Ruling out coronary heart disease in primary care: external validation of a clinical prediction rule

INTRODUCTION
In primary care, 0.7–2.7% of patient encounters are due to chest pain.1–3 While the underlying aetiology in the majority of patients is non-cardiac (for example, musculoskeletal, psychological, oesophageal), coronary heart disease (CHD) accounts for 12.8–14.6% of cases of chest pain in this setting.2–5 GPs must reliably identify serious cardiac disease, while also protecting patients from unnecessary investigations and hospital admissions. Based on medical history taking and physical examination, they decide whether further diagnostic procedures are indicated.

Bösner and his colleagues at the University of Marburg, Germany, developed a simple clinical prediction rule (CPR) proposed to assist GPs in ruling out CHD in patients presenting with chest pain.4,5 The Marburg Heart-Score (MHS) is based on five findings of the medical history and physical examination (Table 1).

The authors derived the CPR using the data of 1199 unsel ected and consecutive patients aged ≥35 years, who presented with chest pain to 74 GPs in Germany (first Marburg chest pain study). The data were gathered in 2005 to 2006. The overall prevalence of CHD, chronic stable CHD, or acute coronary syndrome (ACS), was 15.0% in this sample. The area under the receiver operating characteristic curve (AUC) as a measure of overall discrimination was 0.87 (95% confidence interval [CI] = 0.83 to 0.91). The best discrimination was with a cut-off value of 3, which had a sensitivity of 86.4% (95% CI = 78.5% to 91.7%), a specificity of 75.2% (95% CI = 71.8% to 78.3%), a positive predictive value (PPV) of 34.9% (95% CI = 29.3% to 40.9%), and a negative predictive value (NPV) of 97.3% (95% CI = 95.5% to 98.4%).

The authors externally validated the MHS using the data of 672 unselected and consecutive patients aged ≥16 years who presented with chest pain in 58 primary care practices in Switzerland (TOPIC — Thoracic Pain in Community Study).6,7 These data were gathered in 2001. The overall prevalence of CHD in this study was 12.6%. Four out of the five variables of the MHS could be directly derived from the data. For the fifth variable (‘patient assumes cardiac origin of pain’), Bösner et al used ‘anxiety’ — defined as a positive answer to the question ‘Are you feeling very worried about your chest pain?’ — as a proxy variable. In contrast to common findings in prediction rule research,7 the MHS showed a higher predictive power in this sample than in the derivation cohort. The AUC was 0.90 (95% CI = 0.87 to 0.93). At the proposed threshold of 3 points, the score showed a sensitivity of 87.1% (95% CI = 79.9% to 94.2%), a specificity of 80.8% (95% CI = 77.6% to 83.9%), a PPV of 39.6% (95% CI = 32.6% to 46.6%), and a NPV of 97.7% (95% CI = 96.4% to 99.1%).
In this, second, Marburg chest pain study, the researchers aimed to investigate the diagnostic accuracy of the MHS in an independent sample and to evaluate the generalisability to new patients.

**METHOD**

**Study design and participants**

The researchers approached 208 GPs in the state of Hesse, Germany; 56 (26.9%) agreed to participate in the study. Over a period of 12 weeks, participating GPs were required to consecutively recruit every patient fulfilling the inclusion criteria. Patients had to be included if they had pain localised in the anterior chest either as the presenting complaint or on questioning, if they were ≥35 years, and if they agreed to participate. Patients were not eligible if chest pains had subsided for more than 1 month, or had already been investigated. Patients with traumatic chest pains were excluded from analysis. Data were collected between July 2009 and February 2010, following a pre-specified study protocol, which was approved by the ethics committee of the Faculty of Medicine, University of Marburg, Germany. All patients gave informed consent.

**Marburg Heart Score**

During the index consultation, all participating GPs gathered data on 16 items of the medical history and clinical examination including the variables of the MHS. To calculate the score, 1 point was assigned to each item and the points were totalled. The cut-off value used in the study for ruling out CHD was 3, as proposed by Bösner and colleagues. The GPs were blinded to the results of the reference standard.

