




# Validity of Routinely Reported Rutherford Scores Reported by Clinicians as Part of Daily Clinical Practice

Laura L.M. van der Heijden, MSc<sup>1,2</sup>  Perla J. Marang-van de Mheen, PhD<sup>2</sup> Louis Thielman, MD<sup>1</sup>  
Pieter Stijnen, PhD<sup>3</sup> Jaap F. Hamming, MD, PhD<sup>4</sup> Inge Fourneau, MD, PhD<sup>1</sup>

<sup>1</sup> Department of Vascular Surgery, University Hospitals Leuven, Leuven, Belgium

<sup>2</sup> Department Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, The Netherlands

<sup>3</sup> Management Information and Reporting, University Hospitals Leuven, Leuven, Belgium

<sup>4</sup> Department of Vascular Surgery, Leiden University Medical Centre, Leiden, The Netherlands

**Address for correspondence** Inge Fourneau, MD, PhD, Department of Vascular Surgery, University Hospitals Leuven, Herestraat 49, 3000 Leuven, Belgium (e-mail: inge.fourneau@uzleuven.be).

Int J Angiol

## Abstract

Routinely reported structured data from the electronic health record (EHR) are frequently used for secondary purposes. However, it is unknown how valid routinely reported data are for reuse.

This study aimed to assess the validity of routinely reported Rutherford scores by clinicians as an indicator for the validity of structured data in the EHR.

This observational study compared clinician-reported Rutherford scores with medical record review Rutherford scores for all visits at the vascular surgery department between April 1, 2016 and December 31, 2018. Free-text fields with clinical information for all visits were extracted for the assignment of the medical record review Rutherford score, after which the agreement with the clinician-reported Rutherford score was assessed using Fleiss' Kappa.

A total of 6,633 visits were included for medical record review. Substantial agreement was shown between clinician-reported Rutherford scores and medical record review Rutherford scores for the left ( $k = 0.62$ , confidence interval [CI]: 0.60–0.63) and right leg ( $k = 0.62$ , CI: 0.60–0.64). This increased to the almost perfect agreement for left ( $k = 0.84$ , CI: 0.82–0.86) and right leg ( $k = 0.85$ , CI: 0.83–0.87), when excluding missing clinician-reported Rutherford scores. Expert's judgment was rarely required to be the deciding factor (11 out of 6,633).

Substantial agreement between clinician-reported Rutherford scores and medical record review Rutherford scores was found, which could be an indicator for the validity of routinely reported data. Depending on its purpose, the secondary use of routinely collected Rutherford scores is a viable option.

## Keywords

- ▶ validation
- ▶ medical record review
- ▶ peripheral artery disease
- ▶ Rutherford

When data are registered as structured data in the electronic health record (EHR), this could potentially enable secondary use for other purposes, that is, for research or quality assessment purposes.<sup>1–3</sup> Most studies use inclusion criteria

based on data that are frequently registered as structured data, such as age and severity of the disease.<sup>4</sup> However, it is unknown how correct and consistent structured data are which are routinely collected as part of the EHR.

DOI <https://doi.org/10.1055/s-0043-1761280>.  
ISSN 1061-1711.

© 2023. International College of Angiology. All rights reserved.  
This is an open access article published by Thieme under the terms of the Creative Commons Attribution License, permitting unrestricted use, distribution, and reproduction so long as the original work is properly cited. (<https://creativecommons.org/licenses/by/4.0/>)  
Thieme Medical Publishers, Inc., 333 Seventh Avenue, 18th Floor, New York, NY 10001, USA

The Rutherford classification is an example of routinely reported structured data in the EHR that is frequently used in studies as part of their inclusion criteria but is also relevant in daily practice as it may guide further treatment.<sup>5</sup> The Rutherford classification captures the different stages of peripheral artery disease (PAD), a frequently studied, common and progressive disease due to narrowed arteries increasingly reducing the blood flow to the legs.<sup>6</sup> Six stages of increasing PAD severity and one asymptomatic stage are distinguished.<sup>7</sup> Each stage is defined by patient symptoms (i.e., pain intensity and location), Doppler pressures, and the presence or absence of ulcers. Where the difference between some stages may be clear, for example, because they can be distinguished by the presence of wounds, it may be more subtle for other stages and less straightforward without the criteria at hand, resulting in interclinician variation in reported scores as part of daily clinical practice.

