INTRODUCTION
Chest pain remains a diagnostic challenge in general practice. Although the underlying aetiology in the majority of patients is non-cardiac, for example, musculoskeletal, psychological, or oesophageal, coronary heart disease (CHD) accounts for 8.0–14.6% of chest pain cases in this setting.\(^1\)\(^2\)\(^3\) GPs must reliably identify serious cardiac disease while also protecting patients from unnecessary investigations and hospital admissions. Based on the results of the medical interview and physical examination, the GP determines the probability of CHD and decides whether additional tests or treatment are needed.

A simple prediction rule assisting GPs to identify patients with a low probability of CHD as the underlying cause of chest pain was developed.\(^4\) The Marburg Heart Score (MHS) is based on five findings of the medical history and physical examination; all five items contribute equally to the score (Table 1). The MHS was externally validated in two independent samples.\(^4\)\(^5\) Among the patients with a negative result (≤2 points), the prevalence of CHD was 2.3%\(^6\) and 2.1%.\(^5\) However, it remains unclear whether using the rule adds to the GP’s clinical judgement. Furthermore, the placement of the MHS in the sequence of diagnostic reasoning about patients with chest pain is uncertain.

In this study, the data of the second validation study\(^7\) were utilised to investigate whether using the rule adds to the GP’s clinical judgement. In order to reach that aim different diagnostic strategies using the MHS were defined and their accuracy compared with the GPs’ clinical judgement.

METHOD
A comparative diagnostic accuracy study was conducted. Over a period of 12 weeks, 56 GPs located in the state of Hesse, Germany, consecutively recruited patients with chest pain. Patients were eligible if they had pain localised in the anterior chest as primary or secondary complaint and if they were aged ≥35 years. Patients were excluded if chest pain had subsided for more than 1 month, had already been investigated, or was caused by a trauma. Patients with a known history of CHD were not excluded. Data were collected between July 2009 and August 2010. Study audits in each practice ensured consecutive recruitment and complete recording of study data. During the index consultation, study physicians gathered standardised data on the medical history and clinical examination including the variables contributing to the MHS.

Index tests
Of the 56 study physicians, 39 were unaware of the MHS (subsample A). After a medical interview and physical examination, but before conducting further tests, these physicians rated the probability of CHD as the underlying cause using a three-level scale:

• 0 ‘low probability: no further diagnostic or therapeutic steps regarding CHD are recommended’;
• 1 ‘intermediate probability: CHD is not the
How this fits in

Previous studies showed that the Marburg Heart Score is a valid, robust, and easy-to-use tool for ruling out CHD in patients presenting with chest pain in general practice. In addition, the current study showed its potential to improve the established diagnostic approach. Considering these points, the authors recommend its use in clinical practice.

most likely cause of the chest pain, but further diagnostic steps are recommended to rule out CHD; and
• 2 ‘high probability: CHD is the most likely cause of the chest pain’.

The physicians used their discretion when deciding on which clinical findings to base their judgement. This test strategy, which represents the ‘established’ diagnostic test in this situation, was referred to as ‘unaided clinical judgement’.

Three alternative test strategies were defined using the MHS:

• The established test strategy, the physician’s unaided judgement, was replaced with the MHS. This test strategy is referred to as ‘MHS’.
• The physician’s clinical judgement and the MHS were combined in a formal and standardised way using the latter as a triage test. Patients with definite MHS results were counted as negative (score ≤2 points) or positive (score ≥4 points).
In patients with an intermediate MHS score (3 points), the final test result was determined by the physicians’ unaided clinical judgement. This test strategy is referred to as ‘triage’.

