USING THE MEDICARE HEALTH OUTCOMES SURVEY TO MANAGE DEPRESSION IN PRIMARY CARE:

A Guide for Medicare + Choice Organizations

AUGUST 15, 2002

HEALTH SERVICES ADVISORY GROUP

1600 E. Northern Avenue, Suite 100 • Phoenix, Arizona 85020

Phone (602) 264-6382 • Fax (602) 241-0757

TABLE OF CONTENTS

USING THE MEDICARE HEALTH OUTCOMES SURVEY TO MANAGE DEPRESSION IN PRIMARY CARE: A GUIDE FOR MEDICARE + CHOICE PLANS

OVERVIEW

INTRODUCTION

The Medicare Health Outcomes Survey (HOS) National Pilot Project on Depression represents a pioneering attempt to use HOS data to help manage one of the most important illnesses that primary care providers confront in their practices. HOS data provide Medicare + Choice Organizations (M+COs) and Quality Improvement Organizations (QIOs) with a unique opportunity to assist primary care providers in recognizing and treating depressive illness in seniors.

The HOS is the first health outcomes measure for the Medicare population in managed care settings. The HOS instrument consists of four components: the SF-36® Health Survey[1] (Ware, Snow, Kosinski, and Gandek, 1993), questions about chronic medical conditions, questions about activities of daily living, and questions designed to collect demographic information. Physical functioning and well being are measured with the Physical Component Summary (PCS) and mental functioning and well being are measured with the Mental Component Summary (MCS) scores, both of which are derived from the SF-36®. The HOS is designed to measure the physical and mental health functioning of Medicare beneficiaries at the beginning and end of a two-year period. Annual baseline collection of HOS data from a randomly selected sample of members from each M+CO began in 1998.

[1] SF-36® is a registered trademark of the Medical Outcomes Trust.

In December 1999, QIOs in Arizona, Florida, and New York began implementing the National Pilot Project. In 2000, QIOs from Michigan, New Mexico, and Ohio joined the project. The analytical goal of the project was to profile the HOS respondents with self-reported low mental health status and provide this information to the M+COs to assist them in identifying current members at risk for depression.

WHY DEPRESSION?

Depression was chosen as the focus of the National Pilot Project because it:

- is prevalent among the senior population (Eaton, 1997; Gurland, Cross and Katz, 1996);
- responds well to treatment (Mulrow et al., 2000);
- is often overlooked by the primary care practitioner (Regier et al., 1993);
- increases the persistence and intensity of symptoms of physical illness (Vaccarino et al., 2001);
- results in increased consumption of resources for physical health care (Koenig and Kuchibhatla, 1998);
- increases the mortality rate for physical illnesses such as myocardial infarction (Frasure-Smith et al., 1995) and cancer (Penninx et al., 1998); and
- increases the mortality rate due to suicide (Conwell, 1996).

For these reasons, the Centers for Medicare & Medicaid Services (CMS) established the National Pilot Project as a partnership among M+COs and QIOs to use the HOS data to identify beneficiaries who are most at risk for depression. Use of the HOS data removes much of the burden of identifying potentially depressed seniors from the primary care providers (PCPs), and allows them to concentrate their efforts on follow up activities with those beneficiaries most likely to suffer from depression. In addition, the Pilot Project has furnished these providers with clinical guidelines and educational materials to assist in the follow up process.

This User's Guide provides M+COs with a written set of procedures for using the HOS to identify and manage their depressed beneficiaries. The recommendations contained in this guide draw on the experiences of the 16 M+COs and 6 QIOs that have participated in the National Pilot Project.

PROJECT PARTICIPANTS

Sixteen plans in six different states participated in the National Pilot Project. Please see Table 1 below for a list of project participants.

TABLE 1 NATIONAL PILOT PROJECT PARTICIPANTS

QIO	M+COs
Health Services Advisory Group (Arizona)	CIGNA Healthcare of Arizona Health Net of Arizona Humana Health Plan Maricopa Integrated Health System
Florida Medical Quality Assurance, Inc.	Health First Health Plan VISTA Health Plan United HealthCare of Florida
MPRO (Michigan)	Health Alliance Plan of Michigan M-CARE
New Mexico Medical Review Association	Lovelace Health Systems
IPRO (New York)	Elderplan Healthfirst HIP Health Plan of New York Univera Healthcare
KePRO (Ohio)	Hometown Health Plan PrimeTime Health Plan

GUIDING PRINCIPLES

The National Pilot Project was designed to minimize the burden on providers by using the HOS data for the initial identification of at risk beneficiaries. Over the course of the National Pilot Project, QIO and M+CO staff requested that the National Pilot staff explore ways to reduce the burden on plan staff even further. In response to this feedback, as well as information from other depression management projects and from the literature, this guide presents a strategy for the efficient use of HOS data to manage depression in primary care. This strategy incorporates the following guiding principles:

1. Use *pre-existing data* to measure the success of depression management activities. In an era of shrinking resources, a depression management strategy that imposes additional data collection burdens will be unlikely to achieve the necessary staff "buy-in" to be successful. Data from the HOS and the plan's existing utilization databases are all that are needed to implement the strategy described in this guide.

2. Measure *process*. The specific interventions that improve depression treatment *outcomes* are well known and well documented in the literature (e.g. Enguidanos and Gibbs, 2001; Kupersanin, 2001; Sherbourne et al., 2001; Unutzer, Rubinstein et al., 2001; Unutzer, Katon et al., 2001). The challenge is not in finding interventions that work, but rather in ensuring that these interventions are fully and correctly implemented (Grol and Grimshaw, 1999). Process measures, rather than outcome measures, should therefore be the focus of the plan's monitoring efforts.