Retrospective studies mostly rely on the Rutherford scores as reported by clinicians in the EHR or the health insurance system to decide on the inclusion of patients.<sup>8,9</sup> Some studies check the correctness of the structured data of all patients manually to investigate who fulfills the inclusion criteria, which is a time-consuming activity and may not be feasible for large databases.<sup>10</sup> To the best of our knowledge, previous retrospective studies did not report about the validity of Rutherford scores assigned by the clinicians in the EHR. Therefore, it is unknown whether routinely collected Rutherford scores by clinicians as part of daily clinical practice are valid to be reused. The aim of the present study is to assess the validity of routinely reported Rutherford scores by clinicians as part of daily clinical practice, as an indicator for the validity of structured data in general that is routinely collected as part of the EHR and their potential for secondary use for research or quality assessment purposes.

## Methods

### Design

An observational study design was used to compare clinician-reported Rutherford scores with medical record review Rutherford scores. This study is part of a larger retrospective observational study to compare treatments among Critical Limb Threatening Ischemia (CLTI) patients, approved by the ethical committee of UZ/KU Leuven (reference number: s64053), for which all CLTI patients had to be identified.

### Terminology

Throughout this article “clinician-reported Rutherford score” refers to the Rutherford scores reported by clinicians as part of daily clinical practice and entered as structured data in the EHR. “Medical record review Rutherford score” refers to the Rutherford score that was assigned by a dedicated reviewer examining the available clinical information in free-text fields where clinicians report relevant symptoms, wounds, and other findings.<sup>3</sup>

### Clinician-Reported Rutherford Scores

From April 2016, the Rutherford scores for all inpatient and outpatient vascular surgery visits of the University Hospitals Leuven (tertiary care hospital, hereafter UH Leuven) could be reported as structured data in the EHR (Nexuzhealth) and remained in use for the entire study period. Since July 2017, the type of consultation, such as outpatient, emergency care, and hospitalized consultation, has been reported per hospital visit as well. Reporting by clinicians was a part of routine care, not forced by any hard stops in the EHR if not filled in, yet promoted by the clinical leadership.<sup>11</sup> Reporting was developed to be aligned with workflow of busy clinicians who spent most of their time taking care of patients. The reporting form contains structured fields to enter information about the patient and their medical situation, including Rutherford scores. At each inpatient or outpatient visit, Rutherford scores need to be reported for both the left and the right leg separately. Clinicians have nine radio buttons to rank the vascular status of each patient. The first seven options are Rutherford 0 to 6, the other two options are “amputation” and “acute ischemia” which were added to ensure that all possibilities could be reported so that missing scores are actual missing values rather than a situation where none of the options applied. For each visit, clinicians also describe the clinical condition of the patient in free-text fields.

### Medical Record Review Rutherford Score

The Rutherford scores and EHR free-text fields with clinical information were extracted for all inpatient and outpatient visits at the vascular surgery department between April 1 2016 and December 31 2018. Unique hospital visits were extracted because the severity of PAD for a given patient may change over time which will be reflected in a different score for a subsequent visit. These data were imported into an electronic case report form (e-CRF) within REDCap (Research Electronic Data Capture) for further review. REDCap is a secure web application for creating and managing databases.<sup>12</sup> From each visit, the clinician-reported Rutherford score was imported as well as all EHR free-text fields with the narrative information and, if available, results of noninvasive vascular tests that were potentially relevant to assign the Rutherford score. These text fields included information about pain intensity, pain location, disease progression, wound situation, medical history, and Doppler pressures, which in combination is the information that may be needed to distinguish between different Rutherford stages. All fields with routinely collected data were locked directly after importing the data in the e-CRF to ensure that these fields could not be changed. The reviewer then assigned a Rutherford score for each leg separately in a new field, following the predefined criteria according to Rutherford et al,<sup>7</sup> for all patients including those with missing Rutherford scores. Records that could not be assigned a Rutherford score based on the available text fields were left empty. All records were reviewed by one reviewer (L.T.) to ensure the most consistent application of scoring criteria. In case of uncertainty, cases were discussed with a second reviewer (L.vdH.); if no

consensus was achieved, a third reviewer (I.F.) who is an experienced vascular surgeon was consulted.