In the third alternative a more integrative and less standardised implementation of the MHS was defined in clinical practice. The remaining 17 study physicians (subsample B) participated in the evaluation of a new clinical practice guideline for chest pain by the German College of General Practitioners and Family Physicians [DEGAM]. The MHS is part of the guideline. These physicians were instructed to explicitly calculate the MHS score and to consider the results when rating the probability of CHD using the same three-level scale. Likewise to the ‘unaided clinical judgement’ the GPs decided on which symptoms or findings they based their judgement on in a particular case. This test strategy is referred to as ‘aided clinical judgement’.
All study physicians were blinded to the results of the reference standard.

Reference standard
It was anticipated that most patients would have a low probability of CHD and that invasive testing would not be allowed for ethical reasons in these patients. A panel diagnosis is recommended as an acceptable and valid alternative when a single reference test is not applicable.6–8 All patients included in the study were contacted by phone after 6 weeks and again at 6 months, and asked about their chest pain, further medical consultations, and treatments including drugs or hospitalisations. Additionally, their GPs were contacted — and specialists and hospitals if referred — to obtain relevant information about further consultations, diagnostic procedures, treatments, and discharge letters. An independent expert panel consisting of at least one GP and one research staff member reviewed each patient’s data and used recommended criteria9–11 to decide whether CHD had been the underlying cause for chest pain. Audits and thorough follow-up guaranteed that all relevant data were available to these experts.
The experts resolved disagreements by discussion.

An assumption of this design is that CHD would manifest within the follow-up. As this 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 found no conclusive diagnosis, cases were accordingly counted as ‘inconclusive’. The experts resolved disagreements by discussion.

Table 1. Components of the Marburg Heart Score

<table>
<thead>
<tr>
<th>Score item</th>
<th>Assigned points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex/age (female ≥65, male ≥55)</td>
<td>1</td>
</tr>
<tr>
<td>Known clinical vascular disease*</td>
<td>1</td>
</tr>
<tr>
<td>Patient assumes cardiac origin of pain</td>
<td>1</td>
</tr>
<tr>
<td>Pain worse with exercise</td>
<td>1</td>
</tr>
<tr>
<td>Pain not reproducible by palpation</td>
<td>1</td>
</tr>
</tbody>
</table>

*Including coronary heart disease, cerebrovascular disease, or peripheral artery disease. 1 point is assigned for each score variable; three different risk categories are derived (low risk = 0–2 points; intermediate risk = 3 points; high risk = 4–5 points).
Table 2. Characteristics of GPs and patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Subsample A, %*</th>
<th>Subsample B, %*</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age, years (SD)</td>
<td>51.1 (6.8)</td>
<td>49.0 (7.3)</td>
<td>0.31</td>
</tr>
<tr>
<td>Range</td>
<td>37–42</td>
<td>36–42</td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>33</td>
<td>24</td>
<td>0.46</td>
</tr>
<tr>
<td>Mean years in practice (SD)</td>
<td>13.9 (7.6)</td>
<td>12.1 (8.3)</td>
<td>0.36</td>
</tr>
<tr>
<td>Range</td>
<td>1–27</td>
<td>1–27</td>
<td></td>
</tr>
<tr>
<td>Urban practice location</td>
<td>46</td>
<td>82</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Patients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age, years (SD)</td>
<td>60.2 (13.9)</td>
<td>57.4 (13.5)</td>
<td>0.007</td>
</tr>
<tr>
<td>Range</td>
<td>35–92</td>
<td>35–89</td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>51.7</td>
<td>50.8</td>
<td>0.80</td>
</tr>
<tr>
<td>History of vascular disease*</td>
<td>27.5</td>
<td>22.8</td>
<td>0.16</td>
</tr>
<tr>
<td>History of hypertension</td>
<td>52.4</td>
<td>44.5</td>
<td>0.035</td>
</tr>
<tr>
<td>History of diabetes mellitus</td>
<td>13.7</td>
<td>12.2</td>
<td>0.78</td>
</tr>
<tr>
<td>History of smoking</td>
<td>17.0</td>
<td>25.6</td>
<td>0.004</td>
</tr>
<tr>
<td>History of dyslipidaemia</td>
<td>40.8</td>
<td>38.6</td>
<td>0.54</td>
</tr>
</tbody>
</table>

*SD = standard deviation. *Unless stated otherwise. *Including coronary heart disease, cerebrovascular disease, or peripheral artery disease.

