Quality of Diabetes Care Predicts the Development of Cardiovascular Events: Results of the AMD-QUASAR Study

MARIA C.E. ROSSI, MSC, PHARM, CHEM[1]
GIUSEPPE LUCISANO, MSC, STAT[1]
MARCO COMASCHI, MD[2]
CARLO COSCELLI, MD[3]
DOMENICO CUCINOTTA, MD[4]
PATRIZIA DI BLASI, MD[5]
GIOVANNI BADER, MD[5]

FABIO PELLEGRINI, MSC, STAT[1]
UMBERTO VALENTINI, MD[6]
GIACOMO VESPASIANI, MD[7]
ANTONIO NICOLUCCI, MD[1]
ON BEHALF OF THE AMD-QUASAR STUDY GROUP*

OBJECTIVE—The QUASAR (Quality Assessment Score and Cardiovascular Outcomes in Italian Diabetes Patients) study aimed to assess whether a quality-of-care summary score predicted the development of cardiovascular (CV) events in patients with type 2 diabetes.

RESEARCH DESIGN AND METHODS—In 67 diabetes clinics, data on randomly selected patients were extracted from electronic medical records. The score was calculated using process and outcome indicators based on monitoring, targets, and treatment of A1C, blood pressure, LDL cholesterol, and microalbuminuria. The score ranged from 0 to 40.

RESULTS—Overall, 5,181 patients were analyzed; 477 (9.2%) patients developed a CV event after a median follow-up of 28 months. The incidence rate (per 1,000 person-years) of CV events was 62.4 in patients with a score of <15, 41.0 in those with a score between 20 and 25 and 36.7 in those with a score of >25. Multilevel analysis, adjusted for clustering and case-mix, showed that the risk to develop a new CV event was 84% higher in patients with a score of <15 (incidence rate ratio [IRR] = 1.84; 95% confidence interval [CI] 1.29–2.62) and 17% higher in those with a score between 15 and 25 (IRR = 1.17; 95% CI 0.93–1.49) compared with those with a score of >25. Mean quality score varied across centers from 16.5 ± 7.5 to 29.1 ± 6.3. When the score was tested as the dependent variable, it emerged that 18% of the variance in the score could be attributed to setting characteristics.

CONCLUSIONS—Our study documented a close relationship between quality of diabetes care and long-term outcomes. A simple score can be used to monitor quality of care and compare the performance of different centers/physicians.

The growing burden of type 2 diabetes mellitus and its cardiovascular complications (1–2) has forced American and European organizations to work for the development, specification, and field-testing of measures for quality of diabetes care (3–9).

Quality measures include process and intermediate outcome indicators, selected under the assumption that they are linked to downstream health outcomes. Process measures denote what is actually done to the patient (i.e., whether A1C [HbA$_{1c}$] has been measured or an ACE inhibitor prescribed in the presence of a specific indication). Outcome measures are the results of a patient health status as a consequence of the care delivered. Intermediate outcomes include laboratory measurements, physical signs, or symptoms.

Although widely used by many organizations worldwide, it is unclear whether selected indicators, which usually include process and intermediate outcome measures, truly reflect the quality of care delivered and its long-term consequences (10). Outcome measures have been criticized because they can be affected by factors other than the quality of care, such as patient characteristics and attitudes (11). However, process measures alone do not fully describe the whole process of care. As an example, regular testing of HbA$_{1c}$ levels does not ensure that the physician will adequately treat persistently elevated values (12). Furthermore, positive results in the short run could fail to be sustained in the long run, even in the presence of scientifically validated indicators. In fact, although widely used, it is unclear to what extent process and intermediate outcome measures are able to predict long-term effects on patients' health. The Associazione Medici Diabetologi (AMD), an Italian diabetes scientific society, developed quality indicators to be used for the routine monitoring of diabetes care (9,13). In parallel, AMD launched the QUASAR (Quality Assessment Score and Cardiovascular Outcomes in Italian Diabetes Patients) study to evaluate whether these indicators were able to predict the incidence of final outcomes. In particular, we assessed whether a quality-of-care summary score estimated at baseline from the combination of the main process and intermediate outcome indicators predicted the development of cardiovascular (CV) events during 30 months.

