Title: Reproducibility of diabetes quality of care indicators as reported by patients and physicians.
Authors: Collet TH, Taffé P, Bordet J, Burnand B, Peytremann-Bridevaux I
Journal: European journal of public health
Year: 2014 Dec
Volume: 24
Issue: 6
Pages: 1004-9
DOI: 10.1093/eurpub/cku011

Reproducibility of diabetes quality of care indicators as reported by patients and physicians

Authors

Tinh-Hai Collet, MD[1,2], Patrick Taffé, PhD[3], Julie Bordet, PhD[3], Bernard Burnand, MD, MPH[3], Isabelle Peytremann-Bridevaux, MD, MPH, DSc[3]

[1] Department of Ambulatory Care and Community Medicine, University of Lausanne, Switzerland; [2] Service of Endocrinology, Diabetes, and Metabolism, Lausanne University Hospital, Switzerland; [3] Institute of Social and Preventive Medicine, Lausanne University Hospital, Switzerland.

Corresponding author

Isabelle Peytremann-Bridevaux, MD, MPH, DSc
Institute of Social and Preventive Medicine (IUMSP)
Route de la Corniche 10
CH-1010 Lausanne
Switzerland
Tel: +41-21-314.72.84
Fax: +41-21-314.49.54
Email: Isabelle.Peytremann-Bridevaux@chuv.ch

Word count

Abstract: 250 words

Main text: 2984 words

References: 25 refs

Tables / Figures: 4 / 0

Funding

This project was funded by the Department of Public Health of the canton of Vaud, Switzerland ("Programme cantonal Diabète") and by the Swiss National Science Foundation. Dr T-H Collet's research is supported by a grant from the Swiss National Science Foundation (PBLAP3-145870). Dr. I. Peytremann-Bridevaux is supported by a grant from the Swiss National Science Foundation [PROSPER N° 32333B-123817 and N° 32333B-139789].

Abstract

Introduction

Self-report of diabetes care has moderate validity and is prone to under- and over-reporting. We assessed reproducibility of a range of processes and outcomes of diabetes care as reported by patients and physicians.

Methods

In a Swiss community-based survey, patients with diabetes and physicians independently reported past 12 months processes of care (HbA1c, lipids, microalbuminuria, blood pressure, weight, foot and eye examinations) and last measured values of HbA1c, height, weight, and blood pressure. For dichotomous variables, we assessed reliability by Cohen's kappa and agreement by uniform kappa. For continuous measures, we used Lin's concordance correlation coefficient and limits of agreement, respectively.

Results

Mean age of the 210 patients was 65 years, 40% were women and 51% had diabetes for >10 years. Agreement was good for recommended processes of care such as blood pressure (uniform kappa=0.94), HbA1c (0.93), weight (0.88), and lipid (0.78), but lower for microalbuminuria, foot and eye examinations (all <0.50). Cohen's kappa values were all low (<0.25). Comparisons of reported continuous variables showed large limits of agreement for height (±6cm) and weight (8-10kg) despite high concordance correlation coefficients (0.93 and 0.97). Concordance correlation coefficients were smaller for HbA1c (0.72) and blood pressure (0.5-0.6) with large limits of agreement (±2% and ±25 mmHg).

Conclusion

While agreement of routine processes of care was good, agreement was less satisfactory for microalbuminuria, foot and eye examinations. Reports of continuous outcomes yielded good reliability but too wide limits of agreement. Quality of care evaluation relying on self-report only should be made cautiously.

Keywords

Quality of care, diabetes, patient and physician report, reliability, agreement

Introduction

Chronic disease care requires good collaboration between healthcare providers, teamwork, self-management education and use of evidence-based medicine.[1] However, chronic care is a complex process often shown to be suboptimal.[2-4] Diabetes, a major public health burden in terms of morbidity, disability and mortality,[5] is not an exception, and improving quality of diabetes care is needed to reduce the health and societal burden of this disease.[6]

Self-report of processes and outcomes of care is often used as a source of information. Indeed, health care providers sometimes need to rely on patient self-report for their monitoring, especially in health care settings lacking shared electronic health records. In addition, self-report is used in community-based surveys since it is an easier mean of collecting data than examination-based surveys.

