Comparison of Weighted Performance Measurement and Dichotomous Thresholds for Glycemic Control in the Veterans Health Administration

LEONARD M. POGACH, MD, MBA[1,2]
MANGALA RAJAN, MBA[1]
DAVID C. ARON, MD, MS[3,4]

OBJECTIVE — Quality measures of glycemic control using threshold values do not assess incremental quality improvement. We compared health care system performance using weighted continuous versus dichotomous measures for glycemic control.

RESEARCH DESIGN AND METHODS — We performed retrospective cross-sectional analysis of chart abstraction data on 37,142 diabetic patients from 141 Veterans Health Administration medical centers in 2000–2001.

RESULTS — Subjects per facility ranged from 163 to 740 (mean 263). Mean overall HbA_{1c} (A1C) was 7.58%. A continuous measure for glycemic control was calculated based on percentage of maximal quality-adjusted life-years saved (QALYsS). Overall mean facility performance using the dichotomous measure was 62% <8% A1C (range 48–75%) and 39% <7% A1C (21–57%), in comparison with 45% maximal QALYsS (31–60%). Correlation between QALYsS and A1C thresholds of <8 (0.848) and <7 (0.838) for facility rankings was excellent; correlation between facility level performance using thresholds of <8 and 7% was poor ($r = 0.13$, $P = 0.14$). Comparison of facility rankings between the <7% dichotomous measure and the QALYsS-weighted measure showed that 22% changed their ranking by ≥ 2 deciles with marked changes in top and bottom deciles.

CONCLUSIONS — Facility rankings vary by threshold or continuous methodology. However, because significant numbers of individuals are unable to reach "optimal" target goals (thresholds) even in clinical trials with extensive exclusion criteria, we propose that a continuous measure assessing improvement toward optimal A1C, rather than a pass/fail optimal target, is both a fairer assessment clinical practice and a more accurate reflection of population health improvement.

Performance measurement is an integral part of health care management. Attributes of good performance measures include their relevance, soundness, and feasibility (1). Governmental, private, and public-private coalitions have defined an increasing number of performance measures to evaluate quality of care (2–6). There are imperatives to consider intermediate outcome measures that are more closely linked to morbidity and mortality and adopting thresholds for adherence closer to guideline-recommended optimal levels (7,8). However, threshold measures may reflect neither the true impact of care on population health nor progress in improvement efforts, thus unfairly evaluating clinician performance. Unintended consequences may include shifting of sicker patients by physicians, nonparticipation in public reporting, and decreased consideration for individual patient preferences (9–11). Potential policy impact of measures and their public reporting upon consumer choice and health care reimbursement (pay-for-performance) poses new challenges in measure development.

Diabetes exemplifies the need for, and difficulty in development of, meaningful and fair intermediate outcome measures. Affecting over 18 million Americans, diabetes is a disproportionate cause of chronic kidney disease, amputations, visual loss, cardiovascular disease, and death (12,13). Not surprisingly, the Diabetes Quality Improvement Project was among the first disease-specific performance measurement coalitions (10, 14). However, although these measures have played a role in improving care and permitting comparison of care across systems (15–17), they are currently under reevaluation by the National Diabetes Quality Improvement Alliance with respect to establishing lower HbA_{1c} (A1C) thresholds to reflect "excellent glycemic control" for public accountability in addition to existing measures for "poor glycemic control" (7). This necessitates reevaluation of the very nature of the measures.

Current accountability measures for intermediate outcomes, e.g., A1C, are dichotomous; they are inconsistent with epidemiological principles and data derived from landmark studies that have demonstrated the efficacy of A1C lowering (18,19). For example, whereas relative risk reduction is linear, absolute risk reduction is log-linear, so that greater absolute risk reduction, and therefore greater population benefit, is derived from treatment initiated at higher than lower values (20). Furthermore, the absolute benefit of

• •

lowering A1C varies with age. Finally, dichotomous measures do not capture the incremental progress that clinicians may have made in improving intermediate outcomes in individual patients not reaching "optimal" levels—levels that are difficult to achieve even in clinical trials. Consequently, dichotomous threshold measures alone for public reporting cannot capture the complexities of clinical management of diabetes and thus may violate core population health principles: valuing different populations appropriately and assessing performance fairly. Another approach is needed.