RESEARCH DESIGN AND METHODS

The Italian health care system
All Italian citizens are covered by a government health insurance and are registered

with a general practitioner (GP). Primary care for diabetes is provided by GPs and diabetes outpatient clinics (DOCs). Patients can choose one of these two ways of accessing the health care system or they can be referred to DOCs by their GPs. Within the Italian health care system more than 600 DOCs are in operation.

Study design

This observational longitudinal study involved DOCs adopting an electronic medical record system for the routine management of outpatients.

Each center was asked to enroll up to 100 patients, selected among individuals having a scheduled visit during 6 months, using random sampling lists. Eligible individuals were men and women patients with type 2 diabetes mellitus, aged ≥18 years, irrespective of diabetes treatment and duration. Newly diagnosed patients were not included. At baseline, clinical data were automatically extracted from electronic clinical records using an ad hoc software. Information on CV events that occurred during three years was collected on ad hoc case report forms (Supplementary Fig. A1). All the data were analyzed anonymously, and patients were identified only by a numeric code.

Patients were enrolled between January 2006 and November 2007 and followed up until September 2009. The study was approved by local ethics committees of all participating centers.

Quality-of-care summary score

A quality-of-care summary score (Table 1), developed and validated in a previous study (14), was applied to the QUASAR population to confirm its ability to predict the incidence of major vascular events.

The score includes those process and intermediate indicators for which a clear link with vascular complications was demonstrated and effective preventive strategies are available.

For each item, the highest score was assigned when the desired goals was attained, whereas the lowest score was assigned when the patient was not treated for the specific condition despite elevated values or when the patient showed unsatisfactory values despite the treatment. Monitoring HbA$_{1c}$, blood pressure, lipids, or microalbuminuria less frequently than once a year was attributed an intermediate scoring under the assumption that these parameters could be less frequently monitored in individuals with less severe diabetes and in better control.

Table 1—Quality-of-care scoring system

Quality-of-care indicator	Scoring
HbA$_{1c}$	
Measurement <1/year	5
≥8.0%	0
<8.0%	10
Blood pressure	
Measurement <1/year	5
Values ≥140/90 mmHg,	
irrespective of treatment	0
Values <140/90 mmHg	10
Lipid profile measurement <1/year	5
LDL cholesterol, mmol/L (130 mg/dL)	
≥3.37, irrespective of treatment	0
<3.37	10
MA measurement <1/year	5
Not treated with ACE inhibitors/ARBs	
despite the presence of MA	0
Treated with ACE inhibitors/ARBs in	
the presence of MA or MA absent	10
Score range	0–40

MA, microalbuminuria; ARBs, angiotensin receptor blockers.

After having developed the score in the QuED (Quality of Care and Outcomes in Type 2 Diabetes) study, we have empirically tested this aspect in the QUASAR study, where we have also collected blood samples, which were centrally analyzed. From these data it emerged that patients monitored less frequently tended to have lower HbA$_{1c}$ and lipid levels and were markedly less likely to have microalbuminuria.

To avoid a strict dependence from the data, the scores attributed to each process and outcome measure were decided a priori and not based on weights derived from regression models applied to the data. Overall, the quality score ranges between 0 and 40, with a higher score reflecting better quality of care. The scoring system was applied to the data collected at study entry and tested as a predictor of incident CV events. Indicators were calculated using the last information relative to the 12 months preceding the recruitment. Because normal ranges for glycated hemoglobin varied among different centers, the percent change with respect to the upper normal value (actual value/upper normal limit) was estimated and multiplied by 6.0.

LDL cholesterol was estimated by the Friedwald equation. The score only included those indicators that could be easily retrieved from medical records. Additional indicators, such as foot or eye examination, although representing key components of diabetes care, were not considered, since information was reported in nonstandardized format in medical records.

Incident CV events

Information on CV events that occurred over three years was collected. The primary composite end point was represented by total CV events, including angina, myocardial infarction (MI), stroke, transient ischemic attack, coronary revascularization procedures, lower-limb complications (claudication, ulcer, gangrene, amputation, aortic-femoral revascularization procedures), and CV mortality.

Secondary end points were represented by major CV events (nonfatal MI, nonfatal stroke, CV mortality), MI, cardiac revascularization procedures, stroke, lower-limb complications, CV deaths, and total deaths.

MI was defined based on the presence of at least two of the following: *1*) typical ischemic chest pain, pulmonary edema, syncope, or shock; *2*) development of pathological Q-waves and/or appearance or disappearance of localized ST-elevation followed by T-wave inversion in two or more of twelve standard electrocardiograph leads; and *3*) increase in the concentration of serum enzymes consistent with MI.