Self-report of diabetes-specific data has been shown to present moderate validity and be prone to both under- and over-reporting.[7-10] However, studies assessing the validity of self-reported diabetes measures have frequently only focused on a small number of indicators, such as foot and eye examinations[10] or levels of glycosylated haemoglobin (HbA1c).[8] To our knowledge, assessment of reliability and agreement has not been conducted on a range of process and outcome indicators of care. A targeted evaluation of more indicators may help define whether some may be more suitable for self-report in future surveys.

Within the context of a regional diabetes program in Switzerland, we conducted a community-based survey to first characterize patients with diabetes who were living in that region and second assess the quality of their care.[11] We aimed to evaluate reproducibility of quality of care indicators, by comparing reliability and agreement of diabetes-related processes and outcomes of care, as independently reported by patients and their physicians.

Methods

Study design and population

Within the development and implementation of a regional diabetes management program, a community-based cross-sectional survey was conducted in the fall of 2011 in the canton of Vaud, a Swiss state of more than 700,000 inhabitants.[12] Briefly, a random sample of registered community pharmacies were contacted and asked to recruit consecutive adults visiting the pharmacy with a diabetes-related prescription.[11] Non-institutionalized adults with type 1 or 2 diabetes for at least 12 months were invited to participate when visiting the pharmacy with a prescription for insulin, oral hypoglycaemic drugs, glycaemic strips, or a glucose meter. Women with current gestational diabetes, patients not fluent enough in French to fill in the questionnaire or unable to give informed consent were excluded. All participating patients gave written informed consent for their involvement in the study and for contacting their treating physicians. This study was approved by the state Institutional Review Board.

Collected data

Patients filled in a paper questionnaire with the following self-reported data: age, gender, education (primary, secondary, high school/university), civil status, smoking status (current, former, non smoker), characteristics of diabetes (type, treatment, duration, complications), and comorbid conditions (heart failure, valvulopathy, hypertension, hyperlipidaemia, chronic respiratory conditions, peptic ulcer, osteoporosis, osteoarthritis, Parkinson disease, malignancy, depression, other chronic condition). Recommended process of care measures[6] reported during the past 12 months were also collected: HbA1c (possible answers: not done, done once, done several times, unknown whether test was performed or not), lipid profile

(not done, done, unknown whether test was performed or not), urine analysis for microalbuminuria (not done, done, unknown whether test was performed or not), foot examination (not done, done, unknown whether it was performed or not), eye examination by ophthalmologist (done in past 12 months, in past 12-24 months, more than 24 months before, never done, unknown whether it was performed or not), blood pressure (not done, done, unknown whether it was performed or not), weight measurement (not done, done, unknown whether it was performed or not) and influenza immunization (not done, done, unknown whether it was performed or not). Patients were also asked to report, if known, the last measured value of HbA1c, height, weight, systolic and diastolic blood pressure.

We then contacted treating physicians (i.e. the physician taking care of patient's health, as reported by the patient) to fill in paper report cards. In addition to some personal characteristics (physician's age, gender, rate of part-time work, board-certified specialty, practice location and type), physicians reported the following data from patients medical charts: dates and last measured values of HbA1c, height, weight, and blood pressure, as well as the dates of last foot and eye examination, urine analysis for microalbuminuria, lipid profile, and influenza immunization.

Patient and physician reported process of care measures were dichotomized into done or not done during the past 12 months. "Unknown whether test was done or not" responses, which represented generally less than a few percents of the answers, were recoded as missing. For continuous outcomes of care, both patients and physicians were asked to report the respective last values, emphasizing the referral to the last encounter or measure.

