Section 6. Cross-Site Analysis Plan

6.1. **Overview of Planned Cross_site Analyses**

The cross-site analyses will make use of information from a variety of sources, representing measurement at multiple levels. In addition to individual-level demographic, outcome and services data obtained through the baseline and follow-up common protocol self-reports and, from most sites, the cost study, project-level data will be available from site visits and the fidelity/implementation ratings. Clusters of sites will also collect data using instruments not included in the multi-site common protocol. The cross-site analyses will use and integrate these data to produce a valid picture of the extent to which participating in COSPs in conjunction with more traditional mental health services improves important outcomes for participants, as well as the costs of these benefits, compared to participating in traditional mental health services alone.

The analyses will proceed through several preparatory and preliminary stages, building a firm foundation for modeling program and participants characteristics, implementation and participant change over time, differences between the experiences and outcomes of participants in traditional mental heath services and COSPs, and the variation in each of these aspects among individual sites and clusters of sites. These steps are depicted in Exhibit 6.1, and will be described in the remainder of this chapter. The rest of this section reviews the steps in turn, and discusses a number of features of multi-site design in general and of the COSP in particular that require special analytical consideration.

6.2 **Multisite Analysis Issues**

Because the COSP Multisite Research Study is a multisite evaluation, the conduct of the cross-site analysis of this evaluation requires analysts to address a number of special data management, analytical and statistical issues:

- Unlike multi-center clinical trials, typical multisite demonstrations or knowledge-development efforts include diverse programs and populations across the sites. Often, the nature of programs is such that program effects depend in part on characteristics of the communities and service systems in which they operate. These are known as "context effects" (Kreft and DeLeuw, 1998). These features demand special attention to both descriptive and process data collection and preparation for integrating multiple types of data in the analyses.

- The variety of programs included in the COSP multisite, as well as the successful effort during the first year of the study to find common ingredients of COSPs mean that data from the COSP characterization of the diverse programs using FACIT (see Section 7) will be crucial in understanding variation in outcomes.

- Concerns about differences in the populations participating in different COSPs as well as the implementation of the screening procedures and eligibility criteria mean that it will be also important to examine cross-site variation in the characteristics of participants, both in deciding whether it is reasonable to pool data from COSPs with similar program characteristics and in understanding and attempting to control for baseline differences.

- In addition, as with other cooperative agreements and collaborative research design efforts, the items in the common protocol are extracted from a wide range of instruments, and arranged in novel ways. Although there were efforts to keep sections or entire pre-existing instruments intact, some sets of items retained have been juxtaposed in novel ways, and some new items and wordings of items have been included. The relationships among the items and scales in the CP must therefore be assessed in preliminary methodological analyses focused on reestablishing the psychometric properties of the instruments as well as sensitivity to change.

- And finally, the COSP study includes a strong cost component which is being conducted within a framework that emphasizes the need to integrate the analysis of costs with analyses of the relationship between program elements and outcomes.

These and other considerations were the basis for planning the pattern of branching analytical paths in Exhibit 6.1. Our approach is one in which the analyses build up from characteristics of the data and populations, adding

characteristics of programs and contexts, and integrating these into statistical tests and summaries of comparative outcomes, while supporting and probing the conclusions as threats to validity are encountered. Some analytical activities may overlap in time or be returned to after later results are obtained. A number of the proposed analyses may be appropriate for ad hoc workgroups comprising site analysts, CC staff, and consultants. This section outlines one approach on the components that are crucial to include, as a basis for discussion of further analysis planning. These include the definition and operationalization of covariates and outcomes; methodological analyses, descriptive population analyses; analysis of implementation/fidelity and other project-level data; multi-level outcome analyses (nonmonetary and monetary); and cost, cost_effectiveness and cost_benefit analyses

6.3 Data Quality Assessment and Development of Measures

Several steps will be required to bring the data to a point of readiness for inclusion in these comprehensive statistical modeling techniques. These will include several preliminary data assessment and descriptive analyses, as well as the development and assessment of composite scores and scales that will serve as key covariates and measures of process and outcome.

Although data quality has been identified as a site-level responsibility, it must be a coordinating center responsibility to provide feedback on of data quality, especially if there are apparent inconsistencies in coding and data handling across sites. Although certain data quality assurance mechanisms have been incorporated into data entry and routine cross-site data management (see Section 11), others are analytical, involving comparing distributions of items and scales scores among sites.

