Validation and comparison of currently available stratification systems for patients with diabetes by risk of foot ulcer development

M Monteiro-Soares¹-², A Vaz-Carneiro²-³-⁴, S Sampaio²-⁵ and M Dinis-Ribeiro²

¹Endocrinology, diabetes and metabolism department – Diabetic Foot Clinic, Centro Hospitalar de Vila Nova de Gaia/Espinho EPE, Vila Nova de Gaia, Portugal, ²CIDES/CINTESIS – Department of Health Information and Decision Sciences, Oporto Faculty of Medicine, Oporto, Portugal, ³CEMBE – Centre for Evidence Based Medicine, University of Lisbon, Portugal, ⁴Cochrane Coordinating Centre Portugal, Lisbon, Portugal and ⁵Vascular Surgery Department, Oporto Faculty of Medicine, Oporto, Portugal

(Correspondence should be addressed to M Monteiro-Soares at CIDES/CINTESIS; Email: mat.monteirosoares@gmail.com)

Abstract

Aims/hypothesis: There are five systems to stratify the risk for the development of a diabetic foot ulcer (DFU). This study aimed to prospectively validate all of them in the same cohort of participants to allow their direct comparison.

Methods: A retrospective cohort study was conducted on all patients with diabetes but without an active DFU attending our podiatry section (n=364) from January 2008 to December 2010. Participants’ characteristics and all variables composing the stratification systems were assessed at baseline. Follow-up was performed for 1 year or until DFU occurred.

Results: Participants had a mean age of 64 years; 99.7% had type 2 diabetes and 48.6% were male. Median follow-up was 12 months (1–12) during which 33 subjects (9.1%) developed a DFU. Age, diabetes duration, foot deformity, peripheral vascular disease, diabetic peripheral neuropathy, previous DFU, and previous lower extremity amputation were associated with DFU occurrence. All systems presented greater DFU occurrence frequency as the risk group was higher (c², P<0.001) and showed good diagnostic accuracy values, especially negative predictive value (≥95%) and area under the receiver operating curve (≥0.73). The lowest performance concerned positive predictive value (≤29.5%).

Conclusions/interpretation: All the currently available stratification systems show high accuracy to detect which patients will develop a DFU with no significant differences among them. Therefore, for diabetic foot screening and resource allocation, it would be desirable to have a single unified system, combining the available systems, prospectively validated in a multicenter context and testing the inclusion of novel predictive variables’ pertinence.

European Journal of Endocrinology 167 401–407

Introduction

Diabetes-related foot complications, namely diabetic foot ulcers (DFUs) and lower extremity amputation (LEA), are very prevalent worldwide (1). Therefore, they have a great impact on health economy and available resource allocation, as well as on patients’ quality of life. Additionally, we are observing a constant and pronounced increase in diabetes prevalence, which represents a number of patients exceedingly superior to the available resources (2). In fact, economical cuts in preventive diabetic care are being proposed and implemented internationally. More than ever, an appropriate stratification of patients by their risk of developing a DFU is crucial for resource allocation, as well as prevention of complications (3, 4).

Although five stratification systems were developed, namely the University of Texas (UT) (5), American Diabetes Association (ADA) (6), International Working Group on the Diabetic Foot (IWGDF) (7), Scottish Intercollegiate Grouping Network (SIGN) systems (8), and the Seattle risk score (9), to date no system has been collectively adopted (10) and their use in clinical practice is still scarce (8).

A systematic review (4), performed to retrieve all the available stratification systems created and their validation studies, showed that: i) although their core variables are very similar, the procedures for selection of variables and risk group stratification varies considerably; ii) some were never externally validated; iii) their prognostic accuracy was not reported; and iv) they were never validated simultaneously in the same cohort.
Therefore, in that systematic review, the authors could not choose which system to apply in daily practice (4). Hence, we have conducted this retrospective cohort study in order to validate and compare all the available systems in terms of structure and validity at 1 year in the same cohort of patients.

**Materials and methods**

**Type of study and selection of participants**

A retrospective cohort study was conducted on all patients with diabetes attending the podiatry section of our Diabetic Foot Clinic, from January 2008 to December 2010. Patients were excluded if they had an active DFU, complete inability to walk, any data missing, and/or a follow-up period of less than 1 year.