Stroke was defined on the basis of the presence of unequivocal signs or symptoms of focal or global neurologic deficit, with sudden onset, lasting more than 24 h, and considered of vascular origin.

The following criteria were adopted to define CV death: death occurring within 28 days after the onset of the documented diagnosis of MI or stroke, in the absence of any other evident cause; sudden deaths; deaths from heart failure; and all other deaths classified as cardiovascular (International Classification of Disease, ninth edition, Clinical Modification). The absence of any evident noncardiovascular cause was sufficient to define a death as cardiovascular.

All of the events were certified by the participating clinicians and were based on clinical documentation. Patients were contacted by phone if their last visit to the center had occurred more than 6 months before.

Statistical analysis

Based on the results of the previous study, the quality score was categorized in three

classes; however, because of a general trend of improvements in the average quality of care in the last years, the cutoffs of the middle class were shifted ahead by 5 points (<15, 15–25, >25). Patients' characteristics according to score classes were compared using the χ^2 test for categorical variables and the Kruskal-Wallis one-way ANOVA for continuous variables.

Incidence rates (IR) by score classes were expressed in terms of events per 1,000 person-years.

Probabilities of CV event according to score classes were calculated using the Kaplan-Meier method and compared using the log-rank test. All reported *P* values are two-sided.

To account for the multilevel nature of the data (patients clustered within diabetes clinic), and to control simultaneously for the possible confounding effects of the different covariates, we used multilevel Poisson regression models to investigate whether the score was an independent predictor of CV events incidence (15). In multilevel analyses patient-related characteristics were tested as level 1 variables, whereas setting-related characteristics represented level 2 variables. The following patients' characteristics were tested: age, sex, BMI, duration of diabetes and treatment, smoking, and history of previous CV event. The risk of CV events according to the score classes was expressed in terms of incidence rate ratio (IRR) and 95% confidence intervals (95% CIs), with a score of >25 as the reference category.

Finally, we estimated the intraclass correlation coefficient (ICC) using the score as dependent variable to evaluate the extent to which the score varies between centers as compared with within-center variation, taking patient case-mix into account (16). The higher the ICC, the greater the influence of the physician-level on the quality-of-care score.

RESULTS—The study was conducted in 67 diabetes clinics, and 5,181 patients were enrolled and followed up for a median of 28 months (interquartile range 24–31). Overall, 6.7% of the patients had a score of <15, 68.2% had a score between 15 and 25, and 25.1% had a score of >25. Patients' characteristics according to the score classes are reported in Table 2. Patients with the lowest score had a longer diabetes duration, and a larger percentage of them were already

Table 2—Patient characteristics according to quality-of-care score

Characteristic	Quality-of-care score			*P* *
	<15	15–25	>25	
N (%)	345 (6.6)	3,531 (68.2)	1,305 (25.2)	
Sex (% men)	51.6	57.8	61.8	0.001
Age (years)	64.7 ± 8.9	64.9 ± 9.2	63.0 ± 9.6	<0.0001
BMI (kg/m^2)	30.7 ± 5.2	29.9 ± 5.0	29.2 ± 4.8	<0.0001
Smoking				<0.0001
No	66.5	61.8	51.6	
Ex	16.3	16.9	21.2	
Yes	17.3	21.3	27.3	
Diabetes duration (years)	11.9 ± 8.5	10.2 ± 8.7	9.1 ± 8.0	<0.0001
Treatment				<0.0001
Diet alone	0.7	7.7	10.2	
Oral agents	60.6	68.5	72.0	
Insulin ± oral agents	38.7	23.8	17.8	
Previous CV event	20.0	18.9	21.3	0.18
HbA$_{1c}$ (%)	9.1 ± 1.3	7.5 ± 1.4	6.8 ± 1.0	<0.0001
Blood pressure (mmHg)				
Systolic	152.2 ± 15.4	144.9 ± 17.0	128.1 ± 13.6	<0.0001
Diastolic	85.6 ± 9.4	82.6 ± 9.2	77.3 ± 7.5	<0.0001
Cholesterol (mg/dL)				
Total	224.3 ± 41.2	193.5 ± 39.9	177.3	<0.0001
LDL	152.3 ± 28.1	115.0 ± 35.0	100.2 ± 25.2	<0.0001
HDL	50.8 ± 11.6	51.6 ± 13.2	50.4 ± 13.3	0.03
Triglycerides (mg/dL)	174.9 ± 107.5	146.6 ± 126.73	138.1 ± 82.8	<0.0001

Data are percentages or mean ± SD. **P* values refer to χ^2 for categorical variables and Kruskal-Wallis one-way ANOVA for continuous variable.

treated with insulin as compared with patients in the other two score classes.

