EVALUATING THE EFFECTIVENESS OF HOME HEALTH AS A DISEASE MANAGEMENT STRATEGY

Ariel Linden, DrPH, MS

Home health (HH) is considered by many to be complementary to existing disease management (DM) programs, or even suitable as a stand-alone DM intervention. The advantage is thought to be with the face-to-face interaction, in contrast to the standard DM telephonic interview. However, much of the literature appears to indicate that telecommunication is as successful as face-to-face contact (typically referred to as "usual care") for administering health surveys, providing counseling, changing health behaviors, and monitoring physiologic functioning. Given the desire to expand into the area of DM, HH agencies will need to identify and demonstrate areas in which they have a clinical and competitive edge over traditional DM models. This article describes and provides examples of three research designs that may assist the HH industry in evaluating their effectiveness in delivering DM services: the randomized controlled trial, the regression-discontinuity design, and case-control matching on the propensity score.

Home health (HH) is considered by many to be complementary to existing disease management (DM) programs, or even suitable as a stand-alone DM intervention. The prevailing thought is that HH holds an advantage over traditional DM telephonic interventions because of the face-to-face contact with patients. During the home visit, agents conduct a comprehensive assessment of the patient's living conditions and provide an objective appraisal of their functional status (physical and mental). In addition, HH caregivers identify and remove barriers and environmental irritants and ensure that patients correctly follow their treatment regimens. Taken at face value, it appears that these activities cannot be adequately replicated by a telephone call.

To make a business case for HH as a suitable addition to existing DM interventions or as an alternative DM model altogether, it must first prove to be at least as effective in the delivery of DM interventions as well as being more cost-effective than standard DM methods. While telephonic and electronic communications can reach hundreds of individuals daily at relatively low cost, HH is subject to time and personnel constraints. The question is whether these fewer HH visits result in better results than their telephonic counterparts?

The results gleaned from the health services research literature poses uncertainties about the advantages of face-to-face interactions over that of telecommunicating with patients.

Several studies have demonstrated that there is no significant difference in how participants respond to health surveys when administered face-to-face versus telephonically (Anie, Jones, Hilton, & Anderson, 1996; Burke, Roccaforte, Wengel, Conley, & Potter, 1995;

Key Words: home health; disease management; randomized controlled trial; regression-discontinuity; propensity score

Author's Note: *Please direct correspondence to Ariel Linden, DrPH, MS, President, Linden Consulting Group, 6208 NE Chestnut Street, Hillsboro, OR 97124; e-mail: alinden@lindenconsulting.org*

Home Health Care Management & Practice / Month 2005 / Volume 15, Number X, 1-7
DOI: 10.1177/1084822305281972

Chwalow, Costagliola, Stern, Mesbah, & Eschwege, 1989; Galan, Rodriguez-Artalejo, & Zorrilla, 2004; Galasso, Panico, Celentano, & Del Pezzo, 1994; Greenfield, Midanik, & Rogers, 2000; Korner-Bitensky, Wood-Dauphinee, Siemiatycki, Shapiro, & Becker, 1994; Lyu et al., 1998; Midanik, Greenfield, & Rogers, 2001; Pridemore, Damphousse, & Moore, 2005; van Wijck, Bosch, & Hunink, 1998), suggesting that maybe there is no advantage to having HH agents performing face-to-face health assessments. That said, one study by Weinberger et al. (1994) found that patients who were elderly taking more than five medications had substantially different scores on the SF-36 Health Related Quality of Life tool (Ware & Sherbourne, 1992) when completing the survey once by telephone and again in a face-to-face interview 1 month later. A second study by Donovan, Holman, Corti, and Jalleh (1997) found significant differences response distributions for health surveys using face-to-face household interviews and telephone interviews. They found that the telephone sample reported significantly lower rates of smoking and unsafe alcohol consumption, had a significantly higher proportion of individuals who were in Prochaska's (Prochaska & Velicer, 1997) "action" stage of change for several health behaviors, and had a significantly greater recall of health messages than in the face-to-face group.

Effectiveness of telephonic interventions for patients with chronic illness is typically compared with "usual care" modalities (i.e., physician or caregiver face-to-face interaction). Riegel et al. (2002) found that telephonic nurse case management reduced the hospitalization rate, hospital days, and the number of multiple readmissions for patients with congestive heart failure (CHF) compared to patients not receiving the calls.