3. Restrict quality improvement (QI) activities to *selected providers*. This reduces the burden on plan resources even further, and ensures that these resources are focused on those providers with a high proportion of beneficiaries who are at risk for depression. Just as it is more efficient to focus depression treatment on those seniors most likely to need the

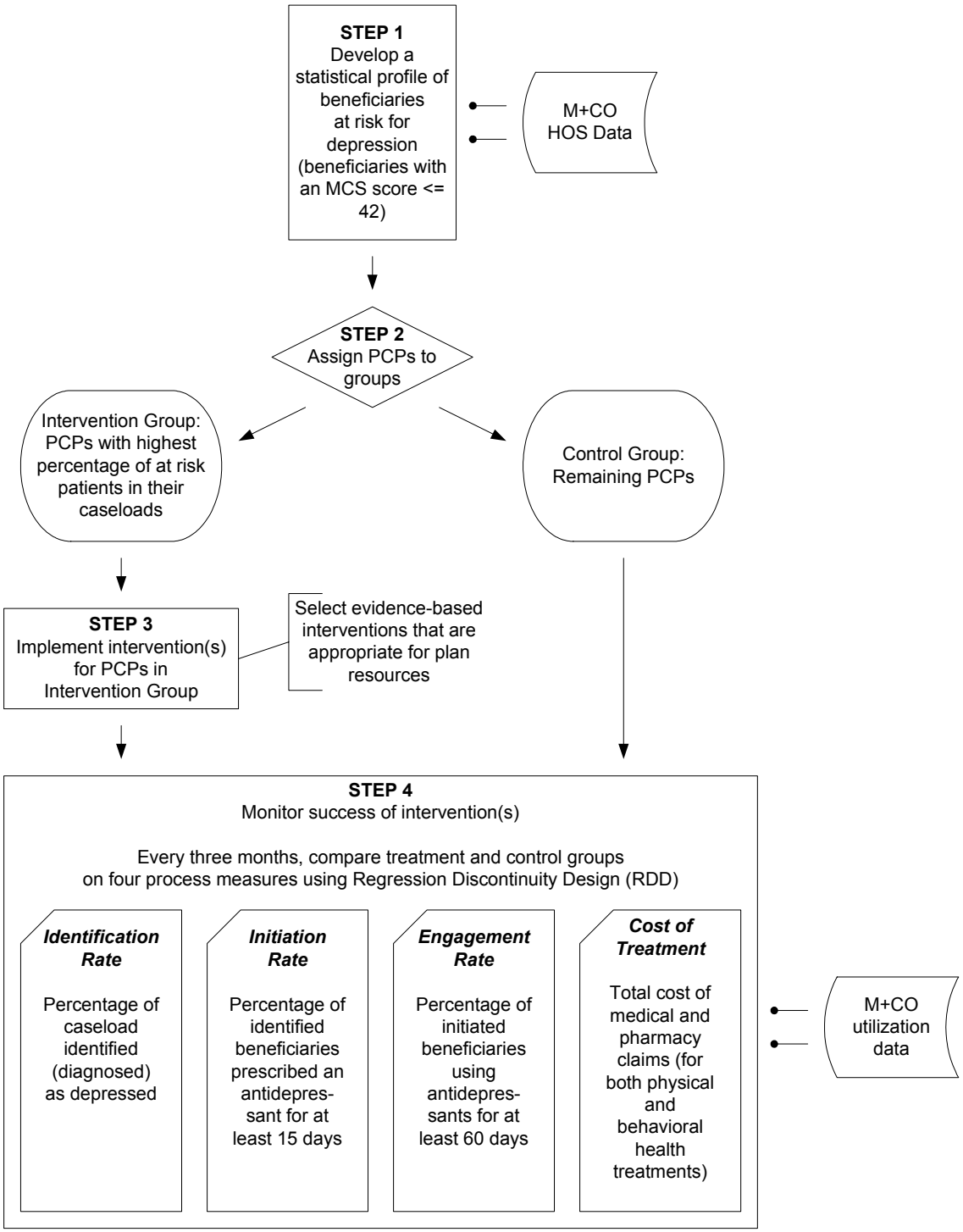
treatment, it is more efficient to focus depression management training on those physicians most likely to need the training. Furthermore, the providers who are not selected for training can be used as a control group to estimate the impact of the QI activities.

4. Estimate the impact of the quality improvement activities using a *comparative evaluation design*. Even if a plan has no systematic depression management effort in place, some depressed seniors will be identified and some will be treated. Using a control group for comparison establishes a valid norm against which to judge the success of any new interventions that are introduced. The evaluation design recommended in this guide is easy to implement, permits plan staff to target their interventions to the providers that need them the most, yet still provides a control group for accurately estimating program impact.

5. *Calculate the cost/benefit* of the chosen QI activities. Identifying and treating depression in the primary care setting improves beneficiary compliance with treatment regimens for physical illnesses, and reduces the costs associated with treatment for these illnesses (Unutzer, Katon et al., 2001). Senior depression can have a substantial impact on overall medical costs; at one Medicare plan, medical costs for depressed seniors were 50% higher than those of seniors without depression ("Seeing 50% higher costs," 1998). The evaluation design described in this guide permits the accurate estimation of the net cost (or, possibly, the net cost savings) of the depression management interventions. This gives the plan's administrators an accurate assessment of the net financial impact of QI activities directed at depression.

Each of these guiding principles is incorporated in the four-step strategy outlined on the following pages. The diagram on the next page provides a visual overview of the strategy. For each step in turn, this guide describes the rationale for the step and the procedures necessary for its implementation. Appendix A uses a question and answer format to provide the user with additional information about these procedures.

STRATEGY FOR USE OF HOS DATA
TO MANAGE DEPRESSION IN PRIMARY CARE



STEP 1: DEVELOP A STATISTICAL PROFILE OF BENEFICIARIES AT RISK FOR DEPRESSION

RATIONALE

The HOS instrument is administered to a randomly selected sample of individuals from each M+CO. For M+COs with Medicare populations of more than 1,000 members, a simple random sample of 1,000 members is selected for the baseline survey. For M+COs with 1,000 members or less, all eligible members are included in the sample for the baseline survey. Each beneficiary completing the HOS can be scored on a measure of mental health status that can be used to determine if the beneficiary is at risk for depression. But, what about the beneficiaries that do not complete the HOS? The plans need some method for determining the risk of depression for the majority of beneficiaries who do not participate in the HOS.