During the follow-up, 477 (9.2%) patients developed one or more CV events. The CV event rate was strictly related to the score, being of 62.1 per 1,000 person-years in patients with a score of <15, 41.0 per 1,000 person-years in those with a score between 15 and 25, and 36.7 per 1,000 person-years in those with a score of >25. Unadjusted Kaplan-Meier curves show that a score of <15 was associated with a significant higher likelihood to develop a CV event (Fig. 1).

Multilevel analysis, adjusted for clustering, age, sex, BMI, duration of diabetes, smoking, and history of CV event, showed that the risk to develop a new CV event was 84% greater in patients with a score of <15 (IRR = 1.84; 95% CI 1.29–2.62) and 17% higher in those with a score between 15 and 25 (IRR = 1.17; 95% CI 0.93–1.49), as compared with those with a score of >25. A trend of increasing overall mortality according to the quality score was also documented, although statistical significance was not reached (IRR = 1.68; 95% CI 0.78–3.63 and IRR = 1.35; 95% CI 0.83–2.21 for a

score of <15 and between 20 and 25, respectively, compared with a score of >25. The same trend of association was found for most of the outcomes, although statistical significance was reached for lower limb complications only (Fig. 2). The strong correlation between the score of quality of care and CV events was evident in patients without previous CV events as well as in those who had already experienced a previous event (Supplementary Table A1).

Finally, the mean quality score varied substantially across centers (from 16.5 ± 7.5 to 29.1 ± 6.3). Mean quality score by center, adjusted for patient characteristics and clustering effect, is in Supplementary Fig. A2. When the score was tested as the dependent variable, it emerged that 25% of the variance in the score (ICC = 0.25) could be attributed to between center variation (i.e., to setting characteristics). When adjusting for case-mix (age, sex, BMI, duration of diabetes, previous CV events), variance attributable to center-level characteristics remained elevated (ICC = 0.18).

CONCLUSIONS—Improving health outcomes is the ultimate goal of the health

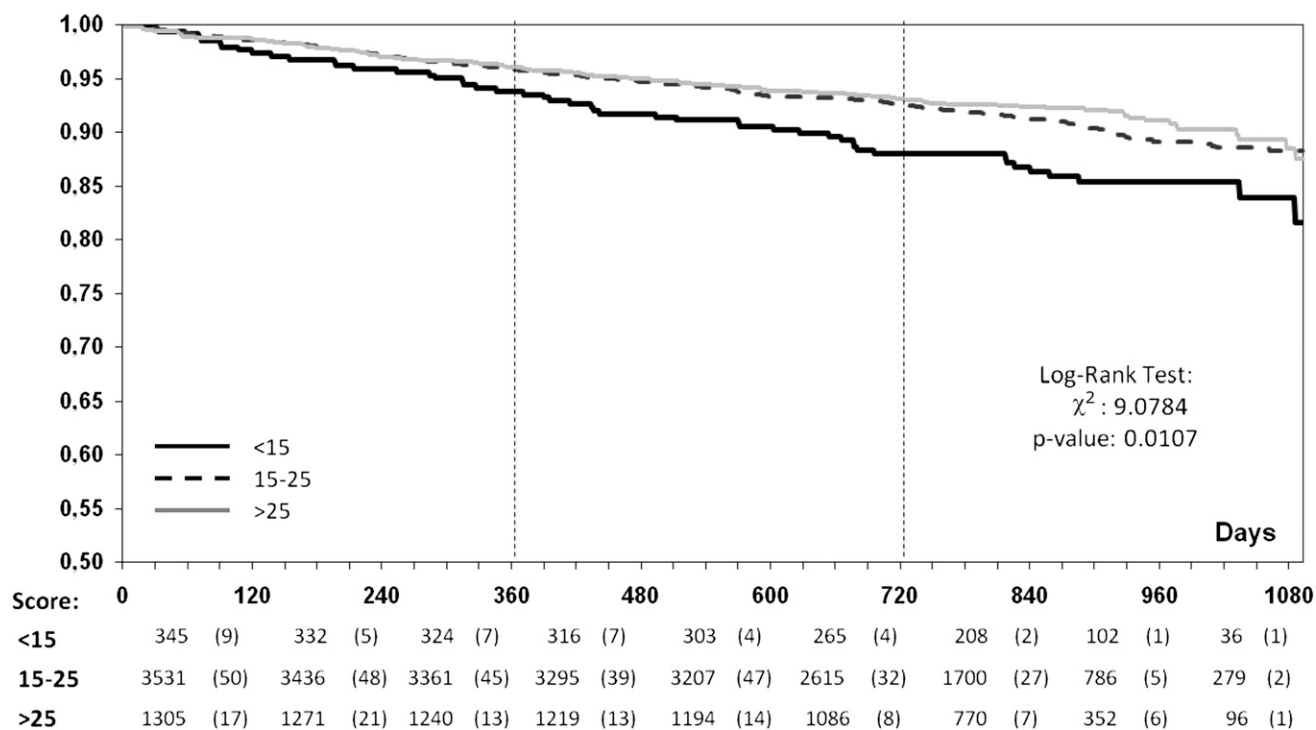| Score:	0		120		240		360		480		600		720		840		960		1080	
<15	345	(9)	332	(5)	324	(7)	316	(7)	303	(4)	265	(4)	208	(2)	102	(1)	36	(1)
15-25	3531	(50)	3436	(48)	3361	(45)	3295	(39)	3207	(47)	2615	(32)	1700	(27)	786	(5)	279	(2)
>25	1305	(17)	1271	(21)	1240	(13)	1219	(13)	1194	(14)	1086	(8)	770	(7)	352	(6)	96	(1)