A series of studies in the Veterans Administration (VA) have compared the impact of automated telephone disease management (ATDM) calls with telephone nurse telephonic follow-up in improving outcomes for people with diabetes compared to usual care (Piette, McPhee, Weinberger, Mah, & Kraemer, 1999; Piette, Weinberger, Kraemer, & McPhee, 2001; Piette, Weinberger, & McPhee, 2000). Universally, patients in the interventions groups reported few symptoms of depression, greater self-efficacy, and fewer bed days than those in the usual care cohorts.

Telephonic interventions appear to be quite successful as a counseling tool. Hunter (2000) noted that telephone support showed promise of offering cost-effective care for persons with psychiatric disabilities. Badger et al. (2005) found that telephone interpersonal counseling for women with breast cancer achieved better symptom management of depression, fatigue, and quality of life compared to a usual care attentional control group. Mohr et al. (2000) found that an 8-week telephone-administered cognitive-behavioral therapy program was successful in reducing depressive symptomatology in patients with multiple sclerosis compared to those receiving usual care.

The use of telemonitoring of patients in lieu of face-to-face care has also been shown to be effective. Sparks, Shaw, Eddy, Hanigosky, & Vantrese (1993) found that a home exercise program using trans-telephonic exercise monitoring for patients in cardiac rehabilitation was just as successful as hospital-based programs. Friedman et al. (1996) evaluated the effect of automated telephone patient monitoring and counseling on patient adherence to antihypertensive medications and on blood pressure control compared to controls receiving usual care. The authors showed that the intervention group improved their adherence to antihypertensive medication and demonstrated a greater decrease in diastolic blood pressure than did controls.

As the aforementioned studies suggest, the value of HH as a DM intervention is far from ensured (the focus is not on HH as a provider of hands-on care but in a role as a DM service provider). Therefore, the task that lies ahead for HH is to prove its worth as an addition to, or in lieu of, a telecommunication process for delivering DM services. That said, the purpose of this article is to provide readers with direction in how best to assess the role of HH in DM. Several research designs are presented with examples, and a discussion of each model's strengths and weaknesses.

THE RANDOMIZED CONTROLLED TRIAL

Any design chosen to evaluate HH program effectiveness is subject to exposure to factors that may influence the results, outside the intervention itself. These factors, termed *biases*, may cause the results to look better or worse than what was actually achieved by the intervention. Consequentially, these potential sources of bias must be identified and controlled for when developing an evaluation strategy.

The primary threat to validity of any health care study is that of *selection bias*. This bias suggests that there are some fundamental differences between individuals enrolled in the program and those suitable for

the program but who are not enrolled. This could be explained by higher motivation levels, fear after a recent acute exacerbation, or any number of socioeconomic factors beyond the researcher's control.

The best available research design to control for this bias is the randomized controlled trial (RCT). Randomization reduces the threat of selection bias by giving each member of the population an equal opportunity to be chosen for inclusion in the study. Therefore, any unknown patient characteristics are distributed equally between groups of participants and nonparticipants.

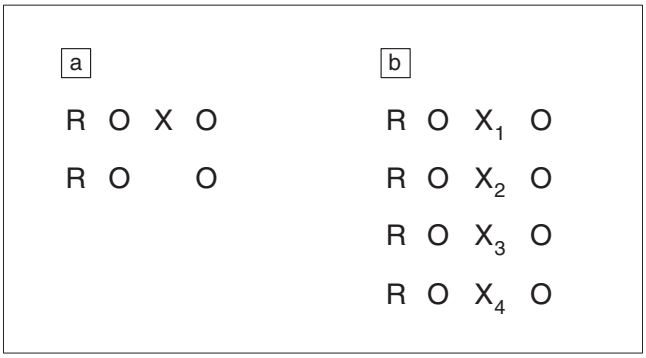
The addition of a control group eliminates another major threat to validity—*regression to the mean* (RTM). Regression to the mean suggests that without the effect of the intervention members with high costs and utilization in one period will tend to cost less and use fewer services in the following period (a move toward the mean). Conversely, members using few services in one period will use more services and accrue higher costs in the subsequent period. With the addition of a control group, it is possible to observe whether changes in the outcome measure occurred differentially between the two groups.