The National Pilot Project addressed this issue by developing a statistical profile of those beneficiaries with an MCS score of less than or equal to 42. Studies have shown that this cutoff score discriminates well between depressed and non-depressed beneficiaries (Ware et al., 1994). This strategy is incorporated into the current procedures.

PROCEDURES

1a. Using the HOS data, generate a statistical model that predicts which beneficiaries are at risk for depression (that is, which beneficiaries are likely to have obtained an MCS score less than or equal to 42, if they had participated in the HOS). This model should utilize both the demographic information and the information on comorbidities available from the HOS. Similar analyses conducted for the National Pilot Project found certain beneficiary characteristics to be most often associated with low MCS scores (see Table 2).

TABLE 2 MOST COMMON RISK FACTORS FOR LOW MCS SCORES	
Comorbid Conditions	*Other Factors*
Diabetes	Dual eligibility
Heart disease	Female gender
Stroke	Age 75 or older

Appendix B describes in detail the analytic strategy that should be used to develop the statistical model.

1b. Once the model has been developed, apply the model to each beneficiary in the plan's enrollment to determine whether or not the beneficiary is at risk for depression. The result is a risk classification for each beneficiary in the plan, not just the proportion of beneficiaries who received and returned the HOS.

STEP 2: ASSIGN PCPS TO GROUPS

RATIONALE

Some PCPs, by virtue of the types of conditions they treat or the characteristics of the beneficiaries in their caseloads, will encounter more potentially depressed beneficiaries than will others. By focusing interventions only on these PCPs, the plan can conserve its resources while directing its depression management efforts where they are most needed.

PROCEDURES

2a. Using the list developed in Step 1b above, calculate the percentage of at risk beneficiaries in each PCP's caseload.

2b. Eliminate all PCPs with small caseloads (fewer than 50 beneficiaries).

2c. Rank order each of the remaining PCPs in terms of the percentage of at risk beneficiaries in their caseload.

2d. Designate the top 25% of PCPs as targets for the plan's depression management strategy. This will be the "Intervention Group," the providers to whom the interventions will be targeted. The remaining providers will serve as the Control Group.

Note: depending on available plan resources, the plan may want to limit the number of PCPs in the Intervention Group even further by targeting, for example, only the top 10% or 15%.

STEP 3: IMPLEMENT QUALITY IMPROVEMENT ACTIVITIES FOR THE PCPS IN THE INTERVENTION GROUP

RATIONALE

The PCP and his or her support staff are the cornerstone of any effort to improve the identification and care of depressed beneficiaries. The interventions aimed at improving depression management depend on the cooperation of the PCPs and their support staff for their success.

The specific interventions, as well as the number of interventions implemented by the plan, will depend on the plan's resources. In recent years, extensive research has clearly identified a number of evidence-based strategies for managing depression. Therefore, there is now a strong consensus among researchers regarding "what works" and what doesn't work in managing depression in primary care. There is no need for plans to re-establish that antidepressant therapy works, or to find more evidence for evidence-based guidelines. Instead, the key challenge for plan staff is to ensure that the interventions they choose to implement are indeed fully and properly implemented.

A detailed summary of effective depression management strategies, including an extensive bibliography, can be found on the Institute for Healthcare Improvement (IHI) web site (IHI, 2002). A brief summary of major findings from the research literature is shown in Table 3. The user is strongly urged to implement only those depression management methods that have sound backing in the research literature.

TABLE 3 WHAT WORKS AND WHAT DOESN'T WORK IN DEPRESSION MANAGEMENT
• Explicit clinical guidelines improve clinical practice and increase the ability to recognize depression, especially if they incorporate patient-specific reminders (Grimshaw and Russell, 1993).
• A simple two question depression screener identifies at risk beneficiaries and minimizes the burden on the PCPs with little loss of screening sensitivity (Whooley et al., 1997).
• Provider education by itself, even in a CME context, has little impact on provider behavior (Davis, 1998) or on the ability to recognize depression (Thompson, Kinmouth, Stevens et al., 2000).
• Interactive provider education, with case studies and role playing, is very effective in changing provider behavior (Cole, Raju, Barrett, Gerrity, and Dietrich, 2000).
• Periodic contacting of patients who initiated antidepressant therapy but who have not refilled their prescriptions helps identify relapsing patients quickly (Katon et al., 2001).
• Use of care managers to monitor the progress of depressed beneficiaries decreases the burden on the PCPs (Sherbourne et al., 2001).
• Training physicians to recognize "red flags" for depression increases the number of potentially depressed beneficiaries that they identify (IHI, 2002).
• A depression registry improves the consistency with which depressed patients are identified and followed (IHI, 2002).

PROCEDURES

3a. Review the literature and select the evidence-based intervention(s) that are most compatible with the plan's organizational structure and available resources.

3b. Establish a quality improvement (QI) committee to oversee implementation of the chosen intervention(s). A multidisciplinary committee that includes a physician champion will improve the committee's opportunity for success.

STEP 4: MONITOR THE SUCCESS OF THE INTERVENTIONS

RATIONALE

The literature has already established the relative efficacy of various interventions in improving outcomes for depressed patients. If appropriate interventions are implemented, improved patient outcomes should follow. Therefore, Step 4 focuses on process measures (management of depression) rather than outcome measures (relief of depressive symptomatology).

A useful model for measuring process implementation comes from the substance abuse treatment literature (McCorry et al., 2000). These authors propose three key indicators, all obtainable from administrative records, as basic process measures for substance abuse management: *Identification* of substance abusers, *Initiation* of treatment for the identified abusers, and *Engagement* of the identified abusers with a course of treatment. This framework is equally applicable to the management of depression. Table 4 describes these measures in detail, as well as an additional indicator to measure costs.