Quality-adjusted life-years (QALYs) are a measure of health utility that integrate morbidity and mortality into equivalents of well years of life on a scale of 0 to 1 (21). One year of perfect health is scored as 1, and years of life at less than perfect health are represented as <1. Because QALYs measure years of life loss in diabetes because of premature death, as well as years of healthy life lost to disabling complications, they have been widely used to assesses the impact of A1C control on lifetime health (22–24). We propose use of a weighted continuous measure based on QALYs saved (QALYsS) to evaluate cross-sectional performance for A1C in a health plan context, i.e., at a population level, not the individual patient level. Such a measure would complement, not replace, A1C measures for individual patients and quality improvement. We now report the feasibility of this approach. We analyzed the effect of giving partial credit based on QALYsS toward achievement of the proposed A1C threshold of 7% by comparison to dichotomous performance measures of <7 and <8% in terms of measure adherence and league table facility rankings.

RESEARCH DESIGN AND METHODS

We obtained Veterans Health Administration Office of Quality and Performance External Peer Review Program chart abstraction data on all veterans identified as having diabetes during the chart review process between 1 October 2000 and 30 September 2001 (16). The sampling frame included patients with 2 years of continuous enrollment in the Veterans Affairs (VA) who had made one or more visits in the previous 12 months. We analyzed a deidentified dataset consisting of 37,142 individuals aged ≥24 years from 141 facilities with a diagnosis of diabetes based on HEDIS (Health Employer Data Information System) criteria: one inpatient hospitalization or two outpatient visits with diabetes-specific (250.xx, 357.2, 362.0, and 366.41) diagnosis codes or an antiglycemic medication (73% sensitivity and 98% specificity compared with patient self-report (25). The dataset included the last recorded A1C level as a continuous variable or noted that no A1C had been performed. Institutional review boards at Cleveland and East Orange VA medical centers approved the study.

Weighted continuous A1C performance measure

We determined QALYsS for A1C reduction from 7.9 to 7.0% for different age-groups using published values from the Centers from Disease Control and Prevention (22). The following age strata were used: 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, and ≥85 years. We calculated QALYsS resulting from A1C reduction for each individual by assigning a value of zero to A1C values >7.9%, and, because the relative and absolute risk reduction for all complications is linear within this range of A1C, we used linear interpolation to assign QALYsS between 7.9% and the value between 7.0 and 7.9% for each individual patient based on the maximal possible QALY reduction for each age-group (26). This approach is more conservative than a log-linear approach, which would provide fewer QALYsS for each increment as it approached the 7% threshold. Values <7% were assigned maximal credit. Consistent with performance measure criteria initially used by the National Committee for Quality Assurance/HEDIS Comprehensive Diabetes Care, subjects were assigned a score of 0 if an A1C test was not obtained. We calculated the percentage of the maximal QALYsS for each subject within each facility by dividing the extrapolated QALY score for that individual by the maximal possible QALYsS value by lowering A1C from 7.9 to 7.0% based on that individual's age.

Dichotomous thresholds

We defined two dichotomous A1C thresholds: <8 and <7%, based on the 2000 American Diabetes Association Clinical Practice Recommendations contemporaneous with the study period. These recommendations noted that whereas <7% was the target, clinicians should take action for an A1C ≥8%. An individual subject met adherence criteria to an individual measure if he or she achieved an A1C value below the dichotomous threshold. In developing a measure of "excellent" control, we assigned an upper boundary of 8%, above which no partial credit would be provided.