**Reference standard**

The reference diagnosis was established using a delayed-type reference standard in combination with an independent expert panel. Study nurses contacted all patients by phone after 6 weeks and 6 months and asked about the course of chest pain, further medical consultations, and treatments including drugs or hospitalisations. Additionally, they contacted all GPs to receive relevant information about further consultations, diagnostic procedures, treatments, and discharge letters from specialists, or hospitals. If necessary, specialists and hospitals were approached directly. An expert panel consisting of two members of the research team (at least one GP and another research staff member) reviewed each patient’s data and decided if CHD had been the underlying cause for chest pain, using recommended criteria. As the delayed-type reference standard is based predominantly on follow-up data, patients were counted as ‘loss to follow-up’ if they could not be reached by phone and the GP had no relevant data after 6 months. If data were available but the expert panel achieved no conclusive diagnosis, these cases were accordingly counted as ‘inconclusive’. The expert panel was not blinded to the results of the index tests.

**Statistical analysis**

Losses to follow-up and cases with missing values in the score variables were assumed to be completely missing at random and were excluded from the analysis. Sensitivity was plotted against 1–specificity in the receiver operating characteristics (ROC) space for each cut-off value, and the AUC was calculated to assess the overall discriminative power of the MHS. Additionally, the sensitivity, specificity, likelihood ratios, and predictive values for the recommended cut-off of 3 points for ruling out CHD were calculated as proposed by Bösner and colleagues. Lower and upper limits of CIs for proportions were calculated using the Wilson procedure without a correction for continuity, and for likelihood ratios using the procedure recommended by Simel and colleagues. In the main analysis, inconclusive cases...
were excluded, but two sensitivity analyses were performed, treating inconclusive cases as ‘CHD positive’ or ‘CHD negative’, respectively. To compare the results of these analyses, and to compare the results of the current study with those of the derivation and the first validation cohort, the accuracy achieved in the different analyses was plotted in the ROC space and the AUC was calculated. Sample size calculation aimed to achieve a precise estimation of the score’s sensitivity.

RESULTS

In total, the GPs approached 939 patients fulfilling the inclusion criteria during the study period (Figure 1). Among these patients, 59 (6.3%) refused to participate, 15 (1.6%) presented with traumatic chest pain, and 12 (1.3%) were losses to follow-up. In 9 (1.1%) of the remaining 853 patients, the score could not be calculated due to missing values, leaving the data of 844 patients for analysis. In 480 patients (56.9%) the score was \( \leq 2 \).

The mean age of patients was 59.5 years (standard deviation [SD] = 13.9 years) and 435 (51.5%) were female. The reference diagnosis was CHD in 92 (10.9%) patients, including 21 (2.5%) with ACS. In 12 patients (1.4%), the reference diagnosis was ‘inconclusive’. Table 2 shows GPs’ and patients’ characteristics.

At the proposed cut-off value of 3 [positive result 3–5 points], the sensitivity, specificity, and NPV of the MHS were 89.1% (95% CI = 81.1% to 94.0%), 63.5% (95% CI = 60.0 to 66.9%), and 97.9% (95% CI = 96.2 to 98.9%), respectively (Table 3).

The ROC curves for the main analyses and the two sensitivity analyses were congruent, indicating that the results of the analysis were not significantly affected by excluding the 12 patients with an inconclusive diagnosis (Figure 2).

The AUC was 0.84 (95% CI = 0.80 to 0.88) in this sample. This was slightly lower when compared with the results for the first validation sample (Figure 3). However, the ROC curves of both validation studies were situated near the ROC curve drawn from the data of the derivation study, indicating that the accuracy of the MHS is generally robust.

For a cut-off value of 3, the differences in the individual ROC curves were mainly caused by variation in specificity, while the sensitivity proved to be very robust overall three samples.

Ten patients were falsely classified as ‘CHD negative’ by the MHS. One patient scored 1 point while the others scored 2 points. None of them died during follow-up.

In four of these patients, an ACS was diagnosed. Three males aged 48–53 years, two with history of smoking and one with a

