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

Tinh-Hai Collet¹,², Patrick Taffe³, Julie Bordet³, Bernard Burnand³, Isabelle Peytremann-Bridevaux³

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

Correspondence: Isabelle Peytremann-Bridevaux, 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, e-mail: Isabelle.Peytremann-Bridevaux@chu.ch

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 (±6 cm) and weight (8-10 kg) 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.

Introduction

Chronic disease care requires good collaboration between health care 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,⁷ 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, as it is an easier means 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.²−¹⁰ 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 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 >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 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), 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 patient 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 percent 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 and Brennan, we used Cohen’s kappa15 and uniform kappa13 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, as it appears to be more appropriate in such situations. For continuous outcome results, we used Lin’s concordance correlation coefficient17 as a measure of reliability and Bland & Altman’s 95% limits of agreement18 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 $2 \times 2$ 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 patient and physician 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 ($k_U$) = 0.941], HbA1c ($k_U = 0.932$), weight ($k_U = 0.882$) and lipid profile ($k_U = 0.780$), but it was less satisfactory for procedures such as microalbuminuria, and foot and eye examinations (all $k_U < 0.500$). Cohen’s kappa values were generally very low (all $<0.25$), as expected, as cells in the $2 \times 2$ 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 ($\rho_U$) for height and weight ($\rho_U = 0.953$ 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 ($\rho_U = 0.716$) and for blood pressure ($\rho_U = 0.5–0.6$), 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 >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.
Clinical practice location, no (%) (n = 106)  
- Rural 21 (19.8%)  
- Urban 61 (57.5%)  
- Mixed 24 (22.6%)  

Clinical setting, no (%) (n = 107)  
- Solo private practice 60 (56.1%)  
- Group private practice 40 (37.4%)  
- Ambulatory care centre 4 (3.7%)  
- Public hospital 8 (7.5%)  

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)  
  - General or internal medicine board certified 79 (74.5%)  
  - Endocrinology and diabetes board certified 17 (16.0%)  
  - Other 13 (12.3%)  
- Clinical practice location, no (%) (n = 106)  
  - Rural 21 (19.8%)  
  - Urban 61 (57.5%)  
  - Mixed 24 (22.6%)  

Physicians reported diabetes less than one year before the survey were excluded.

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. 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

<table>
<thead>
<tr>
<th>Process of care results (n = 210)</th>
<th>Patients</th>
<th>Physicians</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>HbA1c*</td>
<td>120</td>
<td>98%</td>
</tr>
<tr>
<td>Lipid profile</td>
<td>203</td>
<td>97%</td>
</tr>
<tr>
<td>Microalbuminuria</td>
<td>179</td>
<td>74%</td>
</tr>
<tr>
<td>Foot examination</td>
<td>201</td>
<td>71%</td>
</tr>
<tr>
<td>Eye examination²</td>
<td>201</td>
<td>60%</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>207</td>
<td>99%</td>
</tr>
<tr>
<td>Weight</td>
<td>205</td>
<td>95%</td>
</tr>
<tr>
<td>Influenza immunization</td>
<td>208</td>
<td>67%</td>
</tr>
</tbody>
</table>

Mean outcomes of care (n = 210)

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>N</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c, %*</td>
<td>78</td>
<td>7.4</td>
<td>118</td>
</tr>
<tr>
<td>Height, cm</td>
<td>195</td>
<td>171</td>
<td>174</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>207</td>
<td>88.1</td>
<td>196</td>
</tr>
<tr>
<td>Systolic blood pressure, mmHg</td>
<td>108</td>
<td>133.5</td>
<td>209</td>
</tr>
</tbody>
</table>

a: Among those who knew what glycosylated haemoglobin (HbA1c) was.

b: By an ophthalmologist.