Analyses

We determined average percent of subjects achieving the threshold measure within each age-group and overall. Individual patient data were aggregated by facility. We performed bivariate correlations between percent QALYsS achieved and percentage meeting each dichotomous threshold. Since 90th and 10th percentiles constitute a common industry standard for identifying best/worst performing plans, we ranked facilities into deciles by the two methods (27). We determined how rankings moved based on different thresholds and measurement methods.

RESULTS — Table 1 shows population characteristics. Facilities ranged from 163 to 740 subjects with a mean of 263. The population ($n = 37,142$) was largely male (86.1%) and older (61% of patients ≥65 years). Mean overall A1C was 7.58%; facility means ranged from 6.81 to 8.29%.

Individual level percent adherence for <7%, <8%, and percent of maximal QALYsS increased with age, with smaller gains (4–6%) between the 25–34 and 45–64 age strata than for gains (13–17%) between the 45–64 and 65–74 age strata (8–14%) (Table 2). QALYsS inherently reflect lower benefits for good control for the elderly. However, the adherence measure (percent of maximal QALYsS) is defined as a proportion of actual to maximal and hence negates this consideration in its calculation. Overall mean facility performance assessed by the dichotomous measure was 62% <8% A1C (range 48–75%) and 39% <7% A1C (21–57%), compared with 45% for QALYsS (31–60%). As the dichotomous threshold changes from <8 to <7, adherence levels drop by 23 percentage points. However, because QALYs are based on a system of differential weights ranging from 7.9 to 7.0, there is consistency in the adherence levels for each facility for that point in time. There was high correlation between facility rankings using percentage of QALYsS and dichotomous threshold levels of <8 (0.848) and <7 (0.838). Not surprisingly, because many patients will have A1C levels between 7 and 8, rankings based on these two dichotomous measures would differ. However, the correla-

Table 1—*Characteristics of veterans with diabetes in the 141 facilities*

Patient characteristics	
Age (years)	65.87 ± 11.36
Age ≥65 years (%)	61
Men (%)	86.10
Facility characteristics	
Patients (n)	37,142
Facilities (n)	141
Patients per facility	263 ± 96.99
Patients per facility (range)	163–740
Diabetes quality measures	
A1C test performed (%)	93
A1C test (%)	7.58 (6.81–8.29)
A1C done and value <8% (%)	62 (48–75)
Systolic blood pressure test performed (%)	100
Systolic blood pressure (mmHg)	137.24 (131.25–141.84)
Systolic blood pressure done and value <140 mmHg (%)	56 (37–70)
LDL cholesterol test performed (%)	89
LDL cholesterol test (mg/dl)	104.77 (90.76–122.11)
LDL done and value <130 mg/dl (%)	78 (56–89)

Data are means ± SD and mean (range) unless otherwise indicated.

tion between facility level performance with the thresholds of 8 and 7 was extremely poor (Fig. 1, Pearson $r = 0.13$, $P = 0.14$).

To assess the impact of these two kinds of measures on league tables where facilities are ranked based on performance, we assessed rankings for A1C <7 and for the QALYsS measure that gave partial credit for A1C levels between 8 and 7. Overall, 31/141 facilities (22%) changed their decile ranking by >2. There were marked changes among those ranked in the top and bottom 10%. Figure 2 shows the degree of spread from the top and bottom deciles based on one ranking scheme compared with the other. For example, in Fig. 2A, the facility ranked at 129 (141 = best) by the dichotomous threshold measure, ranked 92 by the QALYsS measure; the facility ranked at 19 (1 = worst) rose to a ranking of 59. Similarly, in Fig. 2B, the facility ranked at 135 (141 = best) by the QALYsS measure dropped to a ranking of 63 by the dichotomous threshold measure; the facility

ranked at 13 (1 = worst) rose to a ranking of 47. Overall, of the 15 poorest performers (bottom 10%) assessed by dichotomous threshold, 3 moved up by at least 2 deciles when assessed with weighted measures. Of the 14 best performers (top 10%) by dichotomous measures, 3 moved down by at least 2 deciles when assessed with weighted measures.