TABLE 4 PROCESS MEASURES FOR MEASURING THE IMPACT OF DEPRESSION INTERVENTIONS

INDICATOR	QUESTION	OPERATIONAL DEFINITION
Identification	For all beneficiaries, how frequently is depression identified?	Percentage of beneficiaries diagnosed with depression
Initiation	For identified beneficiaries, how frequently is depression management initiated?	Percentage of identified beneficiaries prescribed antidepressants for at least 15 days
Engagement	For initiated beneficiaries, how frequently is treatment given for the necessary amount of time?	Percentage of initiated beneficiaries who remain in antidepressant treatment for at least 60 days
Cost	What is the net cost to the plan of the interventions?	Pharmacy and claims costs incurred by the beneficiaries

As operationalized above, all of these process measures are readily available from the plan's utilization data. It is recommended that these indicators be measured on a quarterly basis. That is, if the interventions are implemented at the beginning of January 2003, then it is recommended that the indicators be calculated for utilization data for January through March, April through June, July through September, and October through December. If the interventions are not having the expected impact, quarterly measurements will alert plan administrators to this fact and allow them to make adjustments before the next measurement cycle.

Identification, Initiation, and Engagement should be determined using the same diagnostic codes and pharmacy codes as used in the National Committee for Quality Assurance (NCQA) HEDIS® 2002 measures for antidepressant medication management (NCQA, 2001).

The Initiation indicator specifies "…at least 15 days" in recognition of the fact that many PCPs prescribe antidepressants for their patients for a limited period of time for reasons other than depression.

The process indicators listed in Table 4 do not include psychotherapy, only antidepressant therapy. In keeping with the principle of using readily available data, psychotherapy has been eliminated because in many plans it is not consistently documented in the claims database. If the plan keeps comprehensive data on psychotherapy visits, then the Initiation indicator can be modified to "Number of beneficiaries who remain on antidepressants for at least 15 days and/or receive at least one psychotherapy visit for depression." Similarly, the Engagement indicator can be modified to "Number of beneficiaries who remain on antidepressants for at least 60 days and/or receive at least two psychotherapy visits for depression." This will give a more accurate measure of Initiation and Engagement than antidepressant therapy alone.

The final process indicator listed in Table 4 is Cost. Management of depression, while important, is only one of many priorities competing for the attention of plan administrators. In order to make an informed decision about continuing the selected interventions, plan administrators need to know the net cost of these interventions. To arrive at an accurate estimate of cost, the beneficiaries' *total* claims and pharmacy costs need to be calculated. The literature

suggests that the interventions will increase the cost of behavioral health treatment, but decrease the cost of physical health treatment (Von Korff et al., 1998). Cost subcategories (behavioral healthcare versus non-behavioral healthcare, type of care setting, etc.) can also be analyzed separately to better understand where the greatest impacts are occurring.

To assess the impact of the interventions, the four process indicators are calculated for each physician and compared for the Intervention and Control Groups. This means that the physician, not the beneficiary, is the unit of analysis. As described in this guide, the Intervention and Control Groups have *not* been made equivalent through random assignment of physicians to the groups. Rather, group assignments have been made on the basis of the percentage of the physician's caseload deemed at risk for depression. If the physician has a high percentage of at risk beneficiaries in his or her caseload, then that physician is assigned to the Intervention Group, otherwise the physician is assigned to the Control Group.

The practical reason for assigning physicians to groups in this manner is to assure that the beneficiaries that can most benefit from depression interventions will be more likely to receive these interventions. Although the groups are not equivalent, the differences between the two groups can be statistically corrected for by analyzing the data with a Regression Discontinuity Design (RDD). The result is an unbiased, valid measure of the treatment effect (Williams, 1990).

It should be noted that the RDD is a much stronger evaluation design than the more commonly used Nonequivalent Groups Design (NEGD). The NEGD is used to compare preexisting groups for which the assignment process is not controlled by the evaluator and therefore is not fully known to the evaluator. When the NEGD is used, the evaluator typically attempts to adjust the group differences for variables suspected of explaining the differences. Since the assignment process is not under control, the variables that actually explain these differences can never be fully known or measured, and the evaluator can never be sure that the adjustment process has yielded an accurate measure of program impact (Trochim, 2002).

By contrast, with the RDD the evaluator has complete control of the group assignment process, threats to internal validity are minimal, and the measure used to assign subjects to groups permits the calculation of an unbiased measure of the program's impact (Trochim, 2002). In the case described in this guide, the measure used to assign physicians to groups is the percentage of at risk beneficiaries in the physician's caseload.

The RDD therefore provides a practical yet statistically sound approach to measuring the effect of the intervention(s). Further details, and a step-by-step procedure for analyzing the process indicators using the RDD, are outlined in Appendix C.

PROCEDURES

4a. For the time period in question, create a data file for each physician in the study containing group membership (Intervention or Control), and each of the four process measures, calculated for the physician's caseload.

4b. Compare each of the four process indicators (Identification, Initiation, Engagement, and Cost) across the Intervention and Control Groups, using the RDD analytic procedure described in Appendix C.

4c. If no significant differences are found at the end of the first quarter, the intervention(s) should be reviewed and adjusted or augmented accordingly. The analysis should then be repeated at the end of the second quarter to assess the success of these modifications.

CONCLUSIONS

The recommendations found in this guide build upon the methodology tested during the National Pilot Project. In response to feedback from participating QIOs and M+COs, as well as the experience of other depression management projects, this guide presents a refined and simplified version of the National Pilot Project methodology.

The methodology proposed here provides M+COs with an efficient, accurate, and adaptable approach to measuring the success of their depression management activities. Efficiency is achieved by utilizing existing data sources, accuracy is achieved by using the RDD analytical approach, and adaptability is achieved by the ability to adjust cutoff scores according to available plan resources. In fact, the approach suggested here can be applied not only to depression management projects, but to many other kinds of QI projects as well.

APPENDIX A: QUESTIONS AND ANSWERS

Q. In contrast to the evaluation process recommended in this guide, the approaches recommended by both CMS' Quality Improvement System for Managed Care (QISMC) and IHI do not incorporate a Control Group. Is a Control Group really necessary?

A. The Control Group provides plan administrators with a measure of the level of depression management activity that is occurring in the absence of the formal intervention(s). Without such a measure, the estimate of impact due to the intervention(s) is likely to be inflated.