This study was approved by the Ethics Committee of our institution and no adverse events occurred due to its conduction.

**Data collection**

At baseline (the first podiatric appointment), patients’ characterization variables and those included in all the available systems were assessed and registered by two podiatrists with extensive experience in diabetic foot care (more than 8 years). Variables were collected from the patients’ clinical files until the end of December 2011, and all the systems were applied at that moment. Consequently, investigators were blinded to the systems’ stratification during data assessment and collection.

The characterization variables are of age, gender, diabetes duration, type and treatment, previous myocardial infarction and/or stroke, hypertension, nephropathy, and retinopathy (and respective laser treatment). These were collected through a structured interview including the presence of claudication and DFU (5, 6, 7, 8, 9) and/or LEA (5, 6, 7, 8, 9) history.

HbA1c value (9) was collected through blood sample analysis, and we used the one closest to the first appointment (with < 3 months). Visual impairment (8, 9) and physical impairment (8) were assessed through questionnaires during interviews and subjective analysis and were defined as patients’ inability to see or reach their own feet (8) respectively.

Diabetic peripheral neuropathy (DPN) diagnosis varied according to the system used. For the UT system (5), we used the 128 Hz tuning fork sensitivity, at the distal phalanx of the hallux (instead of the biothesiometer); for the SIGN system (8) and Seattle risk score (9), the Semmes–Weinstein monofilament; for the ADA (6) and IWGDF (7) systems, an altered monofilament and/or tuning fork sensation was considered to evaluate the presence of DPN. Monofilament touch perception was tested at the pulp of the hallux, 1st, 3rd, and 5th metatarsal heads in each foot. Absent sensation at one or more sites was considered as the presence of DPN (11).

Regarding the peripheral vascular disease (PVD) diagnosis, although definition variation also occurred (6, 7, 8), it was assessed through direct pulse palpation. One non-palpable foot pulse was considered as PVD for the ADA and IWGDF systems, conversely for the SIGN system was the inability to feel both pulses in one foot. In the UT system and Seattle risk score, the PVD diagnosis is not included. The presence of foot deformity (5, 6, 7, 8), callus (8), edema, onychomycosis, and *tinea pedis* (9) was identified through foot examination.

Follow-up ended as the first DFU occurred in any foot or after 1 year. Participants were reevaluated in variable intervals (from 1 to 6 months), according to podiatrists’ clinical evaluation. However, patients were instructed to contact or return to our clinic if any complication developed before the next scheduled appointment. DFU development was defined as a full-thickness defect distal to the malleoli requiring more than 14 days to heal (9).

**Statistical analysis**

For the association between characterization and systems’ composing variables with DFU occurrence, we used Student’s t-test for continuous variables for independent samples (as all presented a normal distribution), and for categorical variables, we applied the χ² or Fisher’s exact test, when applicable. Significance was defined as $P<0.05$. In those cases presenting an association, a $P$ value between 0.05 and 0.1, odds ratio (OR), and respective 95% confidence intervals (95% CIs) were calculated in addition to the Cramer’s V statistic ($\phi_c$) for categorical variables – this index varies from 0 to 1, ranging from no association between variables to complete association respectively – and for continuous variables. Cohen’s d statistic ($d$) – in which a 0.2 value is interpreted as a small effect size, 0.5 as medium, and 0.8 as large. Participants were stratified according to each system categories. Sensitivity, specificity, likelihood ratios (LRs), predictive values, and area under the receiver operating curve (AUC) with 95% CI were calculated. All statistical analysis was performed using IBM SPSS version 19.0 (Chicago, IL, USA).