### Definitions

Medical record review Rutherford scores are based on the documented clinical symptoms, separately for each leg. If needed, Doppler pressures are checked. For clinician-reported Rutherford scores, the same definitions are applicable but knowledge of these definitions and thereby accuracy of the scores may differ between clinicians. Clinician-reported Rutherford scores are mostly based on clinical symptoms reported by the patient, as Doppler pressures are usually not (yet) available at the time of assigning a score. Rutherford 0 is assigned when no symptoms of PAD are documented. Rutherford 1, 2, and 3 denote patients who have mild, moderate, and severe claudication symptoms, respectively.<sup>7</sup> A walking distance of >100 m is used for mild symptoms and <100 m for moderate symptoms. Since the Rutherford classification does not specify severe claudication symptoms, we defined this as a patient only being able to walk a few meters or only indoors. Rutherford 4 is assigned for clinical symptoms of pain at rest, defined as intractable foot and ankle pain for more than 2 weeks while at rest. Rutherford 5 includes patients with minor tissue loss/ulcerations. Rutherford 6 denotes patients with major tissue loss and gangrene.<sup>7</sup> Amputations are documented if an amputation occurred above the ankle. In case of minor amputations, the classification is reported according to the guidelines.<sup>7</sup> Acute ischemia is registered if ischemia resulted from an acute cause with a sudden increase of pain for several hours to days. CLTI is defined as Rutherford scores 4 to 6, and non-CLTI as 0 to 3 separately for each leg.

When Doppler pressures are required, Rutherford 1 requires >50 mm Hg ankle pressure after exercise, at least 20 mm Hg lower than the resting value, Rutherford 3 an ankle pressure < 50 mm Hg after exercise, and Rutherford 2 intermediate ankle pressures. Rutherford 4 requires a resting ankle pressure of < 40 mm Hg and Rutherford 5 and 6 a resting ankle pressure of < 60 mm Hg.<sup>7</sup> If a wound or crust is present, that leg is assigned a Rutherford score 5 or 6 depending on the size. Until the wound is completely healed, it remains a wound and thus Rutherford score 5 or 6. In case of a venous wound, a Rutherford 0 is assigned.

### Statistical Analysis

Descriptive statistics were used to characterize the patient population, using mean and standard deviation (SD) for normally distributed variables, and median and interquartile range (IQR) for other variables. In all analyses, the medical record review Rutherford scores were considered as the golden standard because Rutherford scores were assigned based on consistent use of the required criteria. For the primary analyses, Rutherford 1 to 3 were combined since the generic nature of the distinction between Rutherford 1 to 3 can result in a difference in the assignment of the Rutherford score between clinicians due to subjectivity. Also, with the Doppler pressures not always available when assigning a score, this would be the reason for a difference between

clinician-reported and medical record review Rutherford scores rather than the difference being due to the reliability of clinician-reported scores. These Doppler pressures are not needed to assign the Rutherford scores 4 to 6 so that we kept the asymptomatic (Rutherford 0) and Rutherford 4, 5 and 6 as individual stages. Since the aim of this study is to evaluate the validity of the clinician reporting Rutherford scores and the options to report an amputation or acute ischemia were merely added to ensure completeness, these were classified as "other scores" and combined with missing scores in the analysis.

