centers.

Guidelines for CEA recommend that ideally a broad range of costs are included and a lifetime time horizon is taken for the analysis (Luce et al. 1996). Compared to this 'gold standard' the case study presented had certain limitations; for example costs outside the capitation contract including, pharmaceuticals were excluded. Another study found that the only difference in pharmaceutical costs was that the DC group used more antipsychotic medication compared to FFS (Wallace et al. 2005). Hence, including these costs would further substantiate the conclusion the DC model was not cost-effective. Of greater concern is the relatively short time frame adopted. While a follow-up study found that the MBHO and DC models had similar costs after six years (Wallace et al. 2006), further research is required to evaluate the long-term cost-effectiveness of different reimbursement mechanisms using the techniques outlined.

The methods presented are of general use to policy-makers aiming to reduce costs without compromising the quality of care. They are particularly relevant for evaluating Medicaid programs where budgetary pressures are perennial (Johnson 2005). CEA highlights tradeoffs between costs and outcomes, allowing policy-makers with differing views on the relative importance of costs versus outcomes to use the same analysis.

In conclusion, this study illustrates appropriate methods for estimating and valuing health outcomes, adjusting for differences in patient mix across the intervention groups, and representing the sampling uncertainty surrounding the results. The case study found that a capitation model with a for-profit element was more cost-effective than either a not-for-profit capitation or FFS model for Medicaid patients with severe mental illness. These techniques can be applied to a wide range of contexts in health services research, to help policy-makers identify which health care programs to prioritize.

Notes

[1] Here CEA is defined broadly to include studies that report outcomes as utilities.

REFERENCES

Abadie, A. 2002. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models." *Journal of the American Statistical Association* 97(457) (March): 284-292.

Alegria, M., Frank R., McGuire T. 2005. "Managed care and systems cost-effectiveness: treatment for depression." *Medical Care* 43(12): 1225-33.

Bloom, J.R., K. Devers, N.T. Wallace, and N.Z. Wilson. 2000. "Implementing capitation of Medicaid mental health services in Colorado: Is "readiness" a necessary condition?" *Journal of Behavioral Health Services and Research* 27(4):437-445.

Bloom, J. R., T. W. Hu, N. Wallace, B. Cuffel, J.W. Hausman, and R. Scheffler. 1998. "Mental health costs and outcomes under alternative capitation systems in Colorado: Early Results." Journal of Mental Health Policy and economics 1(I):1-11.

Bloom, J. R., T. W. Hu, N. Wallace, B. Cuffel, J.W. Hausman, M.L. Sheu, and R. Scheffler. 2002. "Mental health costs and access under alternative capitation systems in Colorado." *Health Services Research* 37(2): 315-40.

Brazier, J., J. Roberts, M. Deverill. 2002. "The estimation of a preferenced-based measure of health from the SF-36." *Journal of Health Economics* 21: 271-292.

Cuffel, B.J., J.R. Bloom, N. Wallace, J.W. Hausman, and T.W. Hu. 2002. "Two-year Outcomes of Fee-For-Service and Capitated Medicaid Programs for people with severe mental illness." *Health Services Research* 37(2): 341-359.

Diamond, A., and J.S. Sekhon "Genetic matching for estimating causal effects: a general matching method for achieving balance in observational studies." Working Paper.

The Society of Political Methodology. [accessed September19th,2006] available at http://sekhon.berkeley.edu/papers/GenMatch.pdf.

Dickey, B. 1997. "Assessing Cost and Utilization in Managed Mental Health Care in the United States." *Health Policy* 41 Suppl: S163-174.

Drummond, M.F., M.J. Sculpher, G.W. Torrance, B.J. O'Brien, and G.L. Stoddart. 2005. *Methods for the economic evaluation of health care programmes. Third edition*: Oxford: Oxford University Press.

Fenwick, E., B.J. O'Brien, and A. Briggs. 2004. "Cost-effectiveness acceptability curves - facts, fallacies and frequently asked questions." *Health Economics* 13: 405-415.

