Statistical Analysis Plan

a. *Primary aim analyses*. The primary hypothesis that at 12 months the intervention will lead to lower (improved) BMI and SCL-20 scores vs. usual care will be tested with a repeated-measures mixed model.[1-3]

$$\mathbf{Y_t = \beta_0 + \beta_1 X + \beta_2 Y_0 + \beta_3 T + \beta_4 (XT) + \Sigma \beta_{4+i} Z_i + \alpha + \gamma + \varepsilon} \quad \mathbf{(1)}$$

Let Y_t be the outcome of interest at follow-up time T on a patient randomized to arm X (intervention or usual care). Given the covariate-adaptive randomization, distributions of baseline values on the outcome variable (Y_0) and key characteristics (Z_i) should be similar between study arms and thus not bias the results. But to the extent they are associated with the outcome, their inclusion in the analysis will account for otherwise unexplained variation and hence increase efficiency.[4] Additional covariates (e.g., ADM changes, out-of-study obesity treatment) will be included in secondary analyses to elucidate the primary ITT findings. α and γ are clinic and PCP random effects. The random error, ε, accounts for the non-independence of repeated measures using a covariance structure within participants to be determined by the least Bayesian information criterion. Similarly, between-group mean differences in BMI and SCL-20 score at 24 months (**Hypothesis 2**), and in secondary outcomes, will be examined using tests of group-by-time interactions in repeated-measures mixed-effects linear (for continuous variables) or logistic models (for categorical variables).

Primary analyses will follow ITT principles. We will verify that mixed model-based results are not sensitive to violations of model assumptions with permutation and bootstrap resampling tests.[5,6] We will document the extent, pattern, and reasons for missing data, and will conduct sensitivity analyses of the impact of missing data on stability of the primary results. For example, we may use weight data up to the point when they are no longer available (e.g., dropouts) or should not be used (e.g., pregnancy) and then employ multiple imputation[195, 196] based on a predictive distribution for future weights.

We will perform cost-effectiveness analyses (**Hypothesis 3**) by extending and combining existing models for obesity and depression that we and others have developed.[7-12] We will compare incremental costs, estimated from the perspectives of health systems (direct medical costs only) and society (direct medical and non-medical costs), to incremental benefits, expressed as QALYs gained. We will also consider the number needed to treat (NNT) as an important and clinically-relevant outcome and estimate the intervention cost per NNT. The use of QALYs allows for comparisons of both changes in morbidity and health status as well as mortality effects based on change in the risk of death and, among survivors, reduction in quality of life due to nonfatal events, given the probabilities of disease progression in the target population.[10,13] We will convert Cohen's *d* effect sizes to estimates of the NNT to have one more patient with better outcomes in the intervention arm vs. usual care.[14] We will exclude from all analyses research-related costs, such as costs of recruitment, screening, and outcome surveillance that are beyond those recommended for routine clinical practice. Intervention start-up costs, fixed costs of sustaining the intervention, and marginal costs of adding additional participants to the intervention will be

differentiated because they are relevant for different decisions: whether to implement the intervention in the first place and whether to sustain or expand it. We will use simulation models, similar to those in DPP and STAR*D,[10,15,16] to analyze incremental cost-effectiveness ratios during the trial and projected into 5-, 10-, 20-, 30-year, and lifetime horizons. Cost-utility estimates with different time horizons will be useful for stakeholders deciding on program implementation. Sensitivity analyses will be performed and results will be interpreted according to standard guidelines.[17-19]

b. *Secondary aim analyses*. We will analyze quantitative process data using standard tests, e.g., Student's *t*-tests and χ^2 tests for continuous and categorical variables, respectively. We will analyze the qualitative data using NVivo.[20] We will organize the qualitative data by source (e.g., participants, coaches, facilitators) then transcribe, and code the data based on the RE-AIM and PARiHS domains assessed. We will use a codebook of codes and definitions to train coders and guide data coding. To identify themes within and across groups (e.g., participants, coaches, facilitators) we will use content analysis methods.[112, 114, 115] We will triangulate data from these different sources to increase the validity of the qualitative data and to draw conclusions about reach, adoption, implementation, and maintenance of the intervention.[21]

We will conduct subgroup analyses to evaluate potential effect modifiers for the primary outcomes by expanding equation 1 (previous page) to include appropriate modifier-by-group interaction terms. In this context, testing whether the β coefficients of the interaction terms are equal to zero is equivalent to testing the null hypothesis that the variable of interest does not independently modify the intervention effect.