Q. In May, 2002, the US Preventive Services Task Force (USPSTF) recommended "…screening adults for depression in clinical practices that have systems in place to assure accurate diagnosis, effective treatment, and followup" (USPSTF, 2002). Since the User's Guide recommends that the plans focus their efforts on a subset of their PCPs, don't the guide's suggestions run counter to this recommendation?

A. As the USPSTF acknowledges, routine screening of all adults for depression can lead to a substantial number of false positives which will incur unnecessary work up and treatment costs. A typical depression screening instrument has a specificity of 80%, meaning that for every 100 non-depressed patients who are screened with the instrument, 20 of these patients will screen positive for depression despite their non-depressed status. If such a screener is applied to a population of 1,000 beneficiaries, 900 of whom are not depressed, the result would be 180 false positives that will needlessly consume staff time and resources. In populations where the prevalence of depression is known in advance to be high, such as post-AMI patients or diabetics, there will be fewer such false positives. In Part D of USPSTF's recommendations, the USPSTF specifically recommends against routinely providing the service to asymptomatic patients.

Q. Does the Health Insurance Portability and Accountability Act (HIPAA) pose any obstacles to implementing the recommendations in this guide?

A. The current HIPAA guidelines do not prevent the collection of identified beneficiary information in support of health care operations. Please note that the HIPAA standards are still evolving and should be monitored frequently for changes. HIPAA updates can be found at http://www.cms.gov/hipaa/hipaa2/default.asp (last accessed on August 15, 2002).

Q. Our plan is already collecting HEDIS measures for antidepressant medication management. Wouldn't it be preferable to use these indicators instead of the process measures used in this guide?

A. Collection of the HEDIS measures only occurs once per year. Unless it is feasible for the plan to calculate these measures on a quarterly basis at a physician level, it is recommended that the indicators proposed by this guide be collected to evaluate the success of the interventions.

Q. Participating plans in the National Pilot Project submitted utilization data to their QIO and the QIO analysts linked the utilization data to the HOS data. Do the procedures in this guide require plans to submit any utilization data to their QIO?

A. The procedures described in the guide can be fully implemented by an M+CO without involvement of the plan's QIO, and do not require linking HOS data with utilization data at the beneficiary level. Instead, the statistical profile developed from the HOS data is applied to each beneficiary (not just the beneficiaries who participated in the HOS) to determine if that beneficiary is at risk for depression. These results are then aggregated for each physician to determine which physicians have the highest percentage of beneficiaries at risk for depression. These are the physicians that will be targeted with the interventions. After the interventions have been implemented, utilization data are

then aggregated to the physician level to measure the four process indicators. None of these steps requires that the HOS and utilization data be linked at the beneficiary level. However, in order to obtain results based upon more recent data, M+COs may want to request that their QIOs conduct the statistical modeling for them, using the more recent data to which only QIOs have access.

Q. At our plan, we have a very active disease management program for depression. Our case managers may potentially work with any one of our physicians on depression issues. How can the proposed methodology be applied to our situation?

A. How best to apply the User's Guide to your situation depends on the intervention(s) you are interested in assessing. If you wish to measure the impact of an intervention that is distinct from the case management program, then you can consider your case management activities to be part of your physicians' standard depression management activities. In this situation physicians in both the Intervention and Control Groups can continue to interact with the case managers, and the RDD results will tell you if the intervention is affecting your physicians' performance over and above the impact of the case management program. If you are interested in assessing the impact of the case management program itself, then in order to apply the methodology in this guide, some of the physicians must be assigned to a Control Group that does not use case management services. This can be accomplished with minimal disruption to the M+CO's current operations by assigning those physicians who use case management services the least to the Control Group. Usage of case management services then becomes the measure from which the cutoff score for group assignment is derived.

Q. Once we are satisfied that our intervention is working well, we will want to "roll it out" to the other physicians in the Control Group. Do we need to continue using the Control Group indefinitely?

A. If the evaluation results indicate that the intervention is having an impact, then data collection may be terminated and the intervention can be introduced to the physicians who belong to the Control Group. If the intervention is substantially modified or augmented with another intervention, then it is recommended that the M+CO restrict the remodeled intervention(s) to a subset of physicians, and recommence the evaluation procedures.

Q. Whom can we contact with questions about the HOS or the information in this guide?

A. Please see Appendix E for technical support information.

APPENDIX B: STRATEGY FOR ANALYSIS OF HOS RESULTS

ANALYSIS OF HOS DATA

From the HOS *Cohort IV Baseline* data, create a beneficiary level data file with data elements that capture key comorbidities, demographic characteristics, and the MCS score. The key comorbidities are those that the literature has shown to be relatively prevalent among seniors and strongly linked to depressive symptoms. Table 5 lists the data elements that should be captured and the coding scheme that should be used for each element.

TABLE 5 DESCRIPTION OF FILE GENERATED FROM HOS DATA

HOS Question Number and Description	Coding Scheme
23. Congestive heart failure	Yes = 1, No = 0
24. A myocardial infarction or heart attack	Yes = 1, No = 0
25. Other heart condition such as problems with heart valves or the rhythm of your heart beat	Yes = 1, No = 0
26. A stroke	Yes = 1, No = 0
29. Arthritis of the hip or knee	Yes = 1, No = 0
30. Arthritis of the hand or wrist	Yes = 1, No = 0
32. Diabetes, high blood sugar, or sugar in the urine	Yes = 1, No = 0
33. Any cancer	Yes = 1, No = 0
46. Age	Less than 75 = 0, 75 or Older = 1
47. Gender	Male = 0, Female = 1
Medicaid status	Medicaid = 1, Not Medicaid = 0
MCS score	Less than or equal to 42 = 1, Greater than 42 = 0

Using the data file created above, run a logistic regression analysis in which the recoded MCS score (dependent variable) is regressed upon the remaining variables (independent variables) in the data file. Verify that this model has sufficient explanatory power by examining the goodness of fit statistics.

ANALYSIS OF UTILIZATION DATA

From claims and enrollment data, create a data file that contains the following information for each beneficiary.