As a practical matter, the RCT is costly and very difficult to implement correctly and, therefore, is used very infrequently outside of an academic center. It requires tremendous collaboration from participants and project administrators over the course of at least 6 months, but more typically 1 year. A number of participants will inevitably fall out of the study for one reason or another (called "lost to attrition"), which may reduce the power to detect meaningful differences in outcomes between the groups. Depending on the type of study or where it is conducted, individuals from the various cohorts may find out which group they were assigned to and behave differently than expected (contamination bias). While there are many issues that must be controlled for, the RCT nonetheless remains the most powerful of research designs.

Figure 1 illustrates (a) the basic RCT with one treatment and one control group, and (b) a multi-intervention design where four different interventions are evaluated. There are many variations of the RCT including *crossover* and *nested* designs. For a more comprehensive discussion of the RCT, interested readers should refer to Shadish, Cook, and Campbell (2002).

An appropriate design to test the effectiveness of a face-to-face HH intervention versus traditional DM would have three study groups (a) a DM treatment group receiving telephonic interventions alone, (b) a DM treatment group receiving HH intervention alone, and (c) an control group receiving neither intervention. This design could be further enhanced by dividing the HH group into two, one receiving "hands-on care" without any DM, and the second receiving hand-on care and DM.

FIGURE 1



Depiction of the Randomized Controlled Trial (RCT). Individuals are randomly (R) assigned to treatment or control group. There are two observation points (O): (a) shows a design with one treatment and one control group, where the treatment group receives the intervention (X) and (b) shows a design in which there are four treatment groups, each receiving a different intervention.

In light of the fact that DM and HH services overlap in only certain types of activities (e.g., assessment, education, etc.), the choice of appropriate outcome measures is paramount. HH agencies participating in Medicare or Medicaid are already required to collect and report specific patient outcome data as part of the Outcomes Assessment and Information Set (OASIS), which was instituted in 1999 as a component of the Balanced Budget Act of 1997. These data items encompass sociodemographic, environmental, support system, health status, and functional status attributes of adult patients (Centers for Medicare and Medicaid Services, 2005). It is, therefore, possible to use these existing measures as part of the overall analysis. However, if the evaluation is meant to compare costs or cost savings between the various study arms, additional outcome measures may include hospitalizations or readmissions, emergency department visits, and other acute events, which are cost drivers and considered to be affected by DM and HH interventions.

In general there are relatively few inherent weaknesses in the RCT design. However, it is complex and resource intensive, and because it must be tightly con-

FIGURE 2

O_A C X O_2

O_A C O_2

Depiction of the Regresssion Discontinuity (RD) Design. O_A indicates the pretest measure, C is cutoff score on that pretest measure that determines assignment to treatment (X) or control. O_2 is the posttest measure, which may or may not be the same as the pretest measure.

trolled over a long observation period, there are many opportunities for bias to be introduced. In addition to the issue of attrition, *measurement error* can also be a concern if data must pass through many individuals on the way to the final evaluation, or if measurement systems change during the observation period. Another weakness of the RCT is a lack of generalizability to other persons, settings, treatments or outcomes (Linden, Adams, & Roberts, 2004b). With typically narrowly defined participation criteria or treatment parameters, results may not be valid or useful in disparate circumstances.

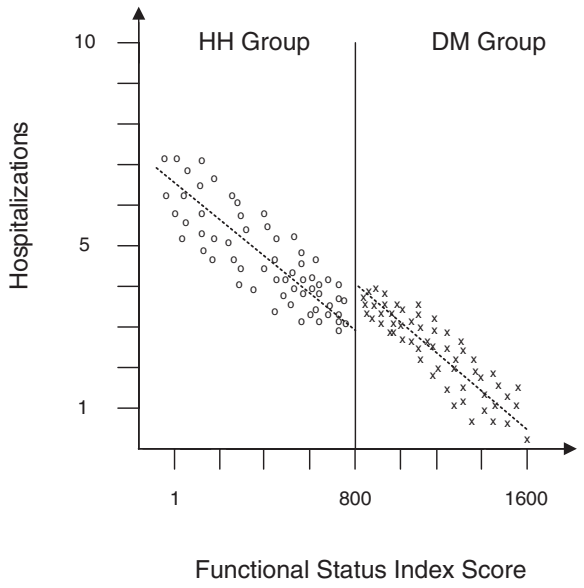
THE REGRESSION DISCONTINUITY DESIGN

As described above, there are logistic, practical, and even ethical limitations that preclude the widespread use of the RCT in evaluating most large-scale programs. The regression discontinuity (RD) design may be the most suitable alternative to randomized studies in evaluating HH as a DM intervention (Linden, Adams, & Roberts, in press-a **UPDATE** ; Trochim, 1984, 1990).