Figure 1—*Event-free survival; all patients by group score.*

care system and should represent the main aim for any quality measurement program. From the data on more than 5,000 patients with type 2 diabetes mellitus routinely cared for by diabetes clinics and followed up for a median of 28 months, we confirmed that a quality-of-care summary score, based on a restricted number of indicators, is able to predict total cardiovascular events. The trend of increase in the event rates with the progressive reduction of the score is clear for all the selected outcomes, including overall mortality, although statistical significance was not reached.

Our study substantially reproduces the body of evidence emerged by the QuED study (14), confirming the validity and applicability of the instrument. In comparison with previous studies (14,17), conducted ten years ago in comparable settings, quality of care has improved, as reflected by the lower percentage of patients scoring below 15 and the general shift of the score distribution toward higher values. Patients with a particularly high risk of negative outcomes, i.e., those with a score of <15, now represent a small minority (6.7%). Acting to improve the performance in this small subgroup would hardly translate into tangible results at the population level, although individual patients would benefit from any quality improvement. On the

other side, a large proportion of patients (i.e., 68.2%) lies in the 15–25 score range, associated with a nearly 20% increased risk of major events. A substantially greater benefit could thus be expected by improving the monitoring and treatment of major cardiovascular risk factors in such a large population. The score can help identifying patients whose quality of care is suboptimal and in which aspects of care need to be improved.

Reducing variation in health care delivery represents a key component of any quality improvement initiative (18). In our study, mean quality score widely differed among centers. It has been argued that process and intermediate outcomes can be affected not only by medical interventions but also by patient factors (19); in other words, the score might reflect patient variation rather than disparities in the quality of care. We showed that a relevant proportion of total variance was explained by center-level characteristics (16), even after adjusting for patient case-mix. To this respect, the score seems to efficiently capture differences in the performance of the centers. The wide variation in health care delivery strongly suggests that a substantial proportion of cardiovascular events could be avoided by providing higher quality and more homogeneous levels of care. The use of more

standardized approaches could lead to sizeable savings in care and outcomes, particularly focusing the attention on the large proportion of patients falling in the intermediate category of the score.

The possibility of calculating the quality score from electronic medical records in an easy and quick way makes its use particularly appealing. By allowing its automatic calculation, the score could be used not only to identify patients at risk for poor quality of care but also to monitor the performance of the center across the years. The score can also be used for multicenter, continuous quality improvement initiatives by providing a benchmark to which to compare individual performance. This approach is being adopted in the context of a wide national initiative in Italy, involving over 250 diabetes clinics for a total of half a million patients. The score will be calculated at the national, regional, and individual center levels, and the overall performance will be measured against that of the best performers.