CONCLUSIONS — Our study compared two dichotomous thresholds and a weighted continuous performance measure, based on QALYsS, for A1C. Whereas there was very high correlation between the two methods in facility level rankings using league tables, there were also marked changes in identifying best and worse performing facilities both between the methods and between the two dichotomous thresholds (<7 and <8%). Our findings are significant in demonstrating that even when there are high levels of correlation among dichotomous and weighted continuous measures, absolute rankings are very sensitive to the method used.

Weighted measures have been applied to more global measures, e.g., national population health (28,29). We have applied a weighting scheme for an individual measure for a subset of patients. The differences in rankings can most likely be attributed to the fact that by providing "partial credit" toward achieving the <7% dichotomous measure goals,

Table 2—*Facility level adherence: performance measures and QALYs gained by age category*

	Age (years)								Facility range
	25–34	35–44	45–54	55–64	65–74	75–84	85–94	Total	
n	191	1,202	5,951	7,325	12,459	9,443	571	37,142	—
Performance measure									
A1C <8.0%*	46	46	52	58	66	69	69	62	48–75
A1C <7.0%*	29	29	33	35	41	44	47	39	21–57
QALY maximum weights per patient (A1C <7.0%)†	0.6482	0.4575	0.2527	0.127	0.0507	0.0142	0.0017	0.1043	
Actual QALYsS	0.2417 ± 0.29	0.1736 ± 0.21	0.1082 ± 0.12	0.0596 ± 0.06	0.0275 ± 0.02	0.0082 ± 0.006	0.0010 ± 0.0008	0.0473 ± 0.083	
Percent maximal QALYsS†	37	38	43	47	54	58	60	45	31–60

Data are means ± SD unless otherwise indicated. *For the categories of A1C <8% and <7%, the percent adherence by age-group was calculated by determining the percent of people in that age-group who had A1C levels under the threshold. The number presented is the weighted average across all facilities. †Percent of maximal QALYs was a two-step calculation. In the first step, the maximal QALYs attainable by a facility was calculated by multiplying the number of people in each age category by the QALY weight assigned to that category. This assumes that all patients meet the best criteria. The second step calculates the actual QALYs attained by each facility. This is a function of whether a patient had an A1C test and his or her A1C test value. A1C <7.0%: QALY = QALY maximum weights (QMW). A1C >7.0% and <7.9%: QALY = QMW − [(A1C value − 7.0)/0.9] × QMW. A1C ≥7.9%: QALY = 0. The adherence (percent maximal QALYsS) is a percentage of actual to maximal QALYs. The higher this percentage, the better the levels of A1C testing and the better the A1C control among patients tested. The number presented is the weighted average adherence across all facilities.
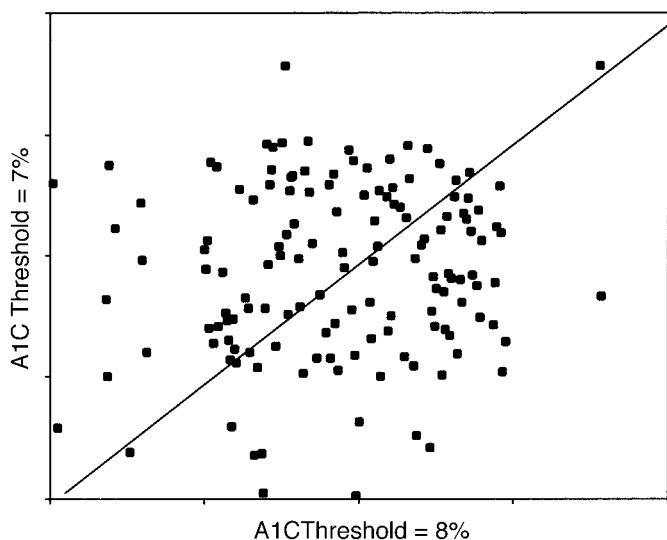
Figure 1—*Correlation between facility rankings by different dichotomous A1C thresholds. Each dot represents a facility.*