Interim analyses of the cross-site characteristics of common protocol and cost items will begin as soon as enough baseline data have been collected to make it feasible and profitable.

6.3.1 Methodological Analyses to Support Data Quality Assurance and Measure Development Efforts

As part of developing the measures that will be used in analyses of outcomes and their relationship to costs, the characteristics of proposed scales and composite scores, as well as their constituent items, will be examined and systematically assessed. In addition, the variation among sites and possible non-equivalence between TMHS and COSP participants at baseline or at follow-up must be assessed before any valid comparative outcome analyses can be conducted.
After some preliminary discussion of initial proposed measures, the working group will determine which shall be primary outcome measures and which shall be used as covariates. Further analyses of sensitivity to change may be required to settle on the final list of outcomes to be included in cross-site analyses. After preliminary results of these analyses are available, it will be appropriate to conduct analyses of the baseline data which describe the COSP participants in terms of the selected and developed measures.

Among the first steps in the cross-site interim analyses is a careful assessment of data quality. Examination of a range of possible data problems will be performed. Results of these checks will be combined with guidance from current theoretical, methodological, and empirical literature regarding what to do about any identified problems. Specific data problems to be examined include, but are not limited to, extreme values as well as "non_normal" distributions on continuous variables. Lack of variability on key predictor, covariate, or outcome variables is also a concern because of the limitations it imposes on the utility of the affected variables for analysis.

The amount of missing data will need to be tracked closely, both in general and as regards particular items. As was done with the pilot study data, the data will be scanned to identify items with disproportionate (we suggest an initial threshold of $\geq$ 10%) amounts of nonresponse. Items meeting this criterion will be scrutinized to determine whether there are aspects of content, form in which a question is asked, or other possibly causal factors in the nonresponse. Attempts will be made during subsequent data collection to alleviate such factors, e.g., by repeated assurances of confidentiality. In the analysis phase, a range of options exist for dealing with missing data, depending on the nature of the “missingness”: missing completely at random, ignorably missing, or nonignorably missing (Little & Rubin).

These include various imputation strategies as well as members of the random regression family of models.

6.4 Analyses Utilizing Baseline Data

6.4.1 Initial Methodological and Descriptive Analyses

The initial round of methodological analyses will follow upon and extend the analysis plan implemented for the pilot data. In addition to checks on data quality and potential data problems as described above, the methodological analyses will begin with basic descriptive statistics, including measures of central tendency and dispersion (means, medians, standard deviations, percentiles, skewness, kurtosis, and ranges). Graphical and statistical indices of the form of the distribution of each continuous variable will also be obtained. Similarly, for categorical variables, the percentages of respondents falling into various response categories will be determined. The psychometric properties of each scale in the CP, including internal consistency reliability, and, to the extent possible, discriminant and convergent validity measured against other components of the CP and/or site-specific instruments as available, will also be determined.

Analyses of pilot data revealed, for the most part, the instruments and measures selected for inclusion within the baseline Common Protocol exhibited very acceptable internal consistency reliabilities, defined as alpha coefficients of ≥ 0.70. The only notable exceptions to this general conclusion included a set of subscales of the Herth Hope Index, the somatization subscale of the Hopkins SCL-25, the "objective" Social Inclusion measure, and the Quality of Life Excerpts covering domains other than social inclusion. These findings will, however, be reassessed utilizing data from randomized participants.

In terms of convergent validity as measured by scale-to-scale correlations, the correlations among the Recovery Assessment, Empowerment, Hope, and Meaning of Life Scales were identified as being of particular interest during the pilot analyses. These were expected to be highly positively correlated with one another, perhaps so much so ($\geq$ 0.80) that the inclusion of all three might appear to result in considerable redundancy. If this had been the case, several investigators would have been inclined to consider deleting one of this set of scales. However, expectations of extremely high correlations among these scales were not totally met. Specifically, there were moderate (.40-.70) correlations between some pairs, including total Herth Hope Index with Empowerment Making Decisions, Herth Hope Index with Recovery Assessment, Empowerment Making Decisions with Recovery Assessment, and Recovery Assessment with Personal Empowerment. But only modest positive correlations were observed between, for example, Empowerment Making Decisions and Personal Empowerment, Herth Hope Index and Personal as well as Organizationally Mediated Empowerment, and Recovery Assessment and Organizationally Mediated Empowerment. The magnitudes of correlations among these scales will continue to be closely tracked as baseline data from randomized participants become available.

