Appendix 2: Further Explanation of Analyses

Scores were constructed at the patient and facility level for each of the 2 methods of explicit quality measurement using two different methods. In the first method, which produced "raw scores," we determined the criteria that made participants eligible for the process specified in each indicator (yes/no). We then determined whether participants had received the specified process each time an indication was noted in their medical record (yes/no/proportion). We determined aggregate indicator scores for each indicator category (i.e., acute, chronic, and preventive care; screening, diagnosis, treatment, and follow-up) by dividing all instances in which participants received recommended care by the total number of possible instances in which the care could have been received. We presented the scores as proportions ranging from 0-100%. This is a close approximation of the method used in the VA's operational use of the focused review instrument and in the original publications of the RAND global tool.

In the second score construction method, which was also applied to the implicit scores, we used an analytic method that took in to account the major sources of measurement error inherent in each instrument to produce scores "filtered" or adjusted for measurement error. Specifically, to account for these sources of measurement error, we used multilevel ordinal and logistic regression models to construct the scores at the patient and site level. For the implicit review the major source of measurement error is between independent reviews of a patient record, or the interrater reliability. Thus we have multiple reviews within patient at the lowest level in the multilevel model to estimate and remove this source of measurement error. For, the explicit reviews, the major source of measurement error is the variation in the number of eligible indicators and the noise inherent in using a group of indicators to measure the quality of care delivered to a particular patient, or the inter-item reliability. The commonly used scoring that calculates the number of indicators passed over number eligible does not reflect this source of measurement error. For physician implicit review, the measurement error of abstracting information from the chart can not be separated from the measurement error in the rating of quality. Thus we have multiple indicators within each patient at the lowest level in the model to estimate and remove this source of measurement error. These models produce estimates of the scores that are narrower in range and often with different rank order than the raw scores, as outlier scores that are estimated imprecisely are shrunk back toward the overall mean score for the entire sample. These "filtered scores" will represent better (in terms of mean square error) the true underlying score that would be obtained if an infinite number of physician raters reviewed the quality of care in a patient record implicitly or an infinite number of process criteria were used to measure the explicit quality of care of an individual. (Snijders and Bosker 1999; Bryk and Raudenbush 1992; Goldstein 1995; Zillich et al. 2005)

For the focused and global explicit measures, because the inter-rater reliability is consistently high, the major source of measurement error is the variable and sometime small number of indicators that go into the construction of the scores for any individual. It has not generally been appreciated that the widely varying number of process indicators that apply to any particular patient introduces substantial amounts of variation into the explicit measurement process. At the site level the measurement error is a function of the number and heterogeneity of the individual scores within each site. Therefore, to produce "filtered scores" for the explicit measures, the dichotomous process measures were clustered within patients who were then clustered within site. In effect, for the explicit review measures this latter procedure weights the process score (# processes performed over # of eligible process measures in that patient) by the number of process measures that went into each persons score. Thus, a score of 1/2 is treated as much less precisely estimated than a score of 7/14.

For implicit review at the patient level, the major source of measurement error is the interrater variability in rating the quality of care. We have published the results of reliability studies done as part of this study, in which 70 cases, each with 1 or more conditions, were reviewed by three different reviewers. (Hofer et al. 2004) The reliability of a single review was as good or better than in any previously published studies using physician implicit review ranging from an ICC of .2 to .5 depending on the condition reviewed. Our analyses used the multiple reviews gathered on a sample of our reviews in generating the scores, thus adjusting the scores for the imprecision of the review. In these models, multiple ordinal ratings of a record were clustered within patient at the second level, who were in turn clustered within sites of care at the third level.

The resultant explicit and implicit patient and facility level quality scores were then transformed onto a probability scale for purposes of presentation. In order to present a "pass rate" for implicit review comparable to the pass rate calculated for the explicit instruments, we estimated the probability for each patient or facility of receiving an overall rating of good or better care (as opposed to "borderline" or worse).

We did a sensitivity analysis to examine the impact of restricting both the focused and global explicit tool scores to only include indicators supported by evidence based on randomized controlled trials (level 1 evidence). (Malin et al. 2000) (Canadian Task Form on the Periodic Health Examination 1979) We reconstructed the scores for each explicit tool using only this subset of indicators and again compared the correlations between each tool, and the facility-level variance explained by the explicit tools with the limited set of indicators. Doing so substantially reduced both the number of possible indicators and the median number of applicable indicators per patient. Most of the correlations were modestly lower (between 10 - 15%) when the scores were restricted to indicators with high levels of evidence supporting their use.

## Appendix Reference List


1. Asch, S. M., McGlynn, E. A., Hogan, M. M., Hayward, R. A., Shekelle, P., Rubenstein, L., Keesey, J., Adams, J., and Kerr, E. A. Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. Annals of Internal Medicine 2004;141: 12: 938-45.
2. Bryk AS and Raudenbush SW. Hierarchical Linear Models Applications and Data Analysis Methods. Sage, Newbury Park,1992.
3. Canadian Task Form on the Periodic Health Examination. The periodi health examination. Canadian Medical Association Journal. 1979; 121:1193-1254.
4. Goldstein, H. Multilevel Statistical Models. Halstead Press, New York, 2nd Edition, 1995
5. Hofer, T. P., Asch, S. M., Hayward, R. A., Rubenstein, L. V., Hogan, M. M., Adams, J., and Kerr, E. A. Profiling quality of care: Is there a role for peer review? BMC Health Services Research 2004;4: 1:9.
6. Kizer KW, Demakis JG, and Feussner JR. Reinventing VA Health Care: Systematizing Quality Improvement and Quality Innovation. Medical Care 2000;38: 6:I7-I16.
7. Malin JL; Asch S; Kerr EA, and McGlynn EA. Evaluating the quality of cancer care: Development of cancer quality indicators for a global quality assessment tool. Cancer.

2000; 88:701-7.

8. McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., and Kerr, E. A. The quality of health care delivered to adults in the United States. New England Journal of Medicine. 2003;348: 26: 2635-45.
9. Mencke NM, Alley LG, and Etchason J. Application of HEDIS Measures Within a Veterans Affairs Medical Center. American Journal of Managed Care 2000;6: 6 :661-668.
10. Perlin, J. B., Kolodner, R. M., and Roswell, R. H. The Veterans Health Administration: quality, value, accountability, and information as transforming strategies for patient-centered care. American Journal of Managed Care 2004;10: 11 Pt 2:828-36.
11. Snijders TAB and Bosker RJ. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelings. Sage, Thousand Oaks, 1999.
12. Zillich, Alan J., Sutherland, Jason M., Kumbera, Patty A., and Carter, Barry L. Hypertension Outcomes Through Blood Pressure Monitoring and Evaluation by Pharmacists (HOME Study). Journal of General Internal Medicine 2005;20: 12:1091-1096.