In contrast to the RCT in which individuals are randomly assigned to treatment or control group, the RD design utilizes a "cutoff" score on a preprogram measure or test to determine who will be assigned to the intervention or control group. The defining characteristic of the RD design is in identifying whether a difference is found in the relationship between the assignment variable and outcome occurring exactly at that cutoff score, where individuals in the treatment and nontreatment groups are most similar.

While the pretest measure does not have to be identical to the outcome measure, it does have to be on either a continuous or ordinal scale. Figure 2 depicts the basic

FIGURE 3



Graphic Display of the Regression-Discontinuity (RD) Design. In this example, individuals are assigned to home-health (HH) or disease management (DM) based on their pretest functional status score relative to the cutoff (vertical line). Regression lines are fitted to each group's data (dotted lines): The graph shows a classic discontinuity at the cutoff, indicating a treatment effect in the HH group.

RD design (Shadish et al. 2002). As shown, individuals are assigned to either treatment or control group based on where their pretest value lies in relation to the cutoff score. At the end of the study period, a posttest measurement will be taken.

As the pretest variable must be on either a continuous or ordinal scale, the selection of instruments available to measure HH effectiveness is somewhat limited. One excellent option is a functional status index created by Keepnews, Capitman, and Rosati (2004), which comprised 16 measures collected in the OASIS system and rescored on a scale from 0 to 100 (thus, the maximum score possible is 1,600 indicating complete independence). For the purpose of this example, the cutoff score will be set at 800, with all individuals below that score receiving a HH face-to-face intervention, and all individuals above that score receiving the DM telephonic intervention. The posttest measure will be the number of hospitalizations in the program year.

Figure 3 illustrates the results of this hypothetical evaluation. All individual pretest and posttest values are charted on the scatterplot as single X-Y coordinates, and a separate regression line is then drawn

through each group's data. As shown, the two lines do not overlap at the cutoff score (i.e., discontinuity), and the HH line intersects with the cutoff at a lower level than the DM line. In other words, the HH intervention appears to have reduced hospitalizations on average, by one hospitalization per patient per year compared to the DM program. Although visual inspection of the data provides a general idea as to the structural form of the RD model, statistical modeling still must be performed to verify whether the discontinuity is indeed statistically significant.

The importance of strict adherence to the cutoff when assigning patients to treatment cannot be emphasized enough. A fundamental criterion necessary for obtaining an unbiased estimate of a treatment effect is to have an assignment process that is completely known and perfectly measured (Shadish et al. 2002). The underlying premise of the RD design is that participants located immediately adjacent to the cutoff are the most similar and, therefore, provide the best comparison units for assessing treatment effect.

If the cutoff is strictly adhered to, the RD design controls for most threats to validity simply because any given bias would have to affect the treatment group causing a discontinuity that coincides with the cutoff. While it is theoretically possible, the likelihood of such an occurrence is remote. Similar to the RCT, the RD design is limited in its generalizability. As most programs are implemented in a particular setting with a given set of individuals under a specific set of circumstances, the results may be limited in their ability to infer a causal relationship across persons, settings, treatments, and outcomes. Therefore, to increase the generalizability potential of the study outcomes, the program intervention must be as inclusive as possible across as many domains as feasible (Linden et al., 2004b).

MATCHED PAIRS USING THE PROPENSITY SCORE

In contrast to academic institutions, most industry-based organizations do not have the resources or structure to perform an RCT as a means of assessing program effectiveness. As such, researchers mostly rely on observational study designs to evaluate large-scale programs.