Conversely, and contrary to expectations, results of pilot data analyses showed the Meaning of Life scale to correlate negatively with Recovery Assessment, all measures of Empowerment, and Hope. The meaning of this finding was not immediately clear. Other, more generally expected results of the pilot analyses included small to moderate negative correlations between measures of empowerment, recovery, and the Herth Hope Index, on the one hand, and measures of symptoms on the other. There were also negative correlations between measures of empowerment and measures of satisfaction/outcome related to traditional mental health services. All these findings will be revisited with respect to baseline data from randomized participants.

In addition, the feasibility and appropriateness of data reduction through methods such as factor analysis will be investigated. Identification of candidate scales/dimensions will be guided by the logic model and applicable theory, as well as other considerations, including the judgment of involved analysts who have experience with the measures nominated and included in the several CP domains.

The basic descriptive statistics obtained in the initial methodological analyses will be vitally important in characterizing the study population overall, by site and, within site, by randomization condition. In addition to

demographics such as gender, age, ethnicity, and level of education, baseline levels of outcome variables such as empowerment, hope, and symptoms will be examined. The comparison of respondents in the TMHS alone with those in the TMHS + COSP conditions, will give an indication of the extent to which the randomization within sites "worked," i.e., the extent to which respondents in the two conditions are comparable with respect to study-relevant characteristics.

In addition, variation among sites in distributions on key baseline variables must be assessed, and decisions made on the methodologically optimal approach to dealing with these, before any valid comparative outcome analyses, especially those involving data pooled across sites, can be conducted. Findings of important nonequivalences at both these levels in the descriptive analyses will drive the identification of potentially important covariates or "control variables" (potential confounders) for inclusion in multivariable models of study outcomes. As well, once follow-up data are available, baseline characteristics will need to be compared between followed and nonfollowed participants at each site, and strategies developed for taking account of losses to follow-up.

6.4.2 Preparing Project-level Data for Inclusion in Analyses

The multisite evaluation will make systematic use of a wide variety of qualitative and quantitative process, implementation, and contextual data about the COS and their community contexts. These data will be maintained in a *project-level database* (PLD), which will be crucial to effective analysis of these data and maximizing their utility for policy guidance. This type of database has been used to links the process and outcome components of multisite evaluations (NIAAA, 1992; CMHS, 1996). It also represents one kind of link between qualitative and quantitative research traditions (Dennis, Fetterman & Sechrest, 1994) in assessing project implementation and in tracking fidelity to program models.

The proposed will also the means by which summary characteristics of projects and their environments will be incorporated into *multilevel analyses*, including hierarchical linear modeling, as detailed below. It can also be used to support quantitative *research synthesis* (meta-analysis) strategies for integrative reviews of evaluations. In a field such as the evaluation of COSPs where little prior formal evaluation exists, the design and development of a project level database can lay an important foundation for the continuing synthesis of subsequent evaluations of such programs.

Data on populations, program models and components, implementation problems, local environmental contexts, and evaluation methods will be rated or coded, with checks on the reliably of the coding (Orwin, 1993; Stock, 1993). The goal is to accommodate all of these sources of analytical complexity and potential problems (e.g., comparison group nonequivalence) by recognizing them at the outset and as they arise, and facilitating solutions to these problems within the analysis. As changes occur and new issues arise, each can also be tracked and coded.

6.4.3 Methodological analysis of the COSP/FACIT scales

The COSP/FACIT rating scales represent a new implementation/fidelity assessment instrument (See Section 7). It will be appropriate to conduct a variety of analyses both of the scales and constituent items and of their relationships with other measures, to assess the utility and validity of the instrument in describing COSP programs and environments.

These are expected to include examination of the item distributions, correlations within and between scales, computation of internal-consistency reliabilities, interrater reliability, and item-analyses. In validating such an instrument, it would be desirable to do further work on the correlational structure, but the number of responses to be generated by the planned site-level data collection will be modest.

6.4.4 Descriptive analyses of the COSPs participating in the study

The main analyses involving the COSP/FACIT instrument will be the average ratings of the COSPs involved in the

multisite study. They will receive multiple ratings, by program and research staff at sites, and by site visitors. Although they will be checked against each other, and efforts will be made understand and possible reconcile extreme differences, the average ratings of programs on dimensions thought to be key in distinguishing COSPs from TMHS programs will computed and made available for inclusion in cross-site analyses of procedures, outcomes and costs.