Statistical analyses

We first performed simple descriptive statistics of collected variables and then assessed reproducibility between patients' and physicians' reports. Following de Mast[13] and Brennan,[14] we used Cohen's kappa[15] and uniform kappa[13] as reliability and agreement parameters, respectively, for dichotomous process indicators. As sample sizes were small and the marginal distributions of the dichotomous variables rather asymmetric, a Bayesian version of Cohen's kappa was also computed[16] since it appears to be more appropriate in such situations. For continuous outcome results, we used Lin's concordance correlation coefficient[17] as a measure of reliability and Bland & Altman's 95% limits of agreement[18] to assess measurement error (agreement). Because limits of agreement need to be clinically interpreted, deciding on whether or not agreement is good must take into account the clinical context. For comparison purposes with other studies, we also computed raw agreement, defined as the raw or unadjusted percentage of agreement in a 2x2 table. Finally, we conducted exploratory stratified analyses to assess whether reproducibility was associated with patient's age (<65 vs. ≥65 years), gender, diabetes type (type 1 vs. type 2 and undetermined type of diabetes), comorbid conditions (none, one, >1), education (primary, secondary, high school/university), or physician's specialty (endocrinologist vs. other) and clinical setting (solo private practice vs. other). We assessed whether there was a statistically significant difference between the strata reliability and agreement parameters by determining whether confidence intervals overlapped. All analyses were performed using Stata 12 (StataCorp, College Station, TX) and WinBUGS (MRC Biostatistics Unit, Cambridge, UK) for the Bayesian approach.

Results

Eight hundred nine of the 1013 consecutive eligible patients visiting pharmacies agreed to participate, of which 406 (50%) returned a filled questionnaire. Out of 186 contacted physicians whose contact details were given by patients, 111 (60%) sent back report cards for a total of 210 patients. Mean age of the 210 patients for whom we had both patients and physicians data, was 65.4 years, and 40% were women (Table 1). Type 1 and 2 diabetes was reported by 13.8% and 66.7% respectively, while diabetes type remained undetermined for nearly 20% of the patients. Half of patients had diabetes for >10 years and half were treated without insulin. Treating physicians were aged 53.5 years on average, and 18% were women (Table 1). About three quarters were general internists and 16% endocrinologists; a majority worked in solo private practices.

Process and outcome of care indicators, as reported independently by patients and physicians, were almost identical, except for foot examination and influenza immunization (Table 2). For these latter two processes of care, physicians' reported data showed more missing values and appeared slightly better than when considering patients' reported data.

Comparisons of diabetes-related process indicators in the past 12 months, as reported by patients and physicians, yielded contrasting results (Table 3). Agreement was good for routine processes of care like measurement of blood pressure (uniform kappa [κ_U]=0.941), HbA1c (κ_U=0.932), weight (κ_U=0.882), and lipid profile (κ_U=0.780), but it was less satisfactory for procedures such as microalbuminuria, and foot and eye examinations (all κ_U<0.500). Cohen's kappa values were generally very low (all <0.25), as expected since cells in the 2x2 tables were not homogenously filled, thereby implying low reliability (i.e. difficulty to

discriminate). Bayesian estimates of Cohen's kappa differed substantially from the standard Cohen's kappas.

Means of continuous outcomes were similar regardless of whether they were reported by patients or physicians (Table 4). Reproducibility results showed high concordance correlation coefficients (ρ_c) for height and weight (ρ_c=0.933 and 0.970, respectively) despite large limits of agreement (about ±6 cm and 8-10 kg, respectively), suggesting moderate agreement overall. Concordance correlation coefficients were substantially smaller for HbA1c (ρ_c=0.716) and for blood pressure (ρ_c=0.5-0.6) with large limits of agreement [±2% for HbA1c and ±25 mmHg for systolic blood pressure].

Stratified analyses did not show any statistically significant differences in reproducibility measures across strata. Nevertheless, we found that uniform kappa values were more than 10% better in women for HbA1c, lipid profile, and weight controls, illustrating a tendency towards better agreement performance for women.

Discussion

The results of this study showed good agreement between patient and physician reports of blood pressure, HbA1c, weight and lipid profile measurements within the past 12 months. Agreement was however lower for reports of microalbuminuria, foot and eye examination. When comparing continuous outcomes from both sources, reliability (i.e. concordance correlation coefficients) was very good for height and weight, but moderate for HbA1c and blood pressure. Despite these high concordance correlation coefficients, limits of agreement remained large for all reported measurements, illustrating that agreement between patients and physicians was rather moderate.