mance measures. This is the case especially with regard to public reporting and physician payment, because of issues of the generalizability of efficacy trials to actual practice. For example, although the mean value of A1C was 7.0% over the 10 years of the U.K. Prospective Diabetes Study, fewer than half of the subjects were able to achieve this value in the last year of the study, at least in part because type 2 diabetes (and the degree of hyperglycemia) worsens with duration of disease (33). Thus, even in well-conducted trials with extensive exclusion criteria, significant numbers of individuals are unable to reach and or sustain "optimal" target goals despite the ongoing effort of a comprehensive team (34,35). Consequently, caution is necessary in applying results to large numbers of people with diabetes
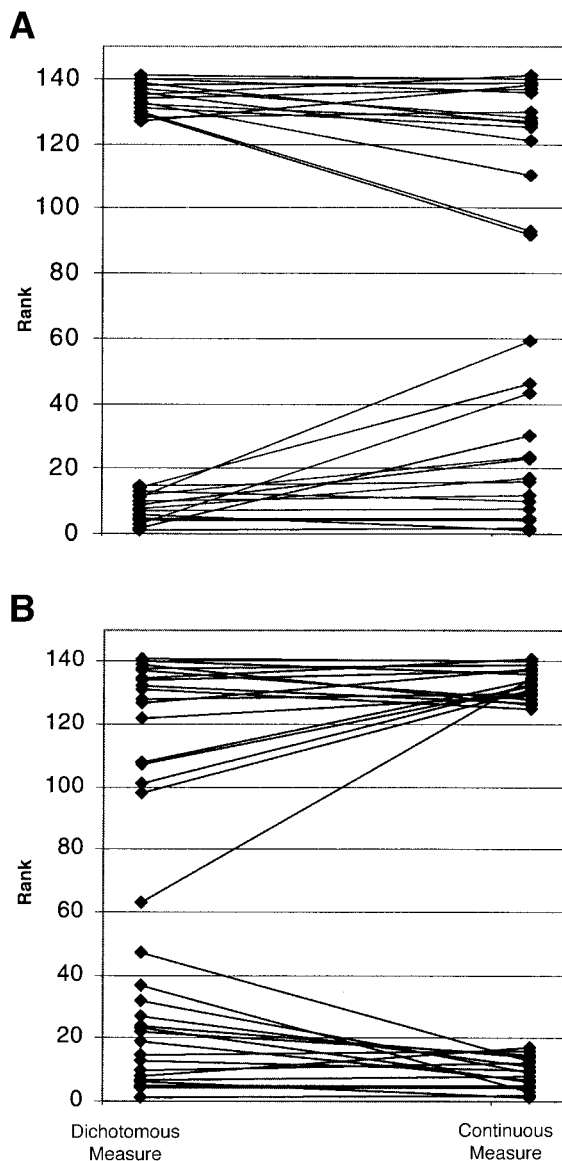
the QALYsS measure is assessing progress toward achieving thresholds rather than whether the targets were completely met. This may in part be due to factors beyond plan control, such as age and duration of diabetes, since QALYsS are age adjusted and dichotomous thresholds are not. Consequently, our weighted measure does risk-adjust for age differences among facilities, which in this study had a 2-SD magnitude of over 11 years, even in this relatively homogeneous health care system (30). There may also be variations in medical and psychiatric comorbidities, and economic status, among facility-level subjects that could affect shared decision-making on target values (31). This assumes particular importance in diabetes, where there is a differential impact of management of each of the major cardiovascular risk factors (22).