**Results**

**Description of participants**

In this study, 364 subjects were included. Median follow-up was 12 (range 1–12) months during which 33 participants (9.1%) developed a DFU. The mean age of the sample was 64 (19–94) years. 48.6% were male, 99.7% had type 2 diabetes, 41.5% used insulin, and the mean diabetes duration was 17 (1–52) years.
**Association of variables with DFU development**

Of the former variables, only age and diabetes duration were significantly associated with DFU occurrence (Table 1). The association between DFU risk stratification system variables and the outcome is also described in Table 1. Foot deformity, PVD diagnosis (through different definitions), DPN diagnosis (through different definitions), previous DFU, and previous LEA were highly associated with DFU development (\(P < 0.001\)). We stress that the HbA1c value (\(P = 0.10\); OR 1.21 (95% CI 0.91–1.60); \(\chi^2\) for association, \(P = 0.3\)), the presence of retinopathy (\(P = 0.05\), OR 2.07 (95% CI 1.00–4.28), \(\chi^2\) for association, \(P = 0.1\)) and visual impairment (\(P = 0.08\), OR 1.80 (95% CI 0.87–3.74), \(\chi^2\) for association, \(P = 0.08\)), callus (\(P = 0.09\), OR 1.84 (95% CI 0.90–3.78), \(\chi^2\) for association, \(P = 0.09\)), and claudication (\(P = 0.10\), OR 1.92 (95% CI 0.87–4.25), \(\chi^2\) for association, \(P = 0.09\)) presented a \(P\) value \(\leq 0.1\), which also represents a potential predictive value, but with small effect size.

For all the systems, DFU occurrence increased as the risk group got higher (\(\chi^2\) for association and trend, \(P < 0.001\); Table 2). For this analysis, some of the categories had to be grouped due to low expected values (\(< 5\)).

**Accuracy of the DFU risk stratification systems**

Analyzing Table 3, one observes that in the highest risk group, the UT system presented the lowest sensitivity and the SIGN system the lowest specificity and positive...
The UT system and the Seattle risk scores presented the highest specificity values.

When assembling the highest and high/medium risk groups, the SIGN system presented the lowest specificity and positive predictive values (PPV) and the ADA system the lowest positive LR. Once again, the UT system and the Seattle risk score had the highest specificity values. When assembling the highest and medium/low risk groups, the ADA system presented the lowest positive LR and the UT system the highest specificity value.

Regarding the systems’ diagnostic accuracy, the respective AUC values were 0.73 (95% CI 0.63–0.83) for the UT system, 0.83 (95% CI 0.79–0.88) for the ADA system, 0.86 (95% CI 0.81–0.91) for the IWGDF system, 0.75 (95% CI 0.68–0.82) for the SIGN system, and 0.82 (95% CI 0.75–0.89) for the Seattle risk score (Fig. 1). All classification systems presented high AUC values and no statistical differences were found between them.

Discussion

There are several risk stratification systems developed for the detection of diabetic patients at higher risk of DFU occurrence (4), and it has been evidenced that they are more sensitive than any individual predictive variable (2). This study was the first where all the DFU risk stratification systems were retrospectively validated in the same cohort. Until now, the ADA and UT were never validated and respective diagnostic accuracy measures reported, and the IWGDF and SIGN were never externally validated.

Additionally, for all the systems (except Seattle risk score), the AUC value was never reported. This measure is considered, for some authors, the best way to determine a system’s discriminatory ability (12).

Our data support the predictive value of the stratification systems’ main variables such as PVD, DPN, and previous foot complications. All the remaining composing variables of the systems, except for tinea pedis and physical impairment, presented a potential predictive value ($\leq 0.1$) but did not achieve statistical significance. One must highlight that, in the two studies assessing the predictive value of tinea pedis for DFU development, statistical significance was only observed in the multivariate analysis (3, 9). Of the collected variables not included in any of the studied systems, only older age and diabetes duration were associated with DFU development.

Also, our results suggest that all the available systems are equally and highly accurate. All systems presented AUC values higher than 0.73 and a trend was observed for increased DFU occurrence in higher risk groups.

We have observed that all the systems, using as cutoff any of the risk groups, presented a PPV value < 30%. This can be interpreted by classifying the subjects as
Table 3 | DFU risk stratification systems’ classification diagnostic accuracy measures.