Some of the potential limitations of our study need to be discussed. First, physicians were selected according to their willingness to participate. They could thus represent those clinicians more interested in diabetes care and therefore not reflect the diabetes care delivered by Italian physicians. It is thus
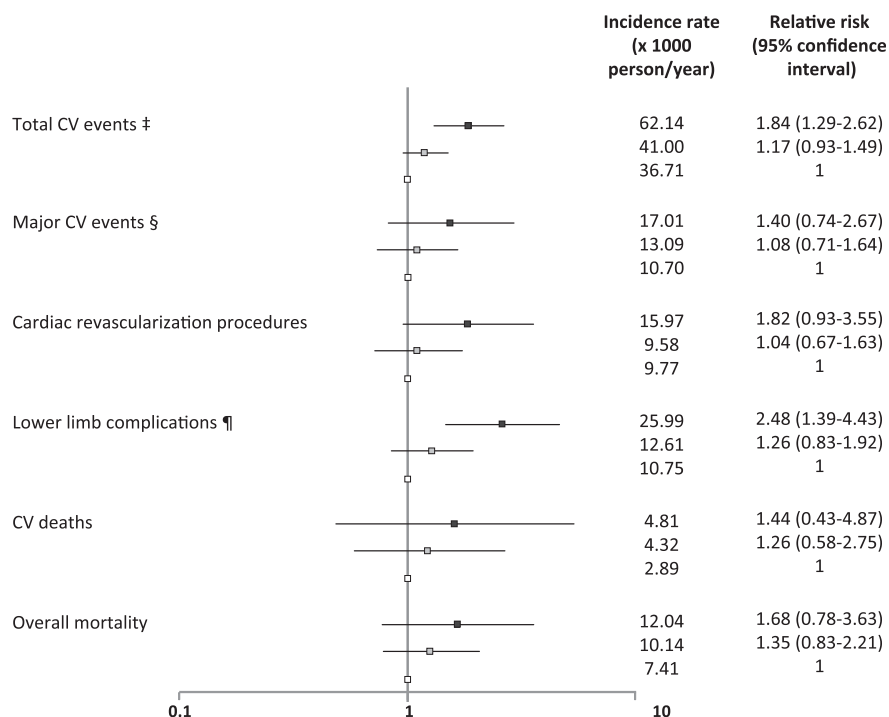
Figure 2—*Risk of developing CV events according to quality of care score classes: results of multilevel analyses. Data are crude incidence rates (IRs) and multilevel IRRs adjusted for age, sex, BMI, duration of diabetes, smoking, and history of previous CV event. Squares denote relative risks, and horizontal lines represent 95% CIs. The position of each square indicates the point estimate of the risk associated with a score of <15 (black squares), between 15 and 25 (gray squares), and >25 (white squares). The latter represents the reference category. ‡Total CV events include angina, MI, stroke, transient ischemic attack, coronary revascularization procedures, lower-limb complications, CV deaths; §major CV events include MI, stroke, and CV deaths; ¶lower-limb complications include claudication, ulcer, gangrene, amputation, and aortic-femoral revascularization procedures.*

The figure reports the following incidence rates (×1000 person/year) and relative risk (95% confidence interval) values:

	Incidence rate (x 1000 person/year)	Relative risk (95% confidence interval)
Total CV events ‡	62.14	1.84 (1.29–2.62)
	41.00	1.17 (0.93–1.49)
	36.71	1
Major CV events §	17.01	1.40 (0.74–2.67)
	13.09	1.08 (0.71–1.64)
	10.70	1
Cardiac revascularization procedures	15.97	1.82 (0.93–3.55)
	9.58	1.04 (0.67–1.63)
	9.77	1
Lower limb complications ¶	25.99	2.48 (1.39–4.43)
	12.61	1.26 (0.83–1.92)
	10.75	1
CV deaths	4.81	1.44 (0.43–4.87)
	4.32	1.26 (0.58–2.75)
	2.89	1
Overall mortality	12.04	1.68 (0.78–3.63)
	10.14	1.35 (0.83–2.21)
	7.41	1

possible that the differences documented in our study are underestimated and that the true variability in process and outcome measures is even greater. Similarly, patients attending diabetes clinics might differ from those primarily cared for by GPs; therefore, they could not be representative of all patients with diabetes. Second, the attainment of the desired therapeutic goals could be at least partially dependent on unmeasured patient factors; therefore, the intermediate outcome measures considered might not fully reflect the care provided. Third, the relatively short follow-up did not allow a robust estimate of the impact of the score on individual cardiovascular outcomes. Nevertheless, our results are highly consistent with those of the QuED study. Furthermore, an update of the study has been scheduled to evaluate the incidence rate of CV events after a 5-year follow-up. Finally, we have developed a score able to predict major cardiovascular events; a comprehensive quality-of-care measure should obviously also include process and outcome measures related to microvascular complications.