6.4.5 Clustering—both analytical and judgmental—of COSPs for analysis purposes

With a sufficient sample of COSPs and TMHSs, it would be possible to conduct a statistical cluster analysis both to confirm and to suggest typologies of programs. With the small number of COSPs included in this first study to use the instrument, clustering and validating typologies will require more judgmental examination of patterns in the COSP/FACIT ratings. A priori differences in program types should be reflected in patterns of ratings. The similarity of programs on important dimensions will play a key role in the decision to pool or cluster data from sets of COSPs.

6.4.6 Descriptive summary of participant characteristics, by cluster

Once clusters of COSPs by model type or COSP/FACIT ratings have been e\established, analyses will describe the characteristics of participants in the clusters of programs, in preparation for deciding whether it is reasonable to analyze pooled data. But it is also possible to analyze patterns of participant characteristics across sites, to suggest other ways of clustering sites within or across program types.

6.5 Approach to Longitudinal, Multi-level, Cross-site Analyses

Analysts have taken many different approaches to analyzing cross-site data. There are differences in statistical modeling approaches, but beyond these, there are differences in disciplinary practice and expert opinion on issues such as when pooling of data from different sites is justified, and ways of incorporating contextual and process information into analyses of outcomes and costs. Some choices have reflected disciplinary preferences and standards of practice or analysts' familiarity with the state of statistical and methodological developments relevant to the analysis of the data typically produced in multisite program studies. Economists , for example have tended to be more confident than others about the ability of multivariate methods to tease out program effects in the presence of complex patterns of covariation— including initial differences in case-mix—in large data sets combining outcomes and service utilization data. Those whose backgrounds include the analysis of multisite clinical trials (of new drugs, for example) have tended to emphasize analyses of pooled data and minimize programmatic differences among sites, focusing on controlling design variation and the benefits to be derived from a common data collection protocol.

In contrast, methodologists who have focused specifically on program evaluation and especially quasi-experimentation have tended to recognize the important differences between a multisite trial of a single well-defined program set of demonstrations or knowledge development efforts that focus on a common problem but may vary widely in program design and focus—and therefore in the populations they serve (NIAAA, I&II, CSAP and NIDA prevention programs). In the latter cases, analysts have often focused on producing valid estimates of program effect for each site, recognizing that sites will typically represent considerable variation in program implementation, contextual features, and participant characteristics, even when the "same" program was to have been implemented in each site. So much the more when sites are selected, permitted or even encouraged to develop variations in programs, in the hope that some will show positive effects and that the variation among sites will be instructive. These differences of perspective were reflected in COSP design and research logistics discussions over the first year of the project. In part to accommodate the valid concerns reflected in differing perspectives, we feel that it will be useful to plan for at least two different general approaches to the COSP cross-site analyses. These are 1) *multi-level analysis* of pooled cross-site data, over all sites or several clusters identified by the means discussed above and 2) *meta-analysis* or *quantitative synthesis* of site-level effect-size estimates of program effects, employing standard comparative techniques and site-specific analyses of threats to validity and adjustments for potentially biasing

factors. The two approaches are statistically related but practically distinct, and proceed in very different ways. In the course of implementing either, the cross-site analysts may consider other statistical and analytical techniques that have been proposed by consultants or site analysts (such asGEE, etc.)

The first statistical strategy will makes use of recently developed statistical and analytical tools for multi-level analysis with *hierarchical linear modeling* (Bryk and Raudenbush, 1992). (Related techniques also referred to as random regression (Hedeker and Gibbons) or mixed-model regression.) Hierarchical linear modeling techniques have been developed to deal with the kinds of "nested" data typically generated by multi-site studies, in which observations are nested both within individual participants who provide information at baseline and follow-up and within sites, which differ in multiple ways. In each site individuals are assigned to traditional or consumer-operated service programs, but are members of programs within communities.