The Fleiss' Kappa was calculated to assess the level of agreement between clinician-reported and medical record review Rutherford scores. This was calculated for the left and right leg separately and both with and without missing values included, as missing values may indicate that clinicians consciously or subconsciously did not enter a score rather than assigning an incorrect score, so these analyses give additional information. The total numbers can be different for each leg for some analyses if a Rutherford score was assigned for one leg (retained in the analyses) but had a missing value for the other leg. The levels of agreement with the Fleiss Kappa values were classified as follows: 0.00 (poor agreement), 0.01–0.20 (slight agreement), 0.21–0.40 (fair agreement), 0.41–0.60 (moderate agreement), 0.61–0.80 (substantial agreement), and >0.80 (almost perfect agreement).<sup>13</sup> In addition, we examined the validity to distinguish between CLTI and non-CLTI patients. Contingency tables were used to calculate the sensitivity, specificity, and positive/negative predictive values (PPV/NPV) for clinician-reported Rutherford scores to correctly identify CLTI versus non-CLTI patients.

As a sensitivity analysis, we hypothesized that there may have been more missing Rutherford scores in 2016 than in later years, since the clinician reporting was introduced in 2016 which may take some time to be fully implemented. To test this hypothesis, we compared the percentage of missing Rutherford scores in 2016 versus all later years using a Chi-square test.

Statistical analysis was performed using SPSS Statistics version 25. The significance level was set at  $p < 0.05$  for all tests.

## Results

A total of 6,633 visits were included for medical record review. There were 3,281 unique patients who had 1 to 28 hospital visits per patient with a median of 3 (IQR, 1.0–5.0). The mean age of patients at the hospital visit was 67.9 years (SD, 13.8), and 2,109 out of 3,281 unique patients were male patients. The type of hospital visits was mostly outpatient consultations (36.9%), and most missing data (2,288/2,446) were due to the late start of registering this variable (► **Table 1**).

► **Tables 2 and 3** show the Rutherford scores assigned by clinicians compared with medical record review Rutherford scores for the left and right leg, respectively. The diagonal gives the number of records where scores were the same,

**Table 1** Baseline characteristics

Patient characteristics			
Unique patients	<i>n</i>	3,281	
Male gender	<i>n</i> (%)	2,109 (64.3%)	
Age (years)	Mean (SD)	67.9 (13.8 SD)	
BMI (kg/m <sup>2</sup> )	Mean (SD)	25.80 (7.86 SD)	
- Missing	<i>n</i> (%)	1,969 (60.0%)	
Smoking status			
- Current smoker	<i>n</i> (%)	699 (21.3%)	
- Former smoker	<i>n</i> (%)	1,232 (37.5%)	
- Nonsmoker	<i>n</i> (%)	886 (27.0%)	
- Missing	<i>n</i> (%)	464 (14.1%)	
- Total	<i>n</i> (%)	3,281 (100%)	
Number of hospital visits per patient	Median (IQR)	3.0 (1.0–5.0)	
Hospital visit characteristics			
Type of consultation			
- Outpatient consultation	<i>n</i> (%)	3,358 (36.9%)	
- Emergency care consultation	<i>n</i> (%)	390 (6.6%)	
- Hospitalized consultation	<i>n</i> (%)	439 (5.9%)	
- Missing	<i>n</i> (%)	2,446 (50.6%)	
- Total		6,633 (100%)	
Rutherford scores per limb		<b>Left</b>	<b>Right</b>
Clinician-reported Rutherford score			
- Rutherford 0	<i>n</i> (%)	2,738 (41.3%)	2,714 (40.9%)
- Rutherford 1	<i>n</i> (%)	553 (8.3%)	583 (8.8%)
- Rutherford 2	<i>n</i> (%)	487 (7.3%)	510 (7.7%)
- Rutherford 3	<i>n</i> (%)	595 (9.0%)	596 (9.0%)
- Rutherford 4	<i>n</i> (%)	101 (1.5%)	110 (1.7%)
Rutherford 5	<i>n</i> (%)	340 (5.1%)	299 (4.5%)
- Rutherford 6	<i>n</i> (%)	61 (0.9%)	37 (0.6%)
- Amputation	<i>n</i> (%)	177 (2.7%)	194 (2.9%)
- Acute ischemia	<i>n</i> (%)	0 (0%)	0 (0%)
- Missing	<i>n</i> (%)	1,581 (23.8%)	1,590 (24.0%)
- Total	<i>n</i> (%)	6,633 (100%)	6,633 (100%)
Medical record review Rutherford score			
- Rutherford 0	<i>n</i> (%)	3,423 (51.6%)	3,393 (51.2%)
- Rutherford 1	<i>n</i> (%)	828 (12.5%)	818 (12.3%)
- Rutherford 2	<i>n</i> (%)	538 (8.1%)	568 (8.6%)
- Rutherford 3	<i>n</i> (%)	496 (7.5%)	523 (7.9%)
- Rutherford 4	<i>n</i> (%)	85 (1.3%)	98 (1.5%)
- Rutherford 5	<i>n</i> (%)	471 (7.1%)	436 (6.6%)
- Rutherford 6	<i>n</i> (%)	76 (1.1%)	58 (0.9%)
- Amputation	<i>n</i> (%)	192 (2.9%)	209 (3.2%)
- Acute ischemia	<i>n</i> (%)	25 (0.4%)	29 (0.4%)
- Missing	<i>n</i> (%)	499 (7.5%)	501 (7.6%)
- Total	<i>n</i> (%)	6,633 (100%)	6,633 (100%)