<table>
<thead>
<tr>
<th>DAM</th>
<th>ADA system</th>
<th>IWGDF system</th>
<th>UT system</th>
<th>SIGN system</th>
<th>Seattle risk score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value (%)</td>
<td>95% CI</td>
<td>RG</td>
<td>Value (%)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Sens</td>
<td>3</td>
<td>90.9</td>
<td>81.1–100.0</td>
<td>3A + 3B</td>
<td>87.9</td>
</tr>
<tr>
<td>Spe</td>
<td>70.4</td>
<td>65.5–75.3</td>
<td>70.7</td>
<td>65.8–75.6</td>
<td>84.6</td>
</tr>
<tr>
<td>PPV</td>
<td>23.4</td>
<td>16.1–30.8</td>
<td>23.0</td>
<td>15.7–30.4</td>
<td>98.3</td>
</tr>
<tr>
<td>NPV</td>
<td>98.7</td>
<td>97.3–100.0</td>
<td>98.3</td>
<td>96.7–100.0</td>
<td>3.1</td>
</tr>
<tr>
<td>LR+</td>
<td>3.1</td>
<td>2.5–3.7</td>
<td>3.0</td>
<td>2.4–3.7</td>
<td>0.1</td>
</tr>
<tr>
<td>LR−</td>
<td>55.9</td>
<td>50.5–61.2</td>
<td>44.7</td>
<td>39.4–50.1</td>
<td>728</td>
</tr>
<tr>
<td>Spe</td>
<td>18.4</td>
<td>12.8–24.1</td>
<td>15.3</td>
<td>10.5–20.1</td>
<td>18.4</td>
</tr>
<tr>
<td>PPV</td>
<td>23.4</td>
<td>16.1–30.8</td>
<td>23.0</td>
<td>15.7–30.4</td>
<td>98.3</td>
</tr>
<tr>
<td>NPV</td>
<td>98.7</td>
<td>97.3–100.0</td>
<td>98.3</td>
<td>96.7–100.0</td>
<td>3.1</td>
</tr>
<tr>
<td>LR+</td>
<td>3.1</td>
<td>2.5–3.7</td>
<td>3.0</td>
<td>2.4–3.7</td>
<td>0.1</td>
</tr>
<tr>
<td>LR−</td>
<td>55.9</td>
<td>50.5–61.2</td>
<td>44.7</td>
<td>39.4–50.1</td>
<td>728</td>
</tr>
<tr>
<td>Sens</td>
<td>1 + 2 + 3</td>
<td>100.0</td>
<td>NC</td>
<td>1 + 2 + 3</td>
<td>100.0</td>
</tr>
<tr>
<td>Spe</td>
<td>13.0</td>
<td>9.0–16.6</td>
<td>38.4</td>
<td>33.1–43.6</td>
<td>65.9</td>
</tr>
<tr>
<td>PPV</td>
<td>10.3</td>
<td>7.0–13.6</td>
<td>13.9</td>
<td>9.5–18.3</td>
<td>17.5</td>
</tr>
<tr>
<td>NPV</td>
<td>100.0</td>
<td>NC</td>
<td>100.0</td>
<td>NC</td>
<td>96.0</td>
</tr>
<tr>
<td>LR+</td>
<td>1.1</td>
<td>1.1–1.2</td>
<td>1.6</td>
<td>1.5–1.8</td>
<td>0.4</td>
</tr>
<tr>
<td>LR−</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>0.4</td>
</tr>
</tbody>
</table>

DAM, diagnostic accuracy measure; Sens, sensitivity; Spe, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR−, negative likelihood ratio; ADA, American Diabetes Association; RG, risk groups; 95% CI, 95% confidence interval; NC, non-calculable; IWGDF, International Working Group on Diabetic Foot; UT, University of Texas; SIGN, Scottish Intercollegiate Guideline Network.
being at risk, and that more than 70% will not develop a DFU, which represents a high cost (especially for tertiary institutions). We believe that it is essential to develop strategies to improve this situation.

Conversely, for the highest risk group or combining the medium with the highest risk group, excellent negative predictive values were obtained in all the systems, and also in most of the systems’ sensitivity values. This means that almost all the patients developing a DFU are predicted by the systems. Therefore, we suggest that those in lower risk groups be followed in the primary care setting.

When comparing the systems with each other we observed that the UT system presented the highest specificity values (only similar to the Seattle risk score) but for the highest risk and the lowest sensitivity. The SIGN system presented the lowest specificity and positive LR in the highest risk group and in the medium plus high-risk group the lowest specificity and PPV.