Although weighted measures have been applied to national population overall health, we are unaware of similar efforts applied to individual health care plans. However, if health is the output of the health care sector, then one rational approach of social policy would be to maximize health for each individual (32). QALYsS by their very nature reflect lower benefits for good control of blood pressure, cholesterol, and glucose for the elderly. However, the adherence measure (percent of maximal QALYsS) is defined as a proportion of actual to maximal and hence negates this consideration in its calculation. Consequently, our results have important implications for the developers and stakeholders in the use of perfor-

Figure 2—*League table changes for top and bottom 10% based on ranking method: A1C <7% dichotomous measure versus QALYsS-weighted continuous measure. A: Top and bottom deciles based on the dichotomous measure (values on left) and the associated spread in rankings for the same facilities based on the continuous measure (value on right). B: Top and bottom deciles based on the continuous measure (values on right) and the associated spread in rankings of the same facilities based on the dichotomous measure (values on left).*

who would not have been eligible for the study because of compliance or medical concerns (36). Therefore, we suggest that the use of a continuous measure that gauges improvement toward objectives rather than a pass/fail target value is more consistent with the concepts of quality improvement in bettering population health through shared physician-patient decision making (37). For example, improvement in A1C of 8.1 to 7.1% would not meet the 7% threshold criterion, whereas a fall from 7.1 to 6.9% (a drop that would result in little benefit in terms of QALYsS) would meet the criterion. In addition, we suggest that because "partial credit" is provided for reducing a modifiable risk factor, there may be fewer objections by clinicians who are not able to achieve "optimal" glycemic thresholds because of legitimate patient-level factors that are not easily obtained from administrative records, e.g., symptoms of frequent hypoglycemia or weight gain.

Our proposal has several strengths. Because the use of QALYsS can be easily calculated using A1C data currently collected for performance measurement, there would be no added burden to health care plans. Whether risk adjustment can be applied and how clinicians can receive additional credit based on intensity of treatment should be the subject of future research (38–40). Similarly, QALYsS may be used to calculate the cost-effectiveness of "evidence-based" interventions in clinical practice to more easily value physician and plan performance (41,42). However, because patients, physicians, administrators, and payors are accustomed to looking at adherence to thresholds, these data can be provided as accompanying information on a report card to make glycemic control more transparent.

Our study also has several limitations. The results depend on the validity of the estimates of utilities for the different health states, including the assumptions underlying QALY calculations. However, the model that we used has been applied to the U.S. population as a whole. Our data are cross-sectional but lend themselves to longitudinal analyses both at the facility level and at the individual patient level; with availability of electronic medical records, individual patient change scores could be calculated from year to year (8,43). We assessed this measure only at this level, where the number of patients exceeded 100 individuals, which has previously been demonstrated to be

reliable in demonstrating difference among practices (44). Our population was a largely male veteran population, which could limit its generalizability. On the other hand, recent articles have established the comparability between Veterans Health Administration and private sector commercial plan performance, and significant sex differences in QALYS have not been reported for A1C (15,16).

In conclusion, our study demonstrates that using percentage of maximal QALYsS as a weighted continuous performance measure is a feasible alternative to dichotomous thresholds in the assessment of quality of care for diabetes and has implications for reporting on and payment for health care plan and physician performance. We suggest that this alternative be investigated for other intermediate outcomes and in other settings.

References

1. National Committee for Quality Assurance: *What's In It and Why It Matters.* Washington, DC, National Committee for Quality Assurance, 1999
2. Agency for Health Care Research and Quality: National Health Care Quality Report [article online], 2005. Available at http://qualitytools.ahrq.gov/qualityreport/download_report.aspx. Accessed 4 February 2005
3. National Committee for Quality Assurance: The state of health care quality [article online], 2005. Available at http://www.ncqa.org/communications/SOMC/SOHC2004.pdf. Accessed 4 February 2005
4. National Quality Forum: National Quality Forum [article online], 2005. Available at http://www.qualityforum.org. Accessed 4 February 2005
5. American Medical Association: Clinical quality improvement and patient safety [article online], 2005. Available at http://www.amaassn.org/ama/pub/category/2936.html. Accessed 4 February 2005
6. Centers for Medicine & Medicaid Services: Nursing home compare. [article on-