Abbreviations: BMI, body mass index; IQR, interquartile range; SD, standard deviation.

**Table 2** Clinician-reported Rutherford scores versus medical record review Rutherford scores for the left leg

		Medical record review Rutherford score left						Total
		RF 0	RF 1–3	RF 4	RF 5	RF 6	Other or missing	
Clinician-reported Rutherford score left	RF0	2,498	114	1	20	1	104	2,738
	RF 1–3	71	1,478	4	16	0	66	1,635
	RF 4	3	9	62	12	0	15	101
	RF 5	5	6	1	309	12	7	340
	RF 6	1	0	0	4	51	5	61
	Other or missing	845	255	17	110	12	519	1,758
Total		3,423	1,862	85	471	76	716	6,633

Abbreviation: RF, Rutherford score.

Note: Other indicates acute ischemia or amputation.

**Table 3** Clinician-reported Rutherford scores versus medical record review Rutherford scores for the right leg

		Medical record review Rutherford score right						Total
		RF 0	RF 1–3	RF 4	RF 5	RF 6	Other or missing	
Clinician-reported Rutherford score right	RF0	2,485	99	3	28	1	98	2,714
	RF 1–3	61	1,534	8	19	0	67	1,689
	RF 4	2	8	71	13	0	16	110
	RF 5	11	6	1	266	9	6	299
	RF 6	1	0	0	1	33	2	37
	Other or missing	833	262	15	109	15	550	1,784
Total		3,393	1,909	98	436	58	739	6,633

Abbreviation: RF, Rutherford score.

Note: Other indicates Acute ischemia or amputation.

which was the case for 4,917 (74.1%) and 4,939 (74.5%) records for the left and right leg, respectively. The first reviewer assigned 655 records (9.9%) as uncertain, of which the majority was because of lack of information (553 records) so the second reviewer could not assign a Rutherford score for these either. The remaining 102 records were discussed between the first and second reviewer, and 11 records remained uncertain for which the third reviewer was consulted. The overall Fleiss Kappa was 0.62 (CI: 0.60–0.63) for the left leg and 0.62 (CI: 0.60–0.64) for the right leg, indicating substantial agreement between clinician-reported Rutherford scores and medical record review Rutherford scores. From the clinician-reported scores in the other/missing category, most were missing scores in 1,581 (89.9%) records for the left leg and 1,590 (89.1%) for the right leg. For the majority of clinician-reported missing Rutherford scores, records were missing for both left and right leg (1492 records). From the clinician-reported missing or other scores, almost half were assigned a Rutherford 0 score at medical record review (48.1% for left leg and 46.7% for right leg) but also a considerable part was still classified as other or missing (29.5% for left leg and 30.8% for right leg) indicating insufficient information or correctly reported amputations. No acute ischemia was reported by clinicians, whereas 25 (left leg) and 29 (right leg) were classified as such during the