Comparing the results of this study with the systems’ validation studies, we observed that the IWGDF presented lower specificity and PPV values in comparison with the study by Peters et al. (13). The SIGN system presented lower specificity and PPV in comparison with one study from Leese et al. (8) and also lower positive LR when compared with a different study by Leese et al. (2). The Seattle risk score was externally validated in the same setting (3), and therefore, no differences were found. Regarding the ADA and UT systems, that we are aware of, no diagnostic accuracy measures were ever reported.

The differences described may be justified by the following limitations. This study was conducted in a tertiary referral center and therefore presents a possible bias – a high prevalence of DFU occurrence (9.1%) when compared with other DFU risk stratification systems’ validation studies, such as the one performed by Leese et al. (8) in a community setting (5%). Conversely, all the remaining derivation or validation studies (2, 3, 5, 7, 9, 13) presented higher frequency (9.5–34%). Our study population was composed mainly of elderly subjects (mean age greater than 65 years) with type 2 diabetes (99.7%), which may affect our results’ generalizability.

For each variable’s prognostic value analyzed, one should include 10–15 subjects (14), which represents a required sample size of 280–450. We enrolled 364 participants, a number containing all the patients available during the study conduction period but not achieving the superior limit for an adequate sample size. Therefore, we have considered that those variables presenting a $P$ value less than or equal to 0.1 presented a potential predictive value. However, for all the variables in this condition, a small effect size was observed.

We believe that a DFU risk stratification system should be equally easy to apply in all sorts of settings and performed using only commonly available material. Therefore, we have chosen to perform the PVD diagnosis only through pulse palpation as proposed by Leese et al. (8). Additionally, we have decided, for the DPN diagnosis, to apply the tuning fork in spite of the biothesiometer as proposed by Peters et al. (13). Despite these modifications and using all the different definitions proposed, these two variables were highly associated with DFU development ($P<0.001$). This corroborates our systematic review results: DPN and PVD clinical collection methods seem relatively unimportant (4, 15). Although several authors have described the foot pulse palpation to have low sensitivity (16, 17), the SWM application procedure does not have consensus (18, 19) and the tuning fork has low reliability (19); several studies have shown that these simple methods can be implemented for both community (2, 8, 15, 19, 20) and high risk (3, 4) setting diabetic foot screening, independently of the variables collection method.

Patients with diabetes should have their feet checked at least once per year (2, 6, 10). However, to our knowledge, no foot reclassification periodicity was ever suggested. Therefore, we decided to implement a 1-year follow-up period. Another limitation was the fact that reliability of variables and systems was not assessed due to the retrospective character of this study.

In conclusion, although the selection of which system to apply is still unclear, this study shows that all these systems include pertinent variables, are easy to apply, present a high accuracy, and therefore are valuable tools to apply in our clinical practice.

Figure 1 Classification systems’ diagnostic accuracy. ROC, receiver operating curve; straight line, University of Texas system; dashed and dotted line, International Working Group on Diabetic Foot system; double dashed and dotted line, American Diabetes Association system; continuous dashed line, Scottish Inter-collegiate Guidelines Network system; discontinuous dashed line, Boyko et al. system.
Nevertheless, further validation studies should be performed on larger samples and in different settings with a longer follow-up period. Additionally, there is a great need to assess the reliability of the systems and their components, the impact of time in the systems’ validity and to consequently propose the most efficient foot examination reevaluation periodicity.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Author contribution statement

M Monteiro-Soares was responsible for data collection and article writing; A Vaz-Carneiro, S Sampaio, and M Dinis-Ribeiro were responsible for the statistical analysis and text editing and revision.

Acknowledgements

The authors would like to thank all staff of the Diabetic Foot Clinic at Centro Hospitalar de Vila Nova de Gaia/Espinho EPE, Portugal, who helped in the execution of this study.

References

15. Jeffcoate WJ. Stratification of foot risk predicts the incidence of new foot disease, but do we yet know that the adoption of routine screening reduces it? Diabetologia 2011 54 991–993. (doi:10.1007/s00125-011-2075-y)

Received 29 March 2012
Revised version received 19 June 2012
Accepted 27 June 2012

www.eje-online.org