Statistical analyses

Study physicians’ three-level rating of CHD probability (‘low probability’ = test negative, ‘intermediate’ or ‘high probability’ = test positive) were dichotomised. For each test strategy sensitivity, specificity, predictive values, and likelihood ratios were calculated. Lower and upper limits of confidence intervals (CIs) were calculated for proportions using the Wilson procedure without a correction for continuity and for likelihood ratios using the procedure recommended by Simel and colleagues.

For each comparison between the established diagnostic test, the ‘unaided clinical judgement’, and one of the new test strategies (‘MHS’, ‘aided clinical judgement’, or ‘triage’), the differences between each strategy’s sensitivity and specificity were calculated, respectively. To compare the accuracy of the ‘MHS’ and the ‘triage’ strategy with that of the ‘unaided clinical judgement’, a paired design was used. Each test strategy was independently applied to the same group of patients (subsample A). Calculations of differences needed to account for the paired nature of the data. However, it was impossible to apply the ‘aided clinical judgement’ and the ‘unaided clinical judgement’ independently to the same sample of patients. Therefore an unpaired design was used applying and comparing the accuracy of these test strategies in two different samples (subsamples A and B).

To illustrate the possible effects of using the different strategies the total number of true positives, false positives, true negatives, false negatives, and overall test positives were modelled in an imaginary population of 1000 patients assuming that the overall prevalence of CHD in this population would be the same as in this study. These numbers and the respective CIs were calculated for each test strategy using the sensitivity, specificity, test-positive fraction, and their respective CIs as obtained from the study data. Calculations were performed using R 2.15.0.

RESULTS

Figure 1 illustrates the flow of patients. In total, 832 patients were included in the current analysis. The reference diagnosis was CHD in 92 (11.1%) patients. The 39 study physicians unaware of the MHS recruited 578 patients (subsample A), while the 17 study physicians who did integrate the MHS into their judgement recruited 254 patients (subsample B). Table 2 shows study physicians’ and patients’ characteristics. It is important to note that, due to the low overall prevalence of CHD in general practice, changes in sensitivity are accompanied by only slight changes in absolute numbers of true positives and false negatives, respectively. In comparison, changes in specificity are accompanied by more distinct changes in...
Table 3. Diagnostic accuracy of the four test strategies

<table>
<thead>
<tr>
<th>Test strategy</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR– (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unaided clinical judgement</td>
<td>82.9 (72.4 to 89.9)</td>
<td>61.0 (56.7 to 65.2)</td>
<td>2.13 (1.83 to 2.48)</td>
<td>0.28 (0.17 to 0.47)</td>
<td>22.7 (18.0 to 28.2)</td>
<td>96.3 (93.6 to 97.9)</td>
</tr>
<tr>
<td>Marburg Heart Score</td>
<td>91.4 (82.5 to 96.0)</td>
<td>60.6 (56.3 to 64.8)</td>
<td>2.32 (2.04 to 2.64)</td>
<td>0.14 (0.07 to 0.30)</td>
<td>24.2 (19.5 to 29.8)</td>
<td>98.1 (95.9 to 99.1)</td>
</tr>
<tr>
<td>Triage</td>
<td>81.4 (70.8 to 88.8)</td>
<td>72.6 (68.6 to 76.3)</td>
<td>2.98 (2.48 to 3.56)</td>
<td>0.26 (0.16 to 0.42)</td>
<td>29.1 (23.2 to 35.8)</td>
<td>96.6 (94.3 to 98.0)</td>
</tr>
<tr>
<td>Aided clinical judgement</td>
<td>90.9 (72.2 to 97.5)</td>
<td>66.8 (60.5 to 72.6)</td>
<td>2.74 (2.19 to 3.43)</td>
<td>0.14 (0.04 to 0.51)</td>
<td>20.6 (13.8 to 29.7)</td>
<td>98.7 (95.5 to 99.6)</td>
</tr>
</tbody>
</table>

LR+ = positive likelihood ratio. LR– = negative likelihood ratio. PPV = positive predictive value. NPV = negative predictive value.