medical record review. From the 192 (left leg) and 209 (right leg) amputations reported in medical record review, respectively, 165 (85.9%) and 178 (85.2%) were also clinicians reported with the remaining records mostly having missing clinician-reported scores. Considering only nonmissing clinician-reported Rutherford scores, Fleiss' Kappa showed almost perfect agreement for both left ( $k = 0.84$ , CI: 0.82–0.86) and right leg ( $k = 0.85$ , CI: 0.83–0.87).

CLTI patients were identified fairly well by clinician reporting, shown by the PPV of 89.8% for the left and 88.3% for the right leg (► **Tables 4 and 5**). Identification of non-CLTI patients was even better, with NPV of 95.2% for left and 94.9% for the right leg. Important to note is that particularly the NPV was affected by missing medical record review Rutherford scores, meaning that there was insufficient information in the EHR to decide which Rutherford score should be assigned. Sensitivity and specificity were moderate, 71.4 and 78.7% for the left leg and 66.5 and 78.8% for the right leg, respectively. This was mostly due to missing clinician-reported scores.

In the sensitivity analysis, from the 967 clinician-reported scores for the left leg in 2016, 273 (28.2%) scores were missing; for the right leg, 274 out of 967 (28.3%) scores were missing. The proportion of clinician-reported missing values were significantly higher in 2016 than in the

**Table 4** Clinician-reported CLTI diagnosis versus medical record review CLTI diagnosis for the left leg, based on Rutherford scores

Left		Medical record review			
		CLTI	Not CLTI	Other or missing	Total
Clinician reported	CLTI	451	24	27	502
	Not CLTI	42	4,161	170	4,373
	Other or missing	134	1,098	323	1,758
Total		632	5,285	716	6,633

Abbreviations: CLTI, chronic limb threatening ischemia; NPV, negative predictive value; PPV, positive predictive value.

Note: Other indicates acute ischemia or amputation.

Note:  $PPV = 451/502 * 100 = 89.8\%$ .

Note:  $NPV = 4,161/4,373 * 100 = 95.2\%$ .

Note:  $Sensitivity = 451/632 * 100 = 71.4\%$ .

Note:  $Specificity = 4,161/5,285 * 100 = 78.7\%$ .

**Table 5** Clinician-reported CLTI diagnosis versus medical record review CLTI diagnosis for the right leg, based on Rutherford scores

Right		Record review			
		CLTI	Not CLTI	Other or missing	Total
Clinician reported	CLTI	394	28	24	446
	Not CLTI	59	4,179	165	4,403
	Other or missing	139	1,095	550	1,784
Total		592	5,302	739	6,633

Abbreviations: CLTI, chronic limb threatening ischemia; NPV, negative predictive value; PPV, positive predictive value.

Note: Other indicates acute ischemia or amputation.

Note:  $PPV = 394/446 * 100 = 88.3\%$ .

Note:  $NPV = 4,179/4,403 * 100 = 94.9\%$ .

Note:  $Sensitivity = 394/592 * 100 = 66.5\%$ .

Note:  $Specificity = 4,179/5,302 * 100 = 78.8\%$ .

remaining period for the left leg (28.2 vs. 23.1%  $p = 0.001$ ) and right leg (28.3 vs. 23.2%  $p = 0.001$ ).