line], 2005. Available at http://www.medicare.gov/Nhcompare/Home.asp. Accessed 4 February 2005
7. National Diabetes Quality Improvement Alliance: National Diabetes Quality Improvement Alliance [article online], 2005. Available at http://www.medicine.ucsd.edu/fpm/ndqia/organizations.html. Accessed 4 February 2005
8. Thompson B, Harris J: Are we measuring what matters? *Prev Med* 20:291–293, 2001
9. Landon B, Normand S, Blumenthal D, Daley J: Physician clinical performance assessment: prospects and barriers. *JAMA* 290:1183–1189, 2003
10. Werner R, Asch D: The unintended consequences of publicly reporting quality information. *JAMA* 293:1239–1244, 2005
11. Institute of Medicine: *Envisioning the National Healthcare Quality Report: Designing the National Healthcare Quality Report.* Washington, DC, National Academy Press, 2001
12. Centers for Disease Control and Prevention: National diabetes fact sheet [article online], 2005. Available at http://www.cdc.gov/diabetes/pubs/factsheet.htm. Accessed 4 February 2005
13. American Diabetes Association: Economic costs of diabetes in the U.S. in 2002. *Diabetes Care* 26:917–932, 2003
14. Fleming B, Greenfield S, Engelgau M, Pogach L, Clauser S, Parrott M: The diabetes quality improvement project: moving science into health policy to gain an edge on the diabetes epidemic. *Diabetes Care* 24:1815–1820, 2001
15. Kerr EA, Gerzoff RB, Krein SL, Selby JV, Piette JD, Curb JD, Herman WH, Marrero DG, Narayan KM, Safford MM, Thompson T, Mangione CM: Diabetes care quality in the Veterans Affairs Health Care System and commercial managed care: the TRIAD study. *Ann Intern Med* 141:272–281, 2004
16. Jha A, Perlin J, Kizer K, Dudley R: Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. *N Engl J Med* 348:2218–2227, 2003
17. Geiss L, Engelgau M, Pogach L, Acton K, Fleming B, Roman S, Han L, Wang J, Vinicor F: A national progress report on diabetes: successes and challenges. *Diabetes Technol Ther* 7:198–203, 2005
18. Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986, 1993
19. UK Prospective Diabetes Study (UKPDS) Group: Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 352:837–853, 1998

20. Vijan S, Hofer T, Hayward R: Estimate benefits of glycemic control in microvascular complications in type 2 diabetes. *Ann Intern Med* 127:788–795, 1997

21. Gold M, Siegel J, Russell L, Milton W: *Cost-Effectiveness in Health and Medicine.* New York, Oxford University, 1996

22. CDC Diabetes Cost-Effectiveness Group: Cost effectiveness of intensive glycemic control, intensified hypertension control, and serum cholesterol level reduction for type 2 diabetes. *JAMA* 287:2542–2551, 2002

23. Eastman R, Javitt J, Herman WH, Dasbach E, Dong F, Manninen D, Zbrozek A, Kotsanos J, Garfield S, Harris M: Model of complications of NIDDM. II. Analysis of the health benefits and cost-effectiveness of treating NIDDM with the goal of normoglycemia. *Diabetes Care* 20:735–744, 1997

24. Gray A, Raikou M, McGuire A, Fenn P, Stevens R, Cull C, Stratton I, Adler A, Holman R, Turner R: Cost effectiveness of an intensive blood glucose control policy in patients with type 2 diabetes: economic analysis alongside randomised controlled trial (UKPDS 41): United Kingdom Prospective Diabetes Study Group. *BMJ* 20:1373–1378, 2000