The second proposed analysis strategy consists of two steps: the synthesis of "best estimates" of program effects on outcomes for each site derived from site-level analyses; and the subsequent analysis of variations among these estimates in the same way that an explanatory meta-analysis examines the variation in effect sizes derived from separate published studies. This strategy should be regarded as a fall-back or minimally feasible approach, to be employed if or when populations or fidelity data indicate that pooling data from clusters of sites would be misleading. With the small numbers of sites likely to fall into any defined cluster, and the small number of sites overall, it will be of interest and feasible to conduct independent but similarly focused methodological assessment and adjustments for each site prior to computing effect sizes and synthesizing these across sites.

6.6 Statistical Modeling Approaches

A range of statistical methods are appropriate to the planned data collection and hypotheses to be tested. The precise form of the methods to be used and the models to be fitted will depend on the results of the preliminary methodological analyses discussed above. It is, however, possible to indicate the general approaches that will be taken depending on the most likely general patterns, and to indicate some of the issues that will need to be addressed in specifying the analyses in detail. One useful model that may be close to the specification that will be reached for the COSP study is described by Byrk and Raudenbush (1992) and Hedeker and Gibbons (1995) among others. If concerns about missing data are raised by preliminary analyses, as described above, pattern-mixture models (e.g., Hedeker and Gibbons, 1995) may provide an approach to handling some missing-data problems in a natural way within the framework.

In all modeling the general approach to model building will be theory_based, and specifically guided by the pre-specified research questions listed in section 4. Variable selection, the specific form of hypotheses, and the specification of the form of effects will depend on examine predictions in logic model. These include the identification of intervening variables, and the focus on short_term and long_term outcomes in relation to selected waves of data. Rather than testing only a pre-specified model, the ad hoc nature of the summary measures and data reduction will be recognized in carefully observing how data behave when each new predictor is added to a model.

6.6.1 Preliminary Regression_based Approaches

If it is judged that the homogeneity of one or more sets of sites warrants pooled analyses, the preliminary analyses comparing means and distributions on the selected covariates and outcomes in the pooled data will be followed by OLS modeling of differences between COSP and TMHS particpants, statistically controlling for baseline values and the set of covariates determined to be important in the preliminary analyses. Altough the aim is to utilize random regression models, and, if feasible, less fragile methods of adjusting for observed differences between the groups, there is still sufficient methodological interest in the performance of these techniques that we judge it worth comparing their results with those of more traditional multiple regression methods.

6.6.2 Power, Statistical Tests, Effect Size

Current practice in program evaluation emphasizes the expression of statistical results in terms of effect sizes, confidence intervals or meaningful measures of effect, so that progam development and policy decisions based on the analyses will be couched in terms of practical rather than statistical significance (Lipsey 1990, 2000; Sechrest et al; Reichardt and Gollub), it is recognized that the statistical hypothesis-testing paradigm still holds sway in the practice and publication standards of some of the home disciplines of the COSP researchers. It will therefore be appropriate to examine and report conventional hypothesis tests associated with these analyses, as well as the ex-post power of the tests, given the achieved sample sizes.

Preliminary power analyses, making few assumptions about analytical methods or choice of statstical tests, and based on planned recruitment in sites' proposals, are presented in the Appendix.

6.6.3 Hierarchical Linear Models (HLM) and Longitudinal Techniques

The recent develpments in random-regression statistical methods referred to above will permit modeling of pooled data from COSP sites judged sufficiently homogeneous to pool, while accounting for the nesting of individual particpants within sites. In addition, baseline and follow-up observations will be nested within individuals. Both the longitudinal and within-site aspects of the design will be modeled within this hierarchical-linear-models framwork, with an intial focus on the set of questions about the differences between COSPs and TMHS over time, on each of the key outcome measuers identified and measured in the preliminary analyses.

Hypothetically, a multi-site study testing a single program model with particpants selected for reasonable homogeneity across sites would permit tests of the differences between the ouctomes of particpants in innovative and comparison programs, pooling across all sites. When there are many (20 or more) sites implementing the same model, as in many multi-site evaluations of edcuatinal programs, it is even possible to conduct statistical tests of interactions between characteristics of program contexts and individual-level outcomes. But it has been more typical to fund 8-12 sites in each study, and the COSP multisite is at the low end of the range withing recent SAMHSA multisites. Because the COSP programs are so diverse in structure and anticipated participant characteristics, it may be that pooled analyses will focus on clusters of two to four sites (see Section 3). In analyses within these clusters, it will be still be possible to fit longitudinal models that include terms to represent the nesting sites, but there will be no expectation of meaningful statistical tests of cross-site variation. Instead, the variation in patterns of outcomes across sites within clusters and, with more judgment, between clusters, will be examined against the known pattern of differences in site-level characteristics, following the multiple-case-study logic (Yin; Miles and Hubmerman) that has been used in several multi-site studies with similar design features.