## Discussion

The present study has shown a substantial agreement between clinician-reported and medical record review Rutherford scores, with an almost perfect agreement when not considering missing values. Clinician-reported Rutherford scores correctly identified CLTI in almost 90% of the cases and non-CLTI in approximately 95%. These are conservative estimates as they were affected by missing medical record review Rutherford scores due to insufficient information in the EHR. A moderate sensitivity and specificity were shown, mostly due to missing clinician-reported Rutherford scores. Possible explanations for insufficient or missing information in the EHR could be, for example, lack of time, not being aware of the importance, not understanding the system, or just not being willing to enter all the data. It rarely occurred that the expert's opinion was necessary to be the decisive factor (11 out of 6,633), which suggests that little discussion is needed about the score. The percentage of missing clinician-reported scores was higher in the first year compared with the remaining period, as shown by the sensitivity analysis, which could be explained by the fact that surgeons needed a period during the first

year to get accustomed to structured reporting of Rutherford scores.

Although the validity of clinician-reported Rutherford scores has not been reported before, a previous study investigated the validity of PAD diagnosis in a national patient registry and found a PPV of 71.9% for vascular surgery departments.<sup>14</sup> The present study found considerably better PPVs of 89.8 and 88.3% for left and right respectively, which indicates a better ability to correctly identify CLTI patients. The high PPVs of the present study are most likely the result of frequent instruction and monthly feedback of (un)completed fields. No other validation studies have been published about the reporting of (a subset of) PAD or about specific routinely reported scores. More broadly, previous studies comparing reporting of a diagnosis by clinicians as part of daily clinical practice compared with medical record review found similar PPVs to the present study.<sup>15-17</sup> This could indicate that the present study and its findings are representative for other routinely reported scores or diagnosis.

Strengths of this study include that all patients who visited the vascular surgery department were included and scores validated, rather than a sample. This resulted in a large sample size, including both the start of the implementation and the period in which everyone was used to the need to report the Rutherford scores, thereby presenting real-world

data. However, some limitations should be noted. The validity of clinician-reported data depends on data completeness and accuracy of reporting. Missing data affected the sensitivity, specificity, and also the NPV, which may have been underestimated if the lack of documentation means that Rutherford 0 is the appropriate score. In addition, while a dedicated reviewer consistently applying the required criteria for the Rutherford score can be considered a strength compared with doctors who may be preoccupied with many different things during a consultation (i.e., treating the patient optimally), having only one dedicated reviewer is also a limitation in the sense that it may have caused observer bias and influenced the level of agreement. By regular discussions with another reviewer, we tried to minimize this bias. However, it would be even more optimal to have two independent reviewers to check all records, since despite strictly following the guidelines there might still be a judgmental difference. Finally, the single-center nature of the study is a limitation as we validated this specific type of clinician reporting within the EHR system of UH Leuven and it is unknown whether it can be generalized to other hospital types with a different patient mix or other countries with a different health care system.

The implications of our findings are that clinician-reported Rutherford scores during daily clinical practice can be used to reliably select CLTI patients for clinical research. Assuming that most of the missing medical record review Rutherford scores will indicate Rutherford 0, we can be rather confident that we will not miss any potentially eligible CLTI patients. These findings may thereby also act as an incentive to raise awareness of the importance of accurate clinician reporting and avoid missing data, particularly knowing that these data need to be reliable if it is reused. The decision whether to use the routinely collected data also depends on how specific patients need to be selected for secondary use. For instance, for studies requiring full agreement of specific Rutherford scores, for example, only 5 and 6, it may not be sufficiently reliable to solely rely on routinely collected data. Still, even in those cases, it may help to start from a smaller selection of CLTI patients rather than reviewing all patients. In case of large quality assessment procedures, a small difference in patients selected might have little impact on the outcome measures. Future research could focus on the validation of other routinely collected scores such as the Wifl (wound, ischemia, and foot infection) Classification System and on having two dedicated reviewers reviewing cases independently to further minimize observer bias and assess the interrater reliability.<sup>18</sup> In addition, other disciplines than surgery can benefit from the results of this study since the nature of the study focuses on routinely collected data.

## Conclusion

A substantial agreement between clinician-reported Rutherford scores and medical record review Rutherford scores was found, which increased to an almost perfect agreement when missing clinician-reported Rutherford scores were

excluded suggesting that those assigned by clinicians were valid. This agreement, together with a good ability to identify CLTI patients, makes reuse of these routinely collected Rutherford scores a viable option, particularly if these findings act to stimulate better EHR documentation and fewer missing clinician-reported scores.