In conclusion, our study documented a close relationship between quality of diabetes care and long-term outcomes. Our findings support the use of the indicators identified as an important tool to assess the level of care provided within systems of care to populations of patients with diabetes.

References

1. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. Diabetes Care 1998;21:1414–1431
2. Zimmet P, Alberti KG, Shaw J. Global and societal implications of the diabetes epidemic. Nature 2001;414:782–787
3. Fleming BB, Greenfield S, Engelgau MM, Pogach LM, Clauser SB, Parrott MA. The Diabetes Quality Improvement Project: moving science into health policy to gain an edge on the diabetes epidemic. Diabetes Care 2001;24:1815–1820
4. Smith JJ. NCQA/HEDIS guidelines for diabetes. Manag Care 2001;10(Suppl. 2):3–5
5. Saaddine JB, Engelgau MM, Beckles GL, Gregg EW, Thompson TJ, Narayan KM. A diabetes report card for the United States: quality of care in the 1990s. Ann Intern Med 2002;136:565–574
6. Jencks SF, Cuerdon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. JAMA 2000;284:1670–1676
7. EUCID - Health & Consumer Protection Directorate - General. Final report European Core Indicators in Diabetes project [article online], 2005. Available from http://ec.europa.eu/health/ph_projects/2005/action1/docs/action1_2005_frep_11_en.pdf. Accessed 27 January 2009
8. Nicolucci A, Greenfield S, Mattke S. Selecting indicators for the quality of diabetes care at the health systems level in OECD countries. Int J Qual Health Care 2006;18(Suppl. 1):26–30
9. Rossi MC, Nicolucci A, Arcangeli A, et al.; Associazione Medici Diabetologi Annals Study Group. Baseline quality-of-care data from a quality-improvement program implemented by a network of diabetes outpatient clinics. Diabetes Care 2008;31:2166–2168
10. Kerr EA, Krein SL, Vijan S, Hofer TP, Hayward RA. Avoiding pitfalls in chronic disease quality measurement: a case for the next generation of technical quality measures. Am J Manag Care 2001;7:1033–1043
11. Kerr EA, Smith DM, Hogan MM, et al. Building a better quality measure: are some patients with 'poor quality' actually getting good care? Med Care 2003;41:1173–1182
12. Hayward RA. All-or-nothing treatment targets make bad performance measures. Am J Manag Care 2007;13:126–128
13. Nicolucci A, Rossi MC, Arcangeli A, et al.; AMD-Annals Study Group. Four-year impact of a continuous quality improvement effort implemented by a network of

diabetes outpatient clinics: the AMD-Annals initiative. Diabet Med 2010;27: 1041–1048

14. De Berardis G, Pellegrini F, Franciosi M, et al.; QuED (Quality of Care and Outcomes in Type 2 Diabetes) Study Group. Quality of diabetes care predicts the development of cardiovascular events: results of the QuED study. Nutr Metab Cardiovasc Dis 2008;18:57–65

15. Snijders TAB, Bosker RJ. Multilevel analysis. An introduction to basic and advanced multilevel modeling. London, SAGE Publications, 1999

16. Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. Improving the reliability of physician performance assessment: identifying the "physician effect" on quality and creating composite measures. Med Care 2009;47:378–387

17. Comaschi M, Coscelli C, Cucinotta D, Malini P, Manzato E, Nicolucci A; SFIDA Study Group—Italian Association of Diabetologists (AMD). Cardiovascular risk factors and metabolic control in type 2 diabetic subjects attending outpatient clinics in Italy: the SFIDA (survey of risk factors in Italian diabetic subjects by AMD) study. Nutr Metab Cardiovasc Dis 2005;15:204–211

18. Hayward RA, Hofer TP, Kerr EA, Krein SL. Quality improvement initiatives: issues in moving from diabetes guidelines to policy. Diabetes Care 2004;27(Suppl. 2): B54–B60

19. Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. Ann Intern Med 2002;136:111–121