Overall, reproducibility of diabetes-related process indicators was consistent with those of previous studies. Among the 8409 participants of the TRIAD study, Beckles *et al.* found only fair reliability between self-reports and medical records (Cohen's kappa 0.25), with twice more self-reported annual examination than found in medical records.[10] Fowles *et al.* found high sensitivity in self-reported eye examination (89%) and HbA1c (99%) with low specificity (65% and 28%, resp.).[8] Comparing combined medical and administrative diabetes indicators with 440 patient self-reports, they also demonstrated more over-reporting than under-reporting by patients, and low Cohen's kappa values (0.371 and 0.678, resp.).[8] In another study, reports were higher than actual medical record data for annual eye examination, but not for annual foot examination and blood pressure.[7] Other studies described high sensitivity, low specificity and low Cohen's kappa of self-reported retinal examination compared with administrative data,[9] and low reproducibility between laboratory results.[19]

The concepts of reliability, agreement and reproducibility are often confused and used interchangeably, despite important conceptual differences.[20] In this study, we used

reproducibility as an umbrella concept for reliability and agreement, referring to the reproducibility of a measurement in a test-retest situation, as suggested by de Vet *et al*.[20] While reliability relates to discrimination, i.e. how well can patients be classified into mutually exclusive categories (e.g., by two observers), agreement addresses measurement errors and assesses how close or similar repeated reports or measurements are.[21]

The concept of reproducibility, which comprises both reliability and agreement, has been the focus of the following two debated issues. First, the often reported paradoxical behaviour of Cohen's kappa (i.e. low values of Cohen kappa despite high raw agreement)[13,20,21], which completely disappears when Cohen's kappa is considered as a measure of reliability instead of a measure of agreement. In fact, uniform kappa exhibits a much more coherent and expected behaviour when used in place of Cohen's kappa as a measure of agreement. Second, the fact that limits of agreement may be large from a clinical perspective even with high concordance correlation coefficients (e.g. 0.95). Indeed, since the concordance correlation coefficient is very sensitive to the range of measurements, only very large concordance correlation coefficients (e.g. 0.999) correspond to narrow limits of agreement; it should therefore be interpreted as a reliability parameter rather than a measure of agreement, even if its name explicitly relates to concordance.

When assessing reproducibility, either high agreement or high reliability may be preferable, depending on the aim of the project. High reliability is preferred to discriminate subjects, because this does not preclude identifying patients with abnormal values even when there are possibly large measurement errors and wide limits of agreement. To establish whether diagnoses, scores or judgments are identical, measurement error is of utmost importance

and discrimination of subjects is no longer relevant. Good measures of agreement are therefore targeted.

In our study results, the apparent discrepancy between agreement and reliability measures may be explained by various reasons. First, the generally low reliability but good agreement of process indicators is an expected consequence of the usually large asymmetry of the marginal distributions of the observations in the cells of 2x2 tables, and this is known to greatly influence Cohen's kappa.[13,20] For some indicators (e.g. HbA1c or blood pressure), agreement was substantial, suggesting that patient report may be quite accurate. Second, the very high reliability of some outcome results demonstrates that patients and physicians reports matched well when it came to magnitude and that these results were good enough for identifying subgroups of patients with abnormal values. However, under the magnifying lens of the limits of agreement, reports were inaccurate and discrepant, and limits of agreement too large to be useful for any clinical decision regarding treatment or patient monitoring. This was typically the case of height and weight reports.