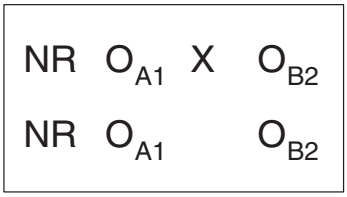
The primary differentiating characteristic between an observational design and an experimental design (e.g., RCT) is the assignment process for study participation. While in the RCT and RD designs the assignment process is completely known and perfectly measured, individuals in observational studies self-select to participate. For example in DM, patients deemed suitable for the program are invited to participate. Whether they choose to enroll may be determined by a myriad of factors, of which some are measurable, and most are not. Therefore, the primary threat to validity of any observational study is that of selection bias. The best that a researcher can do under these circumstances is estimate the magnitude of the selection bias via a sensitivity analysis (Linden et al., in press-b **UPDATE**).

One method to reduce the threat of selection bias is by creating matched pairs of participants and controls (also referred to as a case-control study) who have similar observable characteristics. The theory goes that the more observed characteristics that are captured in the matching process, the fewer unobservable factors remain to bias the study findings. The propensity score, defined as the probability of assignment to the treatment group, conditional on covariates (Dehejia & Wahba, 2002), can control for preintervention differences between the enrolled and nonenrolled groups. The underlying assumption for using the propensity score in DM is that enrollment in the program is associated with observable preprogram variables such as age, sex, utilization, cost, etc. (Linden et al., 2005b). Propensity scores are derived from a logistic regression equation that reduces each participant's set of covariates into a single score, making it feasible to match on what are essentially multiple variables simultaneously.

Figure 4 illustrates the schematic for a propensity score matched case-control study (Shadish et al. 2002). As shown, the propensity score serves as a proxy pretest variable, thereby relinquishing the need for extensive historic data or an identical pretest and posttest measure. To ensure that the propensity-matching technique was successful in creating equivalent groups, baseline characteristics of the two groups should be compared. No observable differences should be noted between groups on any preprogram variable. Outcomes are measured in a manner similar to any other pre–post study design with a control group.

Until recently the propensity score-matching technique was used in only single-treatment studies. Following this methodology, a face-to-face HH intervention could only be compared to one alternative intervention (e.g., telephonic DM) or a control group.

FIGURE 4



Schematic of the Propensity Score Matched Design. As this is an observational study, participants are not randomized (NR) to treatment (X) or control. O_{A1} indicates the proxy pretest (propensity score) and O_{B2} is the posttest measure.

However, recent innovations to the design has allowed for multiple group comparisons using ordinal or categorical regression models (Imai & Van Dyk, 2004; Imbens, 2000; Joffe & Rosenbaum, 1999). Using these methods, a comprehensive multigroup analysis can conducted as described above in the RCT section.

The most significant threat to validity with the propensity score-matching technique (as with any observational study) is that of selection bias. This is because of propensity scores being based solely on observable confounding variables and not for unknown or "hidden" sources of variation. A sensitivity analysis should be performed to estimate the odds of participants assigned to the program intervention having this hidden bias. This analysis allows the researcher to determine if the results are sensitive enough to bias to alter conclusions about a causal relationship between the intervention and the outcome.

The size of the nonparticipant pool available for matching can be the second limitation of the propensity-scoring method (Linden et al., 2005b). There are many situations that may arise in HH where the program group will encompass all patients requiring HH services. In situations such as these, matching current participants with historic controls may be the best alternative; however, adjustments must be made to account for inflation or other secular trends. If there are significantly fewer concurrent controls than cases available for matching, the researcher may choose to use the "matching with replacement" technique in which a given control may be matched to more than one case. This introduces additional statistical issues but can be easily managed via a nonparametric analytic method called "bootstrapping" (Linden et al., 2005a).

As participants are matched based on specific observed baseline characteristics, this design generally lacks generalizability. However, this issue can be somewhat mitigated by creating matches across various persons, settings, or treatments (Linden et al., 2004b).

CONCLUSIONS

Given the desire to expand into the area of DM, and the current body of literature that appears to support telephonic interventions as a viable alternative to face-to-face patient contact, HH agencies will need to identify and demonstrate areas in which they have a clinical and competitive advantage over traditional DM models. This article described in some detail three research designs that may assist the HH industry in evaluating their effectiveness in delivering DM services. Each model has its strengths and weaknesses, and the choice of which evaluation technique to use will depend ultimately on factors such as available resources, data, and expertise. Other designs that are suitable and should be considered for evaluating these specific issues are time series analysis (Linden et al., 2003), survival analysis (Linden et al., 2004a), and instrumental variables (Linden & Adams, in press **UPDATE**).