Longitudinal (e.g., change in mediator, such as depression, from baseline to 6 months and change in outcome, such as weight loss, from 6 to 12 months) and contemporaneous (e.g., changes in mediator and outcome from baseline to 12 months) mediation will be examined separately by MacKinnon's product of coefficients test (αβ).[22] Asymmetric confidence limits will be constructed based on the distribution of the product with the PRODCLIN program.[23] Because multicollinearity may be present in multiple mediator models, we first will test each mediator in single-mediator models. Multiple-mediator models including all variables that are at least marginally significant in the single-mediator models will test for independent and suppression effects. To determine the extent of mediated effect, the percentage of total effect mediated will be calculated for each significant mediator as αβ/(αβ+ γ), where γ is the direct intervention effect on outcome. The effect modification-mediation analyses are hypothesis-generating, but we pre-specify the variables to ensure a focus.

Statistical power. We power this trial on the co-primary endpoints—BMI and SCL-20 score at 12 months. A sample of 202/arm has 90% power to detect a standardized 0.35 mean difference (Cohen's *d*) between the intervention and usual care groups at α=5% (2-sided), assuming at least 85% retention at 12 months based on our prior trial experiences.[24-28] We used a *t*-test with simplified assumptions to estimate power, whereas actual power may be greater due to increased efficiency associated with repeated-measures mixed models with baseline and covariate adjustments.[29] Because treatment success will be judged on both (not either) primary outcomes, multiplicity adjustment is unnecessary.[30] Also, no multiplicity adjustment will be made for secondary analyses, which are

intended to complement the primary findings and to inform future research. They will be interpreted within that context, considering the totality of evidence available.[30,31]

We chose a *d* of 0.35 as the minimum important between-group difference based on our prior studies and other available literature. The *d* effect size for reducing BMI by the E-LITE self-directed intervention vs. usual care was 0.46 (corresponding to a mean of 5.0% vs. 2.6% weight loss over 15 months).[32] Weight loss of $\geq$5% is widely regarded clinically significant,[33] while a weight change of <3% defines weight maintenance.[34] Hence, the net BMI lowering effect for the E-LITE self-directed intervention relates to the minimal clinically important difference in weight reduction. A meta-analysis of behavioral weight-loss studies reported effect sizes of 0.61–0.67 for improvements in depression and self-esteem.[35] The *d* effect sizes in the PEARLS trials ranged from 0.35–0.74.[25,26] Further, the National Institute for Clinical Excellence in the U.K. has defined a threshold of clinical significance for depression treatment as a standardized effect size of 0.50.[36]

References

1. Littell RC, Milliken GA. *SAS System for Mixed Models.* Cary, NC: SAS Institute Inc; 1996.
2. Brown H, Prescott R. *Applied Mixed Models in Medicine.* Chichester, England: John Wiley & Sons, Inc.; 1999.
3. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York: Springer-Verlag; 2000.
4. Shao J, Yu X, Zhong B. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika.* 2010;97(2):347-360.
5. Tang L, Duan N, Klap R, Asarnow JR, Belin TR. Applying permutation tests with adjustment for covariates and attrition weights to randomized trials of health-services interventions. *Stat Med.* 2009;28(1):65-74. PMID18937226.
6. DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Statistic Sci.* 1996;11(3):189-228.
7. Goldhaber-Fiebert JD, Rubinfeld RE, Bhattacharya J, Robinson TN, Wise PH. The utility of childhood and adolescent obesity assessment in relation to adult health. *Med Decis Making.* 2012(Epub ahead of print).
8. Hall KD, Jordan PN. Modeling weight-loss maintenance to help prevent body weight regain. *Am J Clin Nutr.* 2008;88(6):1495-1503. PMID19064508.
9. Hall KD. Predicting metabolic adaptation, body weight change, and energy intake in humans. *American journal of physiology Endocrinology and metabolism.* 2010;298(3):E449-466. PMID19934407.
10. Herman WH, Hoerger TJ, Brandle M, et al. The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. *Ann Intern Med.* 2005;142(5):323-332. PMID15738451.
11. Haji Ali Afzali H, Karnon J, Gray J. A critical review of model-based economic studies of depression: modelling techniques, model structure and data sources. *Pharmacoeconomics.* 2012;30(6):461-482. PMID22462694.
12. Wein LM, Yang Y, Goldhaber-Fiebert JD. Assessing screening policies for childhood obesity. *Obesity (Silver Spring).* 2012;20(7):1437-1443. PMID22240724.
13. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine.* New York: Oxford University Press; 1996.
14. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry.* 2006;59(11):990-996. PMID16368078.
15. The STAR*D Investigative Team. Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Protocol and Analytic Plan. In:2002: http://www.madinamerica.com/MADINAMERICA.COM/STARD%20Documents_files/STAR-D%20Protocol%20with%20Analytic%20Plan.pdf. Accessed April 19, 2011.
16. Within-trial cost-effectiveness of lifestyle intervention or metformin for the primary prevention of type 2 diabetes. *Diabetes Care.* 2003;26(9):2518-2523. PMID12941712.
17. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines

for using clinical and economic evaluations. *CMAJ.* 1992;146(4):473-481. PMID1306034.
18. *Guidelines for the economic evaluation of health technologies [3rd Edition]*. Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health;2006.
19. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA.* 1996;276(15):1253-1258. PMID8849754.
20. *QSR NVivo* [computer program]. Victoria, Australia: QSR International.
21. Patton MQ. *Qualitative research and evaluation methods.* San Francisco: Sage Publications; 2002.
22. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods.* 2002;7(1):83-104. PMID11928892.
23. MacKinnon DP, Fritz MS, Williams J, Lockwood CM. Distribution of the product confidence limits for the indirect effect: program PRODCLIN. *Behav Res Methods.* 2007;39(3):384-389. PMID17958149.
24. Ma J, Yank V, Xiao L, et al. Translating the Diabetes Prevention Program lifestyle intervention for weight loss into primary care: A randomized trial. *Arch Intern Med.* 2012 (accepted for publication).
25. Ciechanowski P, Wagner E, Schmaling K, et al. Community-integrated home-based depression treatment in older adults: a randomized controlled trial. *JAMA.* 2004;291(13):1569-1577. PMID15069044.
26. Ciechanowski P, Chaytor N, Miller J, et al. PEARLS depression treatment for individuals with epilepsy: a randomized controlled trial. *Epilepsy & behavior : E&B.* 2010;19(3):225-231. PMID20609631.
27. Ma J, Berra K, Haskell WL, et al. Case management to reduce risk of cardiovascular disease in a county health care system. *Arch Intern Med.* 2009;169(21):1988-1995. PMID19933961.
28. Katon WJ, Lin EH, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med.* 2010;363(27):2611-2620. PMID21190455.
29. Kraemer HC. To increase power in randomized clinical trials without increasing sample size. *Psychopharmacol Bull.* 1991;27(3):217-224. PMID1775591.
30. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Stat Med.* 1999;18(15):1905-1942. PMID10532877.
31. Bender R, Lange S. Adjusting for multiple testing--when and how? *J Clin Epidemiol.* 2001;54(4):343-349. PMID11297884.
32. Ma J, Yank V, Xiao L, et al. Translating the Diabetes Prevention Program lifestyle intervention for weight loss into primary care: a randomized trial. *JAMA Intern Med.* 2013;173(2):113-121. PMID23229846.
33. Thomas PR. *Weighing the Options: Criteria for Evaluating Weight-Management Programs.* Washington, DC: Committee to Develop Criteria for Evaluating the Outcomes of Approaches to Prevent and Treat Obesity, Institute of Medicine;1995.

34. Stevens J, Truesdale KP, McClain JE, Cai J. The definition of weight maintenance. *Int J Obes (Lond).* 2006;30(3):391-399. PMID16302013.
35. Blaine BE, Rodman J, Newman JM. Weight loss treatment and psychological well-being: a review and meta-analysis. *J Health Psychol.* 2007;12(1):66-82. PMID17158841.
36. National Institute for Clinical Excellence. *Depression: Management of Depression in Primary and Secondary Care.* London, England: National Health Service 2004.