We found good agreement for tests routinely done and easy to understand by patients, such as blood pressure, HbA1c, weight and lipid profile. This contrasted with the low agreement found for microalbuminuria, foot and eye examination. It may be explained by patient and physician factors. Indeed, patient self-report has been shown to be subject to recall bias (patients may forget if they had a specific test), telescoping (remembering an event as more recent than it actually was, e.g. a retinal examination performed in the past 2 years instead of 1 year)[9] and social acceptance bias (expectation of a more desirable answer).[8] Moreover, patients with diabetes may lack specific information or knowledge to recognize which test was performed. This may be especially true for those tests presenting lower agreement

values since several aspects can be assessed while analysing urine, or examining the eyes and feet, without having physicians telling all details to patients. Possible physician factors are that reports in medical records might be subject to bias, such as underreporting. This was shown in a study comparing process-based scores using standardized patients, clinical vignettes, and medical charts that showed that medical chart's scores were 5% lower than those using clinical vignettes, and 10% lower than those using standardized patients.[22] Another explanation includes fragmentation of care and the communication gap between healthcare providers that might lead to poor reproducibility of reports when contacting only one physician per patient.

As opposed to several published studies, we made a clear distinction between agreement and reliability parameters and used appropriate measures of agreement. Moreover, we considered several process and outcome of care indicators. Nevertheless, our study bears some limitations. First, our sample might be different from the general population of patients with diabetes because we targeted only those who visited pharmacies with a diabetes-related prescription and disease duration of at least 12 months. Also, patients not fluent enough in French to fill in a questionnaire, were excluded from the study. However, the proportion of adults with known and untreated diabetes should nevertheless remain small since current guidelines recommend early drug treatment; in addition, characteristics of included patients did not differ much from those of other similar studies.[8,23] Also, participating physicians may present different characteristics from Swiss physicians. Compared to Swiss physicians active in ambulatory care settings in 2011, participating physicians' characteristics were similar except for a lower proportion of female physicians in our sample (33% vs. 18%, respectively).[24] Second, data from both patients and physicians were not available for all possible pairs of respondents. In fact, data was mainly missing

when physicians did not report some information or did not send back the report cards (Table 2); whether non-response to a question represents absence of performance is only speculative. Also, physicians whose contact details were given by patients may not have been the one who performed the last clinical or laboratory test or described process of care. Therefore, patients and physicians may not always have referred to the same clinical encounter even if the word "last" was specified; some may have been tempted to game the survey. However, since both patients and physicians reported data independently, all these limitations should result in non-differential misclassification and, at worst, to lower reliability and agreement results. In addition, characteristics of patients with and without physicians' data were similar in terms of age, gender, type, duration and treatment of diabetes, BMI, smoking status, civil status, education and self-reported health. Third, we considered a 12 month cut-off for process indicators, while some guidelines recommend more than one test a year (such as for HbA1c, blood pressure, or lipid profile).[6] However, this cut-off was chosen because it is commonly used and would help in comparing results across studies.

Our findings may raise healthcare professionals' awareness of which self-reported measurements they may rely on. Indeed, they may rely with confidence on information relating to processes of care (for example whether a previous HbA1c, blood pressure, or lipid profile was performed in another clinical setting (good agreement)), but not on their absolute values since the latter showed lower reliability (i.e. concordance correlation coefficients) in our study.

In conclusion, quality of care evaluation should be made with caution when using self-report measures. For some processes of care, relying solely on patient self-reported values may be unsuitable for monitoring and decision making. Indeed, while they may be appropriately

used when patients are the only source of information (particularly for simple and understandable processes of care), it may be less appropriate when the reports concern processes of care whose purpose patients have more difficulties understanding. Further efforts should target improvement of survey methods (designing and clarifying self-reported questions for example), development of communication skills to help healthcare providers better communicate with each other and their patients, and better access to diabetes education services as well as broader adoption of the use of electronic medical records.

Acknowledgments

We would like to thank all pharmacies, physicians and patients who participated in the study. We also thank Mrs A. Casillas (Institute of Social and Preventive Medicine, Lausanne University Hospital, Switzerland) for copy-editing a previous version of the manuscript. None of the authors has any conflict of interest to declare.

T.H.C. and I.P.B. interpreted the results and wrote the manuscript. P.T. performed statistical analyses, interpreted the results, participated to the writing of the manuscript and provided insightful comments for the discussion. J.B. and I.P.B. designed the survey and collected the data. B.B. contributed to the interpretation of the results and reviewed the manuscript. I.P.B. is the principal investigator of the study and the guarantor of the content of the article.