REFERENCES

Anie, K. A., Jones, P. W., Hilton, S. R., & Anderson, H. R. (1996). A computer-assisted telephone interview technique for assessment of asthma morbidity and drug use in adult asthma. *Journal of Clinical Epidemiology*, *49*, 653-656.

Badger, T., Segrin, C., Meek, P., Lopez, A. M., Bonham, E., & Sieger, A. (2005). Telephone interpersonal counseling with women with breast cancer: Symptom management and quality of life. *Oncology Nursing Forum*, *32*, 273-279.

Burke, W. J., Roccaforte, W. H., Wengel, S. P., Conley, D. M., & Potter, J. F. (1995). The reliability and validity of the Geriatric Depression Rating Scale administered by telephone. *Journal of the American Geriatric Society*, *43*, 674-679.

Centers for Medicare and Medicaid Services. (2005). *OASIS overview*. Retrieved June 26, 2005, from www.cms.hhs.gov/oasis/hhoview.asp

Chwalow, A. J., Costagliola, D., Stern, J., Mesbah, M., & Eschwege, E. (1989). Telephone versus face to face interviewing as a means of collecting data relevant to the management of diabetes among general practitioners in France: A randomized design. *Diabete et Metabolisme*, *15*, 157-160.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for non-experimental causal studies. *Review of Economics and Statistics*, *84*, 151-161.

Donovan, R. J., Holman, C. D., Corti, B., & Jalleh, G. (1997). Face-to-face household interviews versus telephone interviews for health surveys. *Australia and New Zealand Journal of Public Health*, *21*, 134-140.

Friedman, R. H., Kazis, L. E., Jette, A., Smith, M. B., Stollerman, J., Torgerson, J., et al. (1996). A telecommunications system for monitoring and counseling patients with hypertension: Impact on medication adherence and blood pressure control. *American Journal of Hypertension*, *9*(4 Pt 1), 285-292.

Galan, I., Rodriguez-Artalejo, F., & Zorrilla, B. (2004). Telephone versus face-to-face household interviews in the assessment of health behaviors and preventive practices. *Gaceta Sanitaria*, *18*, 440-450.

Galasso, R., Panico, S., Celentano, E., & Del Pezzo, M. (1994). Relative validity of multiple telephone versus face-to-face 24-hour dietary recalls. *Annals of Epidemiology*, *4*, 332-336.

Greenfield, T. K., Midanik, L. T., & Rogers, J. D. (2000). Effects of telephone versus face-to-face interview modes on reports of alcohol consumption. *Addiction*, *95*, 277-284.

Hunter, E. F. (2000). Telephone support for persons with chronic mental illness. *Home Healthcare Nurse*, *18*, 172-179.

Imai, K., & Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, *99*, 854-866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, *87*, 706-710.

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: Propensity scores. *American Journal of Epidemiology*, *150*, 327-333.

Keepnews, D., Capitman, J. A., & Rosati, R. J. (2004). Measuring patient-level clinical outcomes of home health care. *Journal of Nursing Scholarship*, *36*, 79-85.

Korner-Bitensky, N., Wood-Dauphinee, S., Siemiatycki, J., Shapiro, S., & Becker, R. (1994). Health-related information postdischarge: Telephone versus face-to-face interviewing. *Archives of Physical Medicine and Rehabilitation*, *75*, 1287-1296.

Linden, A., & Adams, J. (in press) **UPDATE?**. Evaluating disease management program effectiveness: An introduction to instrumental variables. *Journal of Evaluation in Clinical Practice*, *11* **PAGES**.

Linden, A., Adams, J., & Roberts, N. (2003). Evaluating disease management program effectiveness: An introduction to time series analysis. *Disease Management*, *6*, 243-255.

Linden, A., Adams, J., & Roberts, N. (2004a). Evaluating disease management program effectiveness: An introduction to survival analysis. *Disease Management*, *7*, 180-190.

Linden, A., Adams, J., & Roberts, N. (2004b). The generalizability of disease management program results: Getting from here to there. *Managed Care Interface*, *17*, 38-45.

Linden, A., Adams, J., & Roberts, N. (2005a). Evaluating disease management program effectiveness: An introduction to the bootstrap technique. *Disease Management and Health Outcomes*, *13*, 159-167.