6.7 Potential Problems and Issues

Although the initial analyses will follow the "intent to treat" or "intent to serve" model, tracking of participants after assignment with permit tracking of crossover, in which individuals assigned to TMHS elect to participate in COSPs, or those assigned to COSPs decide to stop participating in COSP activities, and continue to receive TMHS services. The latter category may, for comparative analyses, include those who participate only minimally in COSP programs and activities. Secondary analyses will take account of both types of crossovers in "as implemented" analyses.

6.7.1 Differential loss to follow-up

Overall loss to follow_up threatens statistical power and, even if randomization is well-implemented, moderately high follow-up rates are achieved, and follow-up rates are similar in COSP and TMHS groups, differential attrition of participants with characteristic believed or shown to be related to outcomes can bias the results of analyses. A range of analyses fitting the Foster and Bickman (1998) synthesis of attrition analysis techniques developed over the last decade will be conducted. These include sensitivity analyses that the "bracket" range of possible effects of attrition on observed results.

6.7.2 Procedures, Programs, Types of Program

In addition to the analysis of ratings obtained with the COSP/FACIT instrument, the study will have individual-data from the CP on the program procedures actually experienced by participants. At the level of site summary statistics, these data will provide a valuable partial validation of the pattern of differences among programs based on the COSP/FACIT ratings. At the individual level they will also be available to descriptive analyses of individual differences in participation in the programs and inclusion in analyses of individual outcomes.

6.8 Integrating Cost Data

6.8.1 Approaches to CPPOA

There are two basic approaches the the CPPOA analysis strategy: simultaneous CPPOA and sequential CPPOA.

Simultaneous CPPOA. In addition to describing resources, procedures, processes, and outcomes at various levels of specificity (e.g., per consumer per month for each program, per consumer for their entire period of participation, per month for all consumers, for their program overall), a variety of specific statistical tests can be conducted in the CPPOA model to answer questions of interest to researchers, consumer advocates, policy makers, and funders. In general, the guiding assumption is that outcomes are the result of a complex interaction of processes, procedures, resources, consumer variables, and community variables, i.e.,

$$O = f(p, d, r, c, m),$$

where p = processes, d = procedures, r = resources, c denotes a set of consumer variables, and m denotes a set of community variables. Typically, describing this interaction with precision and testing its fit to the data is exceedingly difficult with the relatively small samples sizes available in most social science research (including, possibly, the same size available in this study). Path analyses can be conducted only when data for several hundred to several thousand consumers are available, depending on the number of variables included and the number of relationships hypothesized between variables.

Sequential CPPOA. An alternative approach to analyzing data on resources, procedures, processes, outcomes, and consumer and community variables is to test relationships between these elements sequentially, in three sets of analyses. The sequence of independent variable _> dependent variable analyses are a) processes _> outcomes, b) procedures _> processes, and c) resources -> procedures. For example, first relationships between outcomes of interest and the various observed processes can be tested for significance using analyses that include main effects and specified interactions of multiple independent variables. Those processes which were found to be significantly related to outcomes would be included in the next round of analyses. Those processes that were significantly related to outcomes would be treated as the dependent variables in analyses that included procedures as the independent variables that might be related to those processes. Finally, relationships would be described (if not tested statistically) between the various resources and the procedures found to be significantly related to processes (and, hence, to outcomes).

Consumer and community variables can be entered into determination of relationships between each pairing of causally adjacent variables, as either additional main effects and, if suggested by theory and if possible given sample sizes, as elements of interactions. For example, each outcome can be described (and tested for accuracy of description) as being a function of processes, client variables, community variables, and a composite of the interactions of process, client, and community.