An oral presentation of the study was given at the Swiss Society of General Internal Medicine Annual Meeting 2013 in Basel, Switzerland, under the title “Comparison of diabetes quality of care indicators as reported by patients and primary care physicians”.

Funding

This work was supported by the Department of Public Health of the canton of Vaud, Switzerland (“Programme cantonal Diabète”), and by the Swiss National Science Foundation. Dr T-H Collet’s research is supported by a grant from the Swiss National Science Foundation (PBLAP3-145870). Dr. I. Peytremann-Bridevaux is supported by a grant from the Swiss National Science Foundation [PROSPER N° 32333B-123817 and N° 32333B-139789].

Conflicts of interest

None declared.

Key points

- Patient and physician agreement was good for routine processes of care within the past 12 months such as blood pressure, HbA1c, weight, and lipid profile measurements, and less so for microalbuminuria, foot and eye examinations.
- Reproducibility, agreement, and reliability are terms often confused and used interchangeably, despite conceptual differences. We used reproducibility as an umbrella concept for reliability (the ability to discriminate, classify individuals into mutually exclusive categories, for example by two observers) and agreement (the assessment of how close or similar repeated reports or measurements are).
- Quality of care evaluation should be made with caution when using self-report measures only, especially for processes of care whose purpose patients have more difficulties understanding.

References

1. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? *Eff. Clin. Pract.* Aug-Sep 1998;1(1):2-4.
2. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N. Engl. J. Med.* Jun 26 2003;348(26):2635-2645.
3. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet.* Jun 16 2012;379(9833):2252-2261.
4. Schoen C, Osborn R, How SK, Doty MM, Peugh J. In chronic condition: experiences of patients with complex health care needs, in eight countries, 2008. *Health Aff. (Millwood).* Jan-Feb 2009;28(1):w1-16.
5. WHO. Global Health Observatory: Deaths from CVD and diabetes. *Global Health Observatory website* 2012; http://www.who.int/gho/ncd/mortality_morbidity/cvd_text/en/index.html. Accessed Dec 01, 2013.
6. ADA. Standards of medical care in diabetes-2013. *Diabetes Care.* Jan 2013;36 Suppl 1:S11-66.
7. Harwell TS, Moore K, Madison M, et al. Comparing self-reported measures of diabetes care with similar measures from a chart audit in a well-defined population. *Am. J. Med. Qual.* Jan-Feb 2001;16(1):3-8.
8. Fowles JB, Rosheim K, Fowler EJ, Craft C, Arrichiello L. The validity of self-reported diabetes quality of care measures. *Int. J. Qual. Health Care.* Oct 1999;11(5):407-412.
9. Thompson BL, O'Connor P, Boyle R, et al. Measuring clinical performance: comparison and validity of telephone survey and administrative data. *Health Serv. Res.* Aug 2001;36(4):813-825.
10. Beckles GL, Williamson DF, Brown AF, et al. Agreement between self-reports and medical records was only fair in a cross-sectional study of performance of annual eye examinations among adults with diabetes in managed care. *Med. Care.* Sep 2007;45(9):876-883.
11. Peytremann-Bridevaux I, Bordet J, Santschi V, Collet TH, Eggli M, Burnand B. Community-based pharmacies: an opportunity to recruit patients? *International journal of public health.* Apr 2013;58(2):319-322.
12. Hagon-Traub I, Hirsiger P, Bovet P, et al. Diabetes state program - Presentation - Technical report [Programme cantonal Diabète - Présentation du programme - Rapport technique]. 2010; http://www.vd.ch/fileadmin/user_upload/themes/sante_social/services_soins/Diabete/fichiers_pdf/110927_Vdiab_Rapport_technique_RF_10_5_7_2010_VC.pdf. Accessed Dec 01, 2013.
13. de Mast J. Agreement and Kappa-Type Indices. *Am Stat.* May 2007;61(2):148-153.
14. Brennan RL, Prediger DJ. Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educ Psychol Meas.* October 1 1981;41(3):687-699.
15. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* April 1 1960;20(1):37-46.
16. Basu S, Banerjee M, Sen A. Bayesian inference for kappa from single and multiple studies. *Biometrics.* Jun 2000;56(2):577-582.

17. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* Mar 1989;45(1):255-268.
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* Feb 8 1986;1(8476):307-310.
19. Harwell TS, Dettori N, McDowall JM, et al. Do persons with diabetes know their (A1C) number? *Diabetes Educ.* Jan-Feb 2002;28(1):99-105.
20. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J. Clin. Epidemiol.* Oct 2006;59(10):1033-1039.
21. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J. Clin. Epidemiol.* Jan 2011;64(1):96-106.
22. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA.* Apr 5 2000;283(13):1715-1722.
23. Firmann M, Mayor V, Vidal PM, et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord.* 2008;8:6.
24. Kraft E, Hostettler S. Medical statistics in 2012 - 31,858 physicians provide healthcare in Switzerland [Statistique médicale 2012 - 31858 médecins garantissent les soins médicaux en Suisse]. *Bulletin des médecins suisses.* 2013;94(12):453-457.
25. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med. Care.* Mar 1996;34(3):220-233.

Tables

Table 1. Patient and physician characteristics

Patients self-reported characteristics (n = 210)	
Age, years, mean (SD)	65.4 (10.5)
Women, no (%)	84 (40.0%)
Body Mass Index, mean (SD) (n = 195)	30.2 (6.1)
Education, no (%) (n = 202)	
Primary school	36 (17.8%)
Secondary school	124 (61.4%)
High school, university	42 (20.8%)
Civil status, no (%) (n = 209)	
Single	18 (8.6%)
Married	129 (61.7%)
Divorced, separated	37 (17.7%)
Widower	25 (12.0%)
Smoking status, no (%) (n = 206)	
Current	39 (18.9%)
Former	88 (42.7%)
Never	79 (38.3%)
Diabetes type, no (%)	
Type 1	29 (13.8%)
Type 2	140 (66.7%)
Other [a]	41 (19.5%)
Diabetes duration, no (%) (n = 208) [b]	
1-10 years	101 (48.6%)
11-20 years	67 (32.2%)
>20 years	40 (19.2%)

Diabetes treatment, no (%) (n = 209) [c]	
Oral hypoglycaemic drugs	107 (51.2%)
Insulin	46 (22.0%)
Oral hypoglycaemic drugs + insulin	56 (26.8%)
Diabetes complications, no (%) (n = 202) [d]	
Macrovascular complications	76 (37.6%)
Microvascular complications	52 (25.7%)
Subjective health, no (%) (n = 205) [e]	
Excellent	4 (2.0%)
Very good	31 (15.1%)
Good	128 (62.4%)
Fair	34 (16.6%)
Poor	8 (3.9%)
Physicians self-reported characteristics (n = 110)	
Age, years, mean (SD)	53.5 (8.8)
Women, no (%)	20 (18.2%)
Activity level ≥ 80% part time or full time, no (%) (n = 106)	76 (71.7%)
Specialty, no (%) (n = 106) [f]	
General or internal medicine board certified	79 (74.5%)
Endocrinology and diabetes board certified	17 (16.0%)
Other [g]	13 (12.3%)
Clinical practice location, no (%) (n = 106)	
Rural	21 (19.8%)
Urban	61 (57.5%)
Mixed	24 (22.6%)
Clinical setting, no (%) (n = 107) [f]	
Solo private practice	60 (56.1%)
Group private practice	40 (37.4%)

Ambulatory care centre	4 (3.7%)
Public hospital	8 (7.5%)

[a] Other diabetes type (0.5%), did not know (12.4%), or no answer (6.7%)

[b] Patients diagnosed diabetes less than one year before the survey were excluded

[c] Patients were eligible if they showed a prescription with oral hypoglycaemic drugs, insulin, glucose meter or strips

[d] Reported macrovascular diabetes complications were angina/myocardial infarction, stroke, and peripheral neuropathy (foot pain/sensation loss, foot ulcers, amputation). Reported microvascular diabetes complications were retinopathy and nephropathy (dialysis or kidney transplant).