Linden, A., Adams, J., & Roberts, N. (2005b). Using propensity scores to construct comparable control groups for disease management program evaluation. *Disease Management and Health Outcomes*, *13*, 107-127.

Linden, A., Adams, J., & Roberts, N. (in press-a) **UPDATE?**. Evaluating disease management program effectiveness: An introduction to the regression-discontinuity design. *Journal of Evaluation in Clinical Practice*, *11* **PAGES**.

Linden, A., Adams, J., & Roberts, N. (in press-b) **UPDATE**. Strengthening the case for disease management effectiveness: Unhiding the hidden bias. *Journal of Evaluation in Clinical Practice*, *11* **PAGES**.

Lyu, L. C., Hankin, J. H., Liu, L. Q., Wilkens, L. R., Lee, J. H., Goodman, M. T., et al. (1998). Telephone vs face-to-face interviews for quantitative food frequency assessment. *Journal of the American Dietetic Association*, *98*, 44-48.

Midanik, L. T., Greenfield, T. K., & Rogers, J. D. (2001). Reports of alcohol-related harm: Telephone versus face-to-face interviews. *Journal of Studies on Alcohol*, *62*, 74-78.

Mohr, D. C., Likosky, W., Bertagnolli, A., Goodkin, D. E., Van Der Wende, J., Dwyer, P., et al. (2000). Telephone-administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. *Journal of Consulting and Clinical Psychology*, *68*, 356-361.

Piette, J. D., McPhee, S. J., Weinberger, M., Mah, C. A., Kraemer, F. B. (1999). Use of automated telephone disease management calls in an ethnically diverse sample of low-income patients with diabetes. *Diabetes Care*, *22*, 1302-1309.

Piette, J. D., Weinberger, M., Kraemer, F. B., & McPhee, S. J. (2001). Impact of automated calls with nurse follow-up on diabetes treatment outcomes in a Department of Veterans Affairs health care system: A randomized controlled trial. *Diabetes Care*, *24*, 202-208.

Piette, J. D., Weinberger, M., & McPhee, S. J. (2000). The effect of automated calls with telephone nurse follow-up on patient-centered outcomes of diabetes care: A randomized, controlled trial. *Medical Care*, *38*, 218-230.

Pridemore, W. A., Damphousse, K. R., & Moore, R. K. (2005). Obtaining sensitive information from a wary population: A comparison of telephone and face-to-face surveys of welfare recipients in the United States. *Social Science and Medicine*, *61*, 976-984.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change [Invited paper]. *American Journal of Health Promotion*, *12*(1), 38-48.

Riegel, B., Carlson, B., Kopp, Z., LePetri, B., Glaser, D., & Unger, A. (2002). Effect of a standardized nurse case-management telephone intervention on resource use in patients with chronic heart failure. *Archives of Internal Medicine*, *162*, 705-712.

Shadish, S. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sparks, K. E., Shaw, D. K., Eddy, D., Hanigosky, P., & Vantrese, J. (1993). Alternatives for cardiac rehabilitation patients unable to return to a hospital-based program. *Heart and Lung*, *22*, 298-303.

Trochim, W. (1984). *Research design for program evaluation: The regression-discontinuity approach.* Beverly Hills, CA: Sage.

Trochim, W. M. K. (1990). The regression discontinuity design. In L. B. Sechrest, E. Perrin, & J. Bunker (Ed.), *Research methodology: Strengthening causal interpretations of nonexperimental data* (AHCPR Conference Proceedings, PHS 90-3545, pp 119-139). Rockville, MD: Agency for Health Care Policy and Research.

van Wijck, E. E., Bosch, J. L., & Hunink, M. G. (1998). Time-tradeoff values and standard-gamble utilities assessed during telephone interviews versus face-to-face interviews. *Medical Decision Making*, *18*, 400-405.

Ware, J. J., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, *30*, 473-483

Weinberger, M., Nagle, B., Hanlon, J. T., Samsa, G. P., Schmader, K., Landsman, P. B., et al. (1994). Assessing health-related quality of life in elderly outpatients: Telephone versus face-to-face administration. *Journal of the American Geriatric Society*, *42*, 1295-1299.

Ariel Linden, DrPH, MS *is a health services researcher whose focus is on evaluating disease management program effectiveness. He is president of the Linden Consulting Group and clinical associate professor in the School of Medicine at the Oregon Health and Science University.*