TABLE 6 DESCRIPTION OF FILE GENERATED FROM BENEFICIARY DATA

Data Element	Equivalent in HOS file (from Table 5)	Coding Scheme
Presence or absence of heart failure	Q23	ICD-9-CM diagnosis code of 402.01, 402.11, 402.91, 404.01, 404.11, 404.91, or 428.x = 1, otherwise = 0
Presence or absence of acute myocardial infarction	Q24	ICD-9-CM diagnosis code of 410.xx, excluding 410.x2 = 1, otherwise = 0
Presence or absence of atrial fibrillation	Q25	ICD-9-CM diagnosis code of 427.31 = 1, otherwise = 0
Presence or absence of ischemic stroke or TIA	Q26	ICD-9-CM diagnosis code of 362.44, 433.xx, 434.xx, 435.0, 435.1, 435.3, 435.8, 435.9, or 436 = 1, otherwise = 0
Presence or absence of arthritis	Q29, Q30	ICD-9-CM diagnosis code of 715.9 or 720.0 = 1, otherwise = 0
Presence or absence of diabetes	Q32	ICD-9-CM diagnosis code of 250.xx, 357.2, 362.01, 362.02, or 366.41 = 1, otherwise = 0
Presence or absence of cancer	Q33	ICD-9-CM diagnosis code of 140 through 239 = 1, otherwise = 0

TABLE 6 DESCRIPTION OF FILE GENERATED FROM BENEFICIARY DATA (CONTINUED)

Data Element	Equivalent in HOS file (from Table 5)	Coding Scheme
Age	Q46	Less than 75 = 0, 75 or older = 1
Gender	Q47	Male = 0, Female = 1
Medicaid status	Medicaid status	Medicaid = 1, Not Medicaid = 0
Predicted risk status	Estimated with predictive model	Predicted probability > .50 = 1, Predicted probability <= .50 = 0
Physician identifier	(none)	Physician code

The next step is to apply the predictive model that was developed from the HOS data file to the beneficiary data file. For each beneficiary in the beneficiary data file, use the predictive model to estimate the probability that this beneficiary's MCS score is less than or equal to 42. If the estimated probability is greater than or equal to 50%, classify the beneficiary as at risk for depression and code the "At Risk" variable as 1. If the estimated probability is less than 50%, classify the beneficiary as not at risk for depression and code the "At Risk" variable as 0.

Once the "At Risk" predictions have been generated for each beneficiary, group the beneficiary file by physician. Drop from the file any physicians with less than 50 beneficiaries in their caseload. For each physician remaining in the file, calculate the percentage of beneficiaries in that physician's caseload who have an "At Risk" variable value of 1. Finally, identify the top 25% of physicians in terms of this percentage. These physicians will be placed in the Intervention Group, and the remaining 75% of physicians will be placed in the Control Group. If plan resources are not sufficient for engaging 25% of physicians, then the top 15%, or top 10%, etc., of physicians can be placed into the Intervention Group.

APPENDIX C: STRATEGY FOR MONITORING IMPACT OF INTERVENTIONS

BACKGROUND

Although not frequently used in impact assessment, the RDD is well suited for this purpose. From a methodological point of view, the conclusions drawn from an RDD design are comparable in validity to the conclusions drawn from randomized experiments. Furthermore, from the administrator's perspective, the RDD is compatible with the goal of getting the intervention to those most in need—a consideration which randomized experiments cannot accommodate (Trochim, 2002). To date, the RDD has been most often used in evaluations of educational programs. For an example of an RDD from the health care arena, see Cappelleri and Trochim (2000).

The defining characteristic of the RDD is that individuals' scores on some measure are used to assign those individuals to groups that will be compared on certain indicator(s) of program success. By contrast, in a randomized design individuals are assigned to groups based on a random process. In the RDD, individuals falling below a certain cutoff score are assigned to one of the groups, while the remaining individuals are assigned to the other group.

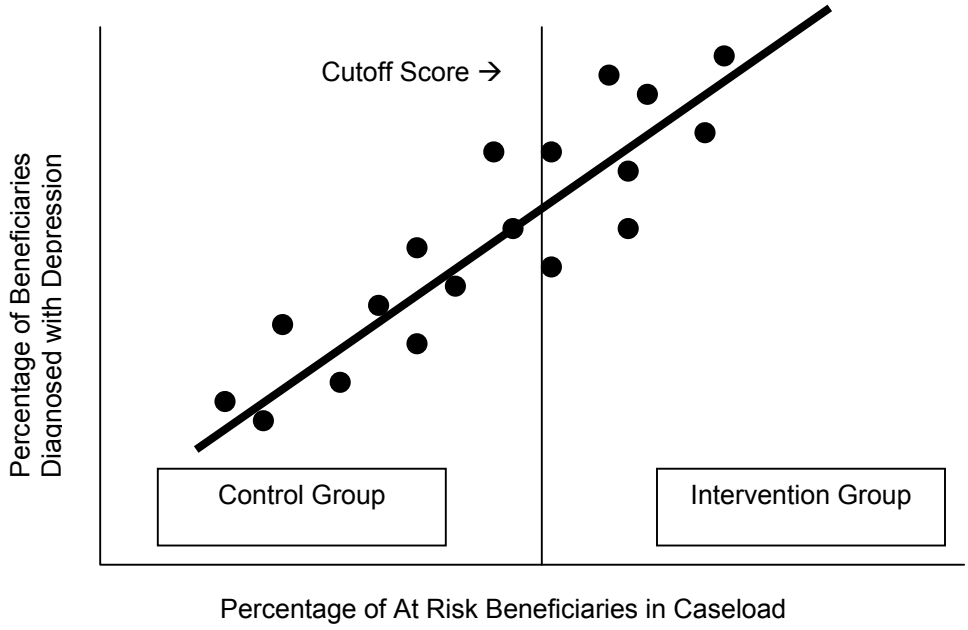
AN EXAMPLE

In terms of the evaluation design presented in this guide, physicians are assigned to either the Intervention or Control Group. The measure used to assign a physician to one of the Groups is the percentage of beneficiaries in the physician's caseload that are predicted to be at risk for depression. The physicians can be sorted in terms of this percentage, and each physician can be assigned a percentile score that indicates where they fall in this listing. If a physician's percentile score is 75% or above, meaning that the physician falls into the top quarter of all physicians in terms of his/her percentage of at risk beneficiaries, then that physician is assigned to the Intervention Group. If the physician's percentile score falls below 75%, then that physician is assigned to the Control Group. The percentile score of 75% is the cutoff score for

determining assignment to groups. Of course, a different cutoff score could be used instead (e.g., 80% or 85%). What is important is that the cutoff score always be used to determine group assignments.