Gold, M.R., J.E. Siegel, L.B. Russell, and W.C. Weinstein (eds).1996. *Cost-Effectiveness in Health and Medicine*. Oxford University Press, NewYork.

Herron, MC. and Wand J (2007) . "Assessing Partisan Bias in Voting Technology: The Case of the 2004 New Hampshire Recount." *Electoral Studies* (in press)

Hu T.W., Jerell J. "Cost-effectiveness of alternative approaches in treating severely mentally ill in California." *Schizophrenia Bulletin* 1991: 17: 461-8.

Johnson, P. "Medicaid: Medicaid: provider reimbursement—2005". 2005. End of Year Issue Brief Issue Brief Health Policy Track Serv. Dec 31;:1-11.

Luce B., W. Manning, J.E. Siegel, and J. Lipscomb. 1996. Estimating costs in cost-effectiveness analysis. In *Cost-effectiveness in health and medicine*. Gold MR, Siegel JE, Russell LB, Weinstein MC (eds.) Oxford University Press, New York.

Manning, W.G., A. Leibowitz, G.A. Goldberg, W.H. Rogers, and J.P. Newhouse. 1984. "A controlled trial of the effect of a prepaid group practice on use of services". *New England Journal of Medicine* 310: (23): 1505-10.

Manning, W., C.F. Liu, T.J. Stoner, D.Z. Gray, N. Lurie, M. Popkin, J.B. Christianson. 1999. "Outcomes for Medicaid Beneficiaries with schizophrenia under a Prepaid Mental Health Carve-Out." *The Journal of Behavioral Health Services and Research* 26: 442-450.

Morgan, S.L., and D.J. Harding. 2006. "Matching estimators of causal effects. Prospects and pitfalls in theory and practice." *Sociological methods and research* 35(1): 3-60.

Mullahy, J. 1998. "Much ado about two: reconsidering transformations and the two-part model in health econometrics." *Health Economics* 17(3): 247-181.

National Institute for Clinical Excellence (NICE). 2004. *Guide to the methods of Technology Appraisal.* NICE: London.

Neuman, P.J. 2004. "Why don't Americans use cost-effectiveness analysis?" *The American Journal of Managed Care* 10:308-312.

Raessler, S. and D.B. Rubin. 2005. "Complications when using nonrandomized job training data to draw causal inferences". Proceedings of the International Statistical Institute.

Ray, W.A., J.R. Daugherty, and K.G. Meador. 2003. "Effect of a mental health "carve-out" program on the continuity of antipsychotic therapy". *New England Journal of Medicine* 2003; 348: 1885-94.

Rennie, D. and H.S. Luft. 2000. "Pharmacoeconomic analyses: making them transparent, making them credible." *JAMA* 2000; 283:2158-2160.

Rubin, D. 2006. *Matched Sampling for Causal Effects*. Cambridge University Press, New York.

Sekhon, J.S. 2007. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software* (in press). Computer program available at http:// sekhon.polisci.berkeley.edu/matching/

Thompson, S.G., and J.A. Barber. 2000. "How should cost data in pragmatic trials be analysed?" *British Medical Journal* 320: 1197-2000

Ubel, P.A., R.A. Hirth, M.E. Chernew, and A.M. Fendrick. 2003. "What is the price of life and why doesn't it increase at the rate of inflation?" *Archives Internal Medicine* 163: 1637-1641.

US Department of Health and Human Services, Health Resources and Services Administration, and Bureau of Health Professions. 2002. Area Resource File (ARF) Rockville, MD.

Wallace, N., J. Hyun, H. Wang, S.H. Kang, T.W. Hu and J. Bloom. 2006. "Six Years of Longitudinal Analyses of Capitation in Colorado" [unpublished manuscript].

Wallace NT, Bloom JR, Hu TW, Libby AM. 2005. "Medication treatment patterns for adults with schizophrenia in Medicaid managed care in Colorado." *Psychiatric Services*. 56(11):1402-8.