Similarly, each process may in turn be a function of procedures and client and community variables separately and in interaction with procedure variables:

$$P = f(d, c, m)$$

Finally, each procedure in tern very likely requires a specific mixture of resources,

$$D = f(r, c, m)$$

The relationship between resources and procedures seems, at first consideration, to be one that can only be described, not test (statistically or otherwise). Upon reflection, however, an argument can be made for the possibility and even desirability of testing resource _> procedure relationships. Whether a procedure such as cognitive behavior therapy can only be conduced in an office, and by people who have a certain level of training and certification, is an important question and a testable one. From some perspectives, a primary question of the present study examines whether therapeutic procedures can be performed successfully and with positive impact by persons other than those who typically provide such procedures. (An alternative perspective on this is that entirely different procedures are being used in COSs __ ones that traditionally trained, educated, and certified professionals could find particular difficult to carry out.)

Even in a study with two groups and two to three repeated measures, a variety of questions and hypotheses are possible in light of the client variables and various types of programs involved. When the three types of relationships between the four categories of variables are added, the variety of questions possible to ask becomes almost overwhelming. Theory and policy can guide us to limit the number and types of hypotheses advanced.

The most basic question in this study is how much do costs and outcomes increase when COS participation is added to TMHS participation. A more complex set of follow_up questions examines whether the change in outcomes worth the change in costs. Moderators of all these questions include 1) type of COS (three groupings have been identified), 2) type of TMHS (groupings may emerge as TMHSs are examined more closely), 3) degree of consumer participation in the TMHS, and 4) degree of consumer participation in the COS. These moderators can be applied to all of the following breakdowns of these questions in the framework provided by the CPPOA model.

6.8.3 Resource _> Procedure relationships

What changes in a) TMHS personnel, b) COS personnel, c) family and friends' time, d) consumer time, e) TMHS facility, f) COS facility, g) TMHS other direct, h) COS other direct, i) TMHS overhead and j) COS overhead costs occur for i) TMHS, ii) COS when consumers are referred for participate in COS as well as TMHS? Are these changes in cost significantly? ...clinically? Do any observed changes depend on the scale of the TMHS or COS, e.g., does a change become reduced or increased as the size of the TMHS or COS increases? Given the potentially large portion of COS costs that may be volunteered or donated, we will want to ask the questions in this section for 1) monetary expenditures only, 2) volunteered and donated resources only, and 3) the estimated monetary value of 2) plus 1), i.e., total cost as estimated from an opportunity costing perspective.

For programs that measure resource _> procedure relationships in more detail, additional questions can be asked. In particular, differences in cost between the three groups of programs and between individual programs can be examined in terms of what resources are being used for what procedures. We may find that COSs are particularly labor_intensive, relative to TMHSs, but that most of their labor costs are volunteered and some of their facilities are donated __ more, by a far margin, that in TMHS. Drop_in programs may have higher "overhead" costs, while advocacy and other programs may be able to adjust their costs more precisely according to the number of consumers wishing to participate in the program. We may find some procedures that are common to most COSs and few TMHSs, and vice versa. We posit that similar mixtures of different types of resources will be used by programs to implement these procedures. To the degree that programs differ in the resources used to provide similar procedures, we may be able to find the least expensive way to provide a procedure.

6.8.4 Procedure _> Process relationships

Basically, we hypothesize that participation in COSs will lead to more of the changes described as Common (COS) Ingredients elsewhere in this manuscript. We also will examine to which these common ingredient processes occur for consumers participating in TMHSs. These relationships may be moderated by the types of clients and

communities.

6.8.5 **Process _> Outcome relationships**

The degree to which biopsychosocial processes identified in the Common Ingredients list are related to outcomes will be examined by multiple regression analyses. We hypothesize that more experience of the changes identified as common ingredients should be related to more positive changes in the major outcomes, including attainment of monetary cost_savings benefits. These predictions, too, should be moderated by consumer and community variables.

6.8.6 Procedures, Programs, Types of Program

In addition to the analysis of ratings obtained with the COSP/FACIT instrument, the study will have individual-data from the CP on the program procedures actually experienced by participants. At the level of site summary statistics, these data will provide a valuable partial validation of the pattern of differences among programs based on the COSPFACIT ratings. At the individual level they will also be available to descriptive analyses of individual differences in participation in the programs and inclusion in analyses of the relationships between costs for program types and individual outcomes.

6.9 Further Analysis and Planning

Regardless of the which analysis pathway is followed, the final shape of the analyses will be determined in a working group of interested analysts and SC members. The process will seek input from diverse perspectives, including consumers, providers, policymakers, and statistical consultants, and, in particular, those analysts at sites who have made a commitment to involvement in one of the work groups that are established to conduct specified analyses.