[e] Subjective health was assessed by the first question of the 12-item Short Form Health Survey:[25] "In general, would you say your health is excellent, very good, good, fair, or poor?"

[f] More than one answer possible

[g] Either non board certified physicians, or other specialties

Table 2. Process and outcomes of care as reported by patients and physicians

	Patients		Physicians	
Process of care results (n = 210)	N	%	N	%
HbA1c [a]	120	98%	207	97%
Lipid profile	203	97%	189	90%
Microalbuminuria	179	74%	132	75%
Foot examination	201	71%	144	92%
Eye examination [b]	201	60%	126	64%
Blood pressure	207	99%	208	99%
Weight	205	95%	191	95%
Influenza immunization	208	67%	86	93%
Mean outcomes of care (n = 210)	N	mean	N	mean
HbA1c, % [a]	78	7.4	118	7.4
Height, cm	195	171	174	170.2
Weight, kg	207	88.1	196	89.1
Systolic blood pressure, mmHg	108	133.5	209	133.9

[a] Among those who knew what glycosylated haemoglobin (HbA1c) was

[b] By an ophthalmologist

Table 3. Comparison of process indicators during the past 12 months, as reported by patients and physicians

	Reported by patient	Reported by physician			Agreement		Reliability	
		Not done	Done	N	Raw agreement	Uniform kappa (95% CI) [a]	Cohen's kappa	Bayesian kappa, median (95% CI)
Glycosylated haemoglobin (HbA1c)	Not done	0	2	117	0.97	0.93 (0.86, 0.98)	-0.02 (-0.05, 0.01)	0.15 (-0.02, 0.58)
	Done	2	113					
Lipid profile	Not done	0	5	182	0.89	0.78 (0.69, 0.86)	-0.04 (-0.08, -0.01)	0.01 (-0.05, 0.20)
	Done	15	162					
Microalbuminuria	Not done	7	10	112	0.69	0.39 (0.22, 0.56)	0.12 (-0.07, 0.31)	0.118 (-0.05, 0.31)
	Done	24	71					
Foot examination	Not done	0	29	136	0.72	0.44 (0.30, 0.59)	-0.11 (-0.18, -0.04)	-0.08 (-0.16, 0.04)
	Done	9	98					
Eye examination [b]	Not done	19	15	123	0.68	0.35 (0.18, 0.50)	0.26 (0.08, 0.43)	0.25 (0.08, 0.42)
	Done	25	64					
Blood pressure	Not done	0	3	205	0.97	0.941 (0.89, 0.98)	-0.02 (-0.04, -0.01)	0.12 (-0.01, 0.50)
	Done	3	199					
Weight	Not done	2	3	187	0.94	0.88	0.24	0.24

	Done	8	174			(0.81, 0.95)	(-0.06, 0.54)	(0.04, 0.53)
Influenza immunization	Not done	0	8	86	0.84	0.67	-0.09	-0.01
	Done	6	72			(0.50, 0.83)	(-0.16, -0.01)	(-0.11, 0.25)

[a] Confidence intervals for uniform kappas were computed using the studentized bootstrap method.

[b] By an ophthalmologist

Table 4. Comparison of continuous clinical and laboratory results, as reported by patients and physicians

	N both reported by patient and physician	As reported by patient, mean (SD)	As reported by physician, mean (SD)	Concordance correlation coefficient (ρ_c) [17]	95% Limits of agreement [18]
Glycosylated haemoglobin (HbA1c), %	82	7.33 (1.32)	7.42 (1.31)	0.716	(-2.03, 1.84)
Height, cm	160	170.7 (9.2)	170.4 (9.2)	0.933	(-6.3, 6.9)
Weight, kg	194	88.45 (19.78)	89.88 (20.90)	0.970	(-10.8, 8.0)
Systolic blood pressure, mmHg	108	133.5 (13.2)	133.7 (16.2)	0.627	(-25.3, 24.8)