## Funding

This research is funded by the joint-PhD program from the Leiden University Medical Center (reference 18-1921) and the Department Cardiovascular Sciences KU Leuven (Belgium).

## Conflict of Interest

None declared.

## Acknowledgments

The authors would like to thank Donna Peerboom for her contribution to data extraction.

## References

- 1 Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg* 2015;102(02):e93–e101
- 2 Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. *BMJ* 2010;341:c4226
- 3 Worster A, Haines T. Advanced statistics: understanding medical record review (MRR) studies. *Acad Emerg Med* 2004;11(02):187–192
- 4 Benchimol EI, Smeeth L, Guttman A, et al; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12(10):e1001885
- 5 Hardman RL, Jazaeri O, Yi J, Smith M, Gupta R. Overview of classification systems in peripheral artery disease. *Semin Intervent Radiol* 2014;31(04):378–388
- 6 Gerhard-Herman MD, Gornik HL, Barrett C, et al; Writing Committee Members ACC/AHA Task Force Members. 2016 AHA/ACC Guideline on the Management of Patients with Lower Extremity Peripheral Artery Disease: executive summary. *Vasc Med* 2017;22(03):NP1–NP43
- 7 Rutherford RB, Baker JD, Ernst C, et al. Recommended standards for reports dealing with lower extremity ischemia: revised version. *J Vasc Surg* 1997;26(03):517–538
- 8 Ricco JB, Gargiulo M, Stella A, et al. Impact of angiosome- and nonangiosome-targeted peroneal bypass on limb salvage and healing in patients with chronic limb-threatening ischemia. *J Vasc Surg* 2017;66(05):1479–1487
- 9 Stella J, Engelbertz C, Gebauer K, et al. Outcome of patients with chronic limb-threatening ischemia with and without revascularization. *Vasa* 2020;49(02):121–127
- 10 Biagioni LC, Pereira L, Nasser F, Biagioni RB, Burihan MC, Wolosker N. Comparison between antegrade common femoral artery access and superficial femoral artery access in infrainguinal endovascular interventions. *J Vasc Surg* 2021;74(03):763–770
- 11 Powers EM, Shiffman RN, Melnick ER, Hickner A, Sharifi M. Efficacy and unintended consequences of hard-stop alerts in electronic health record systems: a systematic review. *J Am Med Inform Assoc* 2018;25(11):1556–1566
- 12 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational

- research informatics support. *J Biomed Inform* 2009;42(02):377–381
- 13 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(01):159–174
  - 14 Lasota AN, Overvad K, Eriksen HH, Tjønneland A, Schmidt EB, Grønholdt MM. Validity of peripheral arterial disease diagnoses in the Danish National Patient Registry. *Eur J Vasc Endovasc Surg* 2017;53(05):679–685
  - 15 De Almeida Chaves S, Derumeaux H, Do Minh P, Lapeyre-Mestre M, Moulis G, Pugnet G. Assessment of the accuracy of using ICD-10 codes to identify systemic sclerosis. *Clin Epidemiol* 2020;12:1355–1359
  - 16 Pimentel MA, Browne EN, Janardhana PM, et al. Assessment of the accuracy of using ICD-9 codes to identify uveitis, herpes zoster ophthalmicus, scleritis, and episcleritis. *JAMA Ophthalmol* 2016;134(09):1001–1006
  - 17 Warwick J, Slavova S, Bush J, Costich J. Validation of ICD-10-CM surveillance codes for traumatic brain injury inpatient hospitalizations. *Brain Inj* 2020;34(13-14):1763–1770
  - 18 Valik JK, Ward L, Tanushi H, et al. Validation of automated sepsis surveillance based on the Sepsis-3 clinical criteria against physician record review in a general hospital population: observational study using electronic health records data. *BMJ Qual Saf* 2020;29(09):735–745