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>GP s (n = 56)</th>
<th>Patients (n = 844)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean years [SD]</td>
<td>50.4 (6.8)</td>
<td>59.5 (13.9)</td>
</tr>
<tr>
<td>Female</td>
<td>17 (30.4)</td>
<td>32 (38.1)</td>
</tr>
<tr>
<td>Years in practice, mean [SD]</td>
<td>13.3 (7.8)</td>
<td>114 (13.5)</td>
</tr>
<tr>
<td>Urban practice location</td>
<td>32 (57.1)</td>
<td>145 (17.4)</td>
</tr>
<tr>
<td>Diagnosed cardiovascular disease</td>
<td>221 (26.2)</td>
<td>342 (40.5)</td>
</tr>
<tr>
<td>History of hypertension</td>
<td>424 (50.2)</td>
<td>342 (40.5)</td>
</tr>
<tr>
<td>History of diabetes mellitus</td>
<td>114 (13.5)</td>
<td>114 (13.5)</td>
</tr>
<tr>
<td>History of smoking</td>
<td>145 (17.4)</td>
<td>145 (17.4)</td>
</tr>
<tr>
<td>History of dyslipidaemia</td>
<td>342 (40.5)</td>
<td>342 (40.5)</td>
</tr>
<tr>
<td>Diagnosis of CHD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHD positive</td>
<td>92 (10.9)</td>
<td></td>
</tr>
<tr>
<td>CHD negative</td>
<td>740 (87.7)</td>
<td></td>
</tr>
<tr>
<td>Inconclusive</td>
<td>12 (1.4)</td>
<td></td>
</tr>
</tbody>
</table>

SD = standard deviation. Data are expressed as n (%) unless stated otherwise.
A 61-year-old female with a history of hypertension was diagnosed with a myocardial infarction caused by arterial embolism and without coronary arteriosclerosis.

The other six falsely negative patients were classified as stable CHD by the expert panel. In a 50-year-old male with a history of smoking and hypertension, the angina was caused by a myocardial bridging without coronary arteriosclerosis. Two females, aged 62 and 70 years, with persistent typical angina refused further investigations. Despite the uncertainty, the expert panel decided that the probability of CHD as an underlying cause outweighed the probability of any other cardiac or non-cardiac cause. A further three patients were a 52-year-old female with a history of dyslipidaemia, diabetes mellitus, and hypertension; a 53-year-old female with a history of dyslipidaemia; and a 75-year-old male with a history of dyslipidaemia. In both females, the chest pain was initially described as ‘not worse during exercise’ at the index visit. Within 14 days, both patients were referred to cardiologists, who classified the pain as effort angina.

In total, 270 patients who received a score value ≥3 were classified as ‘CHD negative’

Table 3. Accuracy of the Marburg Heart Score for a cut-off value of 3 (n = 832)

<table>
<thead>
<tr>
<th>Value</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity, %</td>
<td>89.1</td>
</tr>
<tr>
<td>Specificity, %</td>
<td>63.5</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>2.44</td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>0.17</td>
</tr>
<tr>
<td>Positive predictive value, %</td>
<td>23.3</td>
</tr>
<tr>
<td>Negative predictive value, %</td>
<td>97.9</td>
</tr>
</tbody>
</table>
by the expert panel. The MHS score values were 3 points (59.3%), 4 points (30.7%), and 5 points (10.0%). The mean age was 67.9 years (SD = 10.8 years) and 132 (48.9%) were females. Of these, the GPs referred 28 (10.4%) immediately to hospital and 67 (24.8%) as outpatients to specialists. The reference diagnoses were cardiovascular disorders other than CHD (15.2%), respiratory disorders (7.8%), gastrointestinal disorders (4.1%), chest wall syndrome (53.3%), psychogenic causes (7.0), and no specific diagnosis (12.6%).

**DISCUSSION**

Summary

In this study, the MHS showed a good discriminative power and diagnostic accuracy. The area under the curve was 0.84 (95% CI = 0.80 to 0.88). For a cut-off value of 3, the MHS showed a sensitivity of 89.1% (95% CI = 81.1% to 94.0%), a specificity of 63.5% (95% CI = 60.0% to 66.9%), a PPV of 23.3% (95% CI = 19.2% to 28.0%), and a NPV of 97.9% (95% CI = 96.2% to 98.9%). The sensitivity, negative likelihood ratio, and negative predictive value in particular were shown to be stable when compared to the derivation and the first validation study.

**Strengths and limitations**

This study has several strengths. The patients were highly representative of patients presenting with chest pain in primary care. The large sample size allowed precise estimation of sensitivity, even in this low-prevalence setting. While audits warranted the consecutive recruitment, comprehensive collection of relevant follow-up data reduced the potential for misclassification. Both audits and the comprehensive data collection, resulted in small numbers of losses to follow-up, of cases with missing values, and of cases with inconclusive diagnosis that might otherwise compromise the validity of a study.