Wells, K., W. Manning, and R.B. Valdez. 1990. "The effects of a prepaid group practice on mental health outcomes." *Health Services Research* 25(4): 615-625.

Willan A, Briggs A. 2006. Statistical analysis of cost-effectiveness data. Wiley, Chichester, England.

Table 1: Baseline costs, QALYs, client characteristics and utilization: before and after matching[a]

	FFS vs DC			FFS vs MBHO		
	FFS	DC	P value[b]	FFS	MBHO	P value
Mean costs						
before matching	4820	4524	0.11	4820	6822	0.02
after matching	4820	4805	0.32	4820	4580	0.42
Mean QALYs[c]						
before matching	0.475	0.485	0.10	0.475	0.482	0.29
after matching	0.475	0.476	0.78	0.474	0.474	0.33
% schizophrenia						
before matching	72.2	61.9	0.05	72.2	65.1	0.16
after matching	72.2	70.8	0.68	72.2	72.2	1.00
% bipolar						
before matching	21.2	30.7	0.05	21.2	25.6	0.33
after matching	21.2	22.5	0.48	21.2	21.9	0.56
Mean age						
before matching	43.4	42.3	0.58	43.4	45.1	0.32
after matching	43.4	43.4	0.84	43.4	43.7	0.86
% men						
before matching	44.3	47.7	0.54	44.3	49.7	0.32
after matching	44.3	43.7	0.78	44.3	45.0	0.70
% previous high cost client						
before matching	37.1	36.4	0.89	37.1	31.8	0.31
after matching	37.1	37.1	1.00	37.1	36.4	0.32
% using any service						
before matching	89.4	93.8	0.16	89.4	90.3	0.80
after matching	89.4	89.4	1.00	89.4	89.4	1.00

[a] Before matching: n=522 FFS (n=151), DC (n=176), MBHO (n=195), after matching: n=453 (n=151 in each group).
[b] The tests conducted are non-parametric bootstrap Kolomogorov-Smirnov distributional tests

[c] Note that QALY data were not available for 8 cases before, and 4 cases after matching

Table 2: Utilization of services (%), mean costs ($), HRQOL and QALYs; pre and post capitation.

	Time period	FFS	DC	MBHO
Utilization of services (n=453)				
Inpatient	Pre	15.2	15.2	14.6
	Post (1st)	10.6	3.3	9.9
	Post (2nd)	8.6	8.6	9.9
Outpatient	Pre	89.4	89.4	87.4
	Post (1st)	88.7	82.8	75.5
	Post (2nd)	83.4	78.8	68.2
Costs (n=453)				
Cost per user[d]	Pre	5391	5375	5123
	Post (1st)	4888	7116	3837
	Post (2nd)	4794	9002	4714
Cost per case[d]	Pre	4820	4805	4580
	Post (1st)	4338	5938	2989
	Post (2nd)	4000	7094	3359
Outcomes (n=373)				
HRQOL	Pre	0.63	0.63	0.63
	Post (1st)	0.64	0.62	0.64
	Post (2nd)	0.63	0.61	0.65
QALY	Post 1st + Post 2nd	0.934	0.919	0.954

[d]Note that all cases in the sample used service *prior* to study entry, cost per user gives the cost for those using services in the given period, whereas cost per case reports costs for all those in the sample

.**Table 3: Incremental costs ($), incremental QALYs and INBs ($). Mean estimates (95% CI)**

	DC-FFS	MBHO-FFS	MBHO-DC
Incremental costs	4694(302 to 10170)	-1991(-5801 to 1839)	-6685(-11242 to –1658)
Incremental QALYS	-0.016(-0.061 to 0.026)	0.019(-0.017 to 0.059)	0.035(-0.006 to 0.073)
Incremental net benefit (λ=$50000)	-5477(-10832 to –542)	2950(-1697 to 7078)	8428(3338 to 13297)
Incremental net benefit (λ=$100000)	-6262(-12779 to -13)	3908(-1717 to 9279)	10169(3890 to 16113)

Figure 1: Cost-effectiveness acceptability curves