If the interventions have no impact on the Identification process indicator, what would the data look like? The results might look like the following:

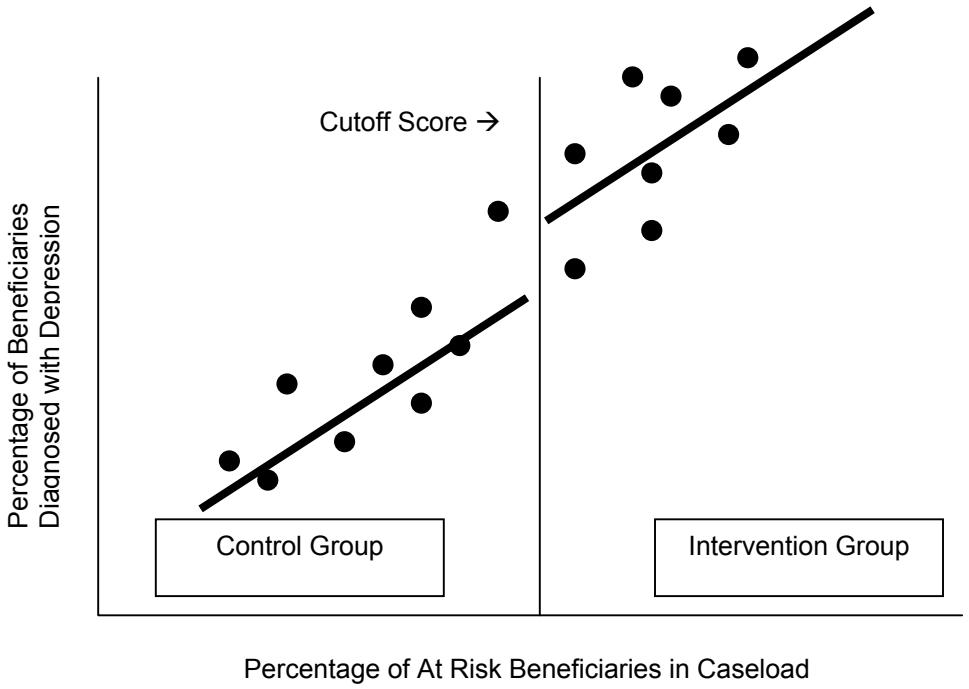
Figure 1 RDD Results – No Impact of Interventions



In the chart above, each dot represents an individual physician's score on the assignment measure (X Axis) and the Identification process measure (Y Axis). Since the interventions did not have an impact, the relationship between the cutoff measure and the Identification process measure looks the same in both the Control and Treatment Groups. The conclusion from the chart is that physicians with higher percentages of at risk beneficiaries are identifying larger numbers of depressed beneficiaries, and furthermore, the interventions have had no impact on this basic relationship (shown by the regression line).

Now suppose that the interventions *did* have a positive impact on the Identification process indicator. That is, the interventions have succeeded in improving the ability of the physicians to identify depressed beneficiaries. What would the data look like?

Figure 2 RDD Results – Positive Impact of Interventions



If the interventions have had a positive impact, then there will be a discontinuity in the regression lines at the cutoff score. This is what gives the Regression Discontinuity Design its name. The vertical distance between the two regression lines at the cutoff score is the measure of program impact.

VALIDITY OF THE RDD

As long as the assignment process dictated by the cutoff score is strictly adhered to, the RDD provides an unbiased measure of program impact. The only way that selection or maturation processes can threaten the validity of the RDD design is if they induce a discontinuity in the

relationship between the assignment score and the measure of program impact that happens to coincide precisely with the cutoff point—a very unlikely scenario. Similarly, regression to the mean does not weaken the validity of this design because the regression to the mean that will occur will be continuous across the entire range of scores and not vary at the cutoff point.

A more significant potential threat to the validity of the conclusions from an RDD study is the possibility that the relationship between the assignment score and the measure of program impact is nonlinear. This can be dealt with by fitting a nonlinear model to the data. See the Trochim reference (2002) for further details regarding the fitting of nonlinear models in RDD analysis.

STATISTICAL ANALYSIS

Statistical analysis of the RDD is easily accomplished using standard regression software. First, subtract the cutoff score from the assignment score to create a modified assignment score:

$$X_{i\,mod} = X_i - X_c$$

Where X_i is the assignment score for the *i*th physician and X_c is the cutoff score.

Second, examine the plot of the assignment and outcome scores visually to determine if the relationship is nonlinear. If the relationship is linear, then fit the model:

$$Y_i = B_0 + B_1X_{i\,mod} + B_2D + B_3X_{i\,mod}D + e_i$$

Where

Y_i = the outcome score for the *i*th physician
$X_{i\,mod}$ = the modified assignment score
D = dummy variable for Group (1 = Intervention, 0 = Control)

B_0 = the coefficient for the intercept

B_1 = the coefficient for the modified assignment score

B_2 = the coefficient for the Group variable

B_3 = the coefficient for the interaction between the modified assignment score and the difference between the Intervention and Control Groups and

e_i = the residual for the *i*th physician

If the relationship is nonlinear, then consult the Trochim reference (2002) for procedures for adding nonlinear terms to the above model.

If the coefficient B_2 for the dummy variable D is statistically significant, then the interventions have had a significant impact on the outcome variable Y, and B_2 provides an unbiased estimate of the impact.

APPENDIX D: REFERENCES

Cappelleri, J.C. & Trochim, W. (2000). Cutoff designs. In: Chow, Shein-Chung (ed.) *Encyclopedia of biopharmaceutical statistics,* 149-156. New York, New York: Marcel, Dekker.