The authors acknowledge that the study has some limitations. The researchers did not interfere in the diagnostic work-up. As a consequence, only some of the patients underwent tests recommended for diagnosing CHD. However, only including patients who were assigned to a comprehensive cardiologic evaluation by their GPs would have resulted in a highly selected sample that was not representative of the clinically relevant population under interest. The delayed-type reference standard is considered a reasonable alternative if the definite reference standard (for example, coronary angiography) is too invasive or otherwise inapplicable, and should be considered the most appropriate choice in a low-prevalence setting. The expert panel establishing the reference diagnosis was not blinded to the baseline data, including the results of the index tests. However, the panel often had to make a decision on the basis of limited data, since there was no requirement for GPs to use defined investigations. If the expert panel had been blinded to baseline data, even less data would have been available. This might have resulted in biased results due to misclassification bias, and a higher rate of cases with an inconclusive reference diagnosis. In the authors’ previous study, a blinded and unblinded reference panel showed a substantial and satisfying agreement (kappa = 0.62).

**Comparison with existing literature**

Before a diagnostic prediction rule can be recommended for use in clinical practice, the accuracy must be investigated in at least one independent sample. For the MHS, both validation studies, including the current study, showed a comparable and satisfying overall accuracy of the MHS based on the AUCs. The values of sensitivity and the corresponding likelihood ratios and NPVs for the recommended cut-off value of 3 were

---

**Figure 3. Empirical ROC curve and area under the curve (AUC) of the current study (validation cohort 2) compared with the results in the derivation cohort and validation cohort 1.**

![Empirical ROC curve](image-url)
nearly identical in both validation studies and the derivation study. This is of particular importance, as safely ruling out CHD is of special concern in the clinical situation under interest.

Both validation studies were comparable in some aspects such as setting, and prevalence of the target disease. In both studies, all patients presenting with chest pain were recruited, regardless of a history of CHD. On the other hand, the studies differed in relation to several aspects: they were conducted within different healthcare systems, different GPs participated, and the time span between data collection was about 9 years. In the validation study reported by Bösner and colleagues, the reference diagnosis was established by the participating GPs and patients to be recruited were aged ≥16 years. In the current study, an independent expert panel established the reference diagnosis, and patients to be recruited were aged ≥35 years.

Validation studies aim to provide evidence that the CPR can be generalised to new patients. Several authors have provided hierarchical frameworks for validation strategies. Although these frameworks differ in detail, all authors agreed that confirmatory results provide stronger evidence for the generalisability of a CPR if the individual studies differed in some aspects, for example, if different inclusion criteria were used, if different physicians participated, or if the studies were conducted in different countries, or within different healthcare systems.

According to these frameworks, and considering both the methodological differences between the validation studies on the one hand and the high degree of agreement of the results on the other hand, this study found strong evidence that the MHS can be generalised to new patients presenting with chest pain in primary care.

Several CPRs have been developed and validated for the clinical assessment of patients presenting with acute chest pain in the emergency department. Since this setting differs with regard to the prevalence and clinical presentation of myocardial ischaemia, their results should not be extrapolated to primary care. The authors are aware of only two other CPRs developed or validated for diagnosing CHD in primary care. Sox and colleagues developed a chest pain score for determining the probability of CHD in patients presenting with chest pain, using the data of patients presenting in secondary care. The score was based on seven findings from the patient’s medical history. Each item was weighted differently, resulting in a score between 0 to 25 points. When comparing the results of validation studies conducted in both primary and secondary care, Sox et al. found substantial differences in the distribution of CHD cases among the chest pain score subgroups. Gencer and colleagues at the University of Lausanne, Switzerland, derived a CPR for ruling out CHD in primary care, using the data of the Swiss TOPIC study. The CHD score is based on eight items from the patient’s medical history, which were weighted differently, resulting in a score between 0 and 11 points. When validated in an independent sample (first Marburg chest pain study), the rule showed an AUC of 0.75. Compared with these CPRs, the MHS proved to be more robust when applied to new patients and easier to use.

**Implications for practice and research**

Future research should determine whether the MHS rule improves patient outcome or reduces costs. However, considering the accuracy, the generalisability, and its ease of application, the authors find it appropriate to recommend its use in clinical practice. While the sensitivity and the NPV were shown to be stable when compared to the derivation and the first validation study, there was a remarkable variation in the specificity across studies. GPs should keep that in mind when using the MHS in practice. While they can largely rely on a negative result, they should consider further clinical assessment in patients with positive results, especially those with a score value of 3 points.
REFERENCES


7. Stevens RJ, Lasserson D. In primary care, when is chest pain due to coronary artery disease? CMAJ 2010; 182(12): 1281–1282.