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)¹³,¹⁴, 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 values (0.371 and 0.678, resp.).²¹ 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.).²² In another study, reports were higher than actual medical record data for annual eye examination, but not for annual foot examination and blood pressure. Other studies described high sensitivity, low specificity and low Cohen’s kappa of self-reported retinal examination compared with administrative data,²³ and low reproducibility between laboratory results.²⁴

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)¹³,¹⁴, 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 values (0.371 and 0.678, resp.).²¹ 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.).²² In another study, reports were higher than actual medical record data for annual eye examination, but not for annual foot examination and blood pressure. Other studies described high sensitivity, low specificity and low Cohen’s kappa of self-reported retinal examination compared with administrative data,²³ and low reproducibility between laboratory results.²⁴

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 $2 \times 2$ tables, and this is known to greatly influence Cohen’s kappa.¹³,²¹ For some indicators (e.g. HbA1c or blood pressure), agreement was substantial, suggesting that patient report may be quite accurate. Second, the high reliability of some outcome results demonstrates that patient and physician 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, as 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 under-reporting. 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.²³ Another explanation includes fragmentation of care and the communication gap between health care 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 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, as current guidelines recommend early drug treatment; in addition, characteristics of included patients did not differ much from those of other similar studies.⁸,²⁴ Also, participating physicians may present different characteristics from Swiss physicians. Compared with

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

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Reported by patient</th>
<th>Reported by physician</th>
<th>Agreement</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not done</td>
<td>Done</td>
<td>N</td>
<td>Raw agreement</td>
</tr>
<tr>
<td>Glycosylated haemoglobin (HbA1c)</td>
<td>Not done</td>
<td>0</td>
<td>2</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>113</td>
<td>0.89</td>
</tr>
<tr>
<td>Lipid profile</td>
<td>Not done</td>
<td>0</td>
<td>5</td>
<td>182</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>71</td>
<td>182</td>
</tr>
<tr>
<td>Microalbuminuria</td>
<td>Not done</td>
<td>7</td>
<td>10</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>71</td>
<td>112</td>
</tr>
<tr>
<td>Foot examination</td>
<td>Not done</td>
<td>0</td>
<td>29</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>98</td>
<td>136</td>
</tr>
<tr>
<td>Eye examination¹ᵇ</td>
<td>Not done</td>
<td>19</td>
<td>15</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>64</td>
<td>123</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Not done</td>
<td>0</td>
<td>3</td>
<td>205</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>199</td>
<td>205</td>
</tr>
<tr>
<td>Weight</td>
<td>Not done</td>
<td>2</td>
<td>3</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>174</td>
<td>187</td>
</tr>
<tr>
<td>Influenza immunization</td>
<td>Not done</td>
<td>0</td>
<td>8</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Done</td>
<td></td>
<td>72</td>
<td>86</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Indicator</th>
<th>N both reported by patient and physician</th>
<th>As reported by patient, mean (SD)</th>
<th>As reported by physician, mean (SD)</th>
<th>Concordance correlation coefficient ($\rho$)¹⁷</th>
<th>95% limits of agreement¹⁸</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycosylated haemoglobin (HbA1c), %</td>
<td>82</td>
<td>7.33 (1.32)</td>
<td>7.42 (1.31)</td>
<td>0.716</td>
<td>(-2.03, 1.84)</td>
</tr>
<tr>
<td>Height, cm</td>
<td>160</td>
<td>170.7 (9.2)</td>
<td>170.4 (9.2)</td>
<td>0.933</td>
<td>(-6.3, 6.9)</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>194</td>
<td>88.45 (19.78)</td>
<td>89.88 (20.90)</td>
<td>0.970</td>
<td>(-10.8, 8.0)</td>
</tr>
<tr>
<td>Systolic blood pressure, mmHg</td>
<td>108</td>
<td>133.5 (13.2)</td>
<td>133.7 (16.2)</td>
<td>0.627</td>
<td>(-25.3, 24.8)</td>
</tr>
</tbody>
</table>
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). Second, data from both patients and physicians were not available for all possible pairs of respondents. In fact, data were 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 ones 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, because 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). However, this cut-off was chosen because it is commonly used and would help in comparing results across studies.

Our findings may raise health care 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, as 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 health care 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.

Acknowledgements

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 (PB LAP3-145870). 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 and 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

Introduction

Vaccinations are an effective way to prevent childhood infectious diseases which may cause severe complications including lasting retardation and death. For some preventable diseases, the vaccination rate should warrant a population-wide protection by the herd effect, i.e., the pathogen should become unable to spread due to a too low proportion of susceptible individuals remaining.

In Germany, vaccination rates have been declining. In some areas, vaccination rates are already insufficient and endemic onsets of...