25. Miller D, Safford M, Pogach L: Who has diabetes? Best estimates of diabetes prevalence in the Veterans Health Administration based on computerized patient data. *Diabetes Care* 27:B10–B21, 2004

26. Stratton I, Adler A, Neil H, Matthews DR, Manley SE, Cull CA, Hadden D, Turner RC, Holman RR: Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35). *BMJ* 321:405–412, 2000

27. NCQA: NCQA report cards [article online], 2005. Available at http://hprc.ncqa.org/menu.asp. Accessed 4 February 2005

28. Mathers C, Sadana R, Salomon J, Murray C, Lopez A: Healthy life expectancy in 191 countries, 1999. *Lancet* 357:1685–1691, 2001

29. Murray C, Frenk J: A framework for assessing the performance of health systems. *Bull World Health Organ* 78:717–731, 2000

30. Kizer K: The "new VA": a national laboratory for health care quality management. *L Med Qual* 14:3–20, 1999

31. Pogach LM, Brietzke SA, Cowan CL Jr, Conlin P, Walder DJ, Sawin CT, the VA/DoD Diabetes Guideline Development Group: Development of evidence-based clinical practice guidelines for diabetes: the Department of Veterans Affairs/Department of Defense Guidelines Initiative. *Diabetes Care* 27 (Suppl. 2):B82–B89, 2004

32. Garber A, Weinstein M, Torrance G, Kamlet M: In *Cost-Effectiveness in Health and Medicine.* Gold M, Siegel J, Russell L, Milton W, Eds. Oxford, U.K., Oxford University, 1996

33. Turner R, Cull C, Frighi V, Holman R, for the UK Prospective Diabetes Study (UKPDS) Group: Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). *JAMA* 281:2005–2012, 1999

34. Fritsche A, Schweitzer M, Haring H, 4001 Study Group: Glimepiride combined with morning insulin glargine, bedtime neutral protamine hagedorn insulin, or bedtime insulin glargine in patients with type 2 diabetes. *Ann Intern Med* 138:952–959, 2003

35. Riddle M, Rosenstock J, Gerich J, Insulin Glargine 4002 Study Investigators: The treat-to-target trial: randomized addition of glargine or human NPH insulin to oral therapy of type 2 diabetic patients. *Diabetes Care* 26:3080–3086, 2003

36. Vijan S, Kent D, Hayward R: Are randomized controlled trials sufficient evidence to guide clinical practice in type II (non-insulin-dependent) diabetes mellitus? *Diabetologia* 43:125–130, 2000

37. Cobin R, Rosof B, Sirio C: Uses of published physician performance data. *JAMA* 290:3194–3195, 2003

38. Chin M: Risk-adjusted quality of care rating for diabetes: ready for prime time? *Diabetes Care* 23:884–886, 2000

39. Zhang Q, Safford M, Ottenweller J, Hawley G, Repke D, Burgess JF Jr, Dhar S, Cheng H, Naito H, Pogach LM: Performance status of health care facilities changes with risk adjustment of HbA$_{1c}$. *Diabetes Care* 23:919–927, 2000

40. Berlowitz, DR, Ash AS, Glickman M, Friedman RH, Pogach LM, Nelson AL, Wong AT: Developing a quality measure for clinical inertia in diabetes care. *Health Serv Res* 40:1836–1853, 2005

41. Eggleston K, Grossman J, Cutler D: Productivity research and healthcare delivery innovation. *Appl Health Econ Health Policy* 3:133–141, 2004

42. Jones N 3rd, Jones S, Miller N: The medicare health outcomes survey program: overview, context, and near-term prospects. *Health Qual Life Outcomes* 2:33, 2004

43. Weiner M, Long J: Cross-sectional versus longitudinal performance assessments in the management of diabetes. *Med Care* 42:II34–II39, 2004

44. Hofer T, Hayward R, Greenfield S, Wagner E, Kaplan S, Manning W: The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA* 281:2098–2105, 1999