Cole, S., Raju, M., Barrett, J., Gerrity, M., & Dietrich, A. (2000). The MacArthur Foundation Depression Education Program for Primary Care Physicians: Rationale, Participants' Workbook, and Facilitators' Guide. *General Hospital Psychiatry*, *22,* 299-358.

Conwell, Y. (1996). *Diagnosis and treatment of depression in late life*. Washington, DC: American Psychiatric Press.

Davis, D. (1998). Does CME work? An analysis of the effect of educational activities on physician performance or health care outcomes. *International Journal of Psychiatry and Medicine, 28(1)*, 21-39.

Eaton, W. W. (1997). Natural history of Diagnostic Interview Schedule/DSM-IV major depression. The Baltimore Epidemiologic Catchment Area follow-up. *Archives of General Psychiatry, 54*, 993-999.

Enguidanos, S. & Gibbs, N. E. (2001). Geriatric case management: an opportunity for improved depression detection and treatment. *Care Management, 10*, 18-23.

Frasure-Smith, N., Lesperance, F., & Talajic, M. (1995). Depression and 18-month prognosis after myocardial infarction. *Circulation, 91,* 999-1005.

Grimshaw, J.M. & Russell, I. T. (1993). Effects of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet, 342(8883)*, 1317-22.

Grol, R. & Grimshaw, J. M. (1999). Evidence-based implementation of evidence-based medicine. *Joint Commission Journal on Quality Improvement, 25*, 503-513.

Gurland, B. J., Cross, P. S. & Katz, S. (1996). Epidemiological perspectives on opportunities for treatment of depression. *American Journal of Geriatric Psychiatry*, *4(Suppl. 1)*, S7-S13.

Institute for Healthcare Improvement. (2002). Depression: changing practice, changing lives. Available at: http://www.ihi.org/collaboratives/Depression_Apr2002.pdf. (Last accessed on July 1, 2002).

Katon, W. J., Rutter, C., Ludman, E. et al. (2001). A randomized trial of relapse prevention of depression in primary care. *Archives of General Psychiatry, 58(3)*, 241-247.

Koenig, H. G. & Kuchibhatla, M. (1998). Use of health services by hospitalized medically ill depressed patients. *American Journal of Psychiatry, 155*, 871-877.

Kupersanin, E. (2001). Depression program promising in primary care settings. *Psychiatric News, 36(21)*, 14.

McCorry, F., Garnick, D., Bartlett, J., Cotter, F. & Chalk, M. (2000). Developing performance measures for alcohol and other drug services in managed care plans. *Joint Commission Journal on Quality Improvement, 26(11)*, 633-643.

Mulrow, C. D., Williams, J.W., Chiquette, E. et al. (2000). Efficacy of newer pharmacotherapies for treating depression in primary care patients. *American Journal of Medicine, 108(1)*, 54-64.

National Committee on Quality Assurance. *HEDIS® 2002 Technical Specification.* 2001: 103-109.

Penninx, B.W. et al. (1998). Chronically depressed mood and cancer risk in older persons. *Journal of the National Cancer Institute, 90,* 1888-1893.

Regier, D. A., Narrow, W. E., Rae, D. S. et al. (1993). The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Archives of General Psychiatry, 50(2)*, 85-94.

Seeing 50% higher costs in depressed seniors, plan begins guideline to manage disorder. (1998). *Public Sector Contracts Report, January(4)*: 13-15.

Sherbourne, C. D., Wells, K. B., Duan, N. et al. (2001). Long-term effectiveness of disseminating quality improvement for depression in primary care. *Archives of General Psychiatry, 58(7)*, 696-703.

Thompson, C., Kinmouth, A., Stevens, L. et al. (2000). Effects of a clinical practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. *Lancet, 15,* 185-91.

Trochim, W. M. (2002). The Research Methods Knowledge Base, 2nd Edition. Available at: http://trochim.human.cornell.edu/kb/index.htm (last accessed on July 3, 2002).

Unutzer, J., Katon, W., Williams, J. W. et al. (2001). Improving primary care for depression in late life: the design of a multicenter randomized trial. *Medical Care, 39(8)*, 785-799.

Unutzer, J., Rubenstein, L., Katon, W. et al. (2001). Two-year effects of quality improvement programs on medication management for depression. *Archives of General Psychiatry, 58(10)*, 935-942.

U.S. Preventive Services Task Force. (2002). Screening for depression. Recommendations and rationale. Agency for Healthcare Research and Quality, Rockville, MD. Availabe at: http://www.ahrq.gov/clinic/3rduspstf/depression/depressrr.htm. (Last accessed on July 2, 2002).

Vaccarino, V., Kasl, S. V., Abramson, J. et al. (2001). Depressive symptoms and risk of functional decline and death in patients with heart failure. *Journal of the American College of Cardiology, 38(1),* 199-205.

Von Korff, M., Katon, W., Bush, T. et al. (1998). Treatment costs, cost offset, and cost-effectiveness of collaborative management of depression. *Psychosomatic Medicine, 60(2)*, 143-149.

Ware, J. E., Snow, K. K., Kosinski, M. & Gandek, B. (1993). *SF-36® Health Status Survey Manual and Interpretation Guide*. Boston: The Health Institute, New England Medical Center.

Ware, J. E. et al. (1994). *SF-36® physical and mental health summary scales: A user's manual.* Boston: The Health Institute, New England Medical Center.

Whooley, M. A., Avins, A. L., Miranda, J. et al. (1997). Case-finding instruments for depression: Two questions are as good as many. *Journal of General Internal Medicine, 12,* 439-445.

Williams, S. V. Regression-discontinuity design in health evaluation. (1990). In Sechrest, L., Perrin, E., and Bunker, J. In *Research methodology: strengthening causal interpretations of nonexperimental data*. (pp. 145-149). Agency for Health Care Policy and Research.

APPENDIX E: TECHNICAL SUPPORT

For assistance or questions regarding the Medicare Health Outcomes Survey (HOS) or this User's Guide, contact the Information and Technical Support Telephone Line at 888-880-0077. Or, you may e-mail azpro.hos@sdps.org.