Discussion

This study showed that using the MHS as a triage test can improve the accuracy of the GP’s clinical judgement with a similar sensitivity and a significantly higher specificity. There was also a trend indicating that integrating the MHS into clinical reasoning [aided clinical judgement] may improve both sensitivity and specificity.

Strengths and limitations

Audits guaranteed the consecutive recruitment, and comprehensive collection of relevant follow-up data reduced the potential for misclassification. Both resulted in a small number of losses to follow-up, cases with missing values, and cases with an inconclusive diagnosis that may otherwise compromise the validity of the study.

This study also has limitations. It was performed in one area of Germany and the generalisability of the results to other national healthcare systems may be limited. The expert panel was not blinded towards the baseline data, including the index test results. Knowing the results of the index tests could have influenced the decision of the expert panel increasing the risk of an incorporation bias. If the expert panel had been blinded, fewer data for decision-making would have been available. This may have resulted in a misclassification bias and a higher rate of cases with an inconclusive reference diagnosis. The study physicians were not randomly assigned to one of the two study groups and there were differences in the location of the practice, and the proportion of patients with a history of smoking and hypertension. This may have compromised the unpaired comparison between the ‘unaided’ and ‘aided clinical judgement’. The calculation of the overall sample size allowed to perform a precise estimation of the measures of the MHS’s diagnostic accuracy. The study was also not powered to investigate the differences between the respective strategies.

Comparison with existing literature

Only a few clinical prediction rules for CHD have been externally validated in general practice. The authors’ knowledge, this is the first study that investigated how one of these clinical decision rules adds to the GP’s judgement. This study gave some

Figure 2. Accuracy of the unaided clinical judgement compared with the accuracy of the three new test strategies that are based on the Marburg Heart Score.

Table 4 shows the absolute numbers of true positives, false positives, true negatives, false negatives, and overall test positives, assuming that all four test strategies were applied to an imaginary population of 1000 chest pain patients with a prevalence or pretest probability of CHD of 11%.

Summary

This study showed that using the MHS as a triage test can improve the accuracy of the GP’s clinical judgement with a similar sensitivity and a significantly higher specificity. There was also a trend indicating that integrating the MHS into clinical reasoning [aided clinical judgement] may improve both sensitivity and specificity.

Strengths and limitations

Audits guaranteed the consecutive recruitment, and comprehensive collection of relevant follow-up data reduced the potential for misclassification. Both resulted in a small number of losses to follow-up, cases with missing values, and cases with an inconclusive diagnosis that may otherwise compromise the validity of the study.

This study also has limitations. It was performed in one area of Germany and the generalisability of the results to other national healthcare systems may be limited. The expert panel was not blinded towards the baseline data, including the index test results. Knowing the results of the index tests could have influenced the decision of the expert panel increasing the risk of an incorporation bias. If the expert panel had been blinded, fewer data for decision-making would have been available. This may have resulted in a misclassification bias and a higher rate of cases with an inconclusive reference diagnosis. The study physicians were not randomly assigned to one of the two study groups and there were differences in the location of the practice, and the proportion of patients with a history of smoking and hypertension. This may have compromised the unpaired comparison between the ‘unaided’ and ‘aided clinical judgement’. The calculation of the overall sample size allowed to perform a precise estimation of the measures of the MHS’s diagnostic accuracy. The study was also not powered to investigate the differences between the respective strategies.

Comparison with existing literature

Only a few clinical prediction rules for CHD have been externally validated in general practice. To the authors’ knowledge, this is the first study that investigated how one of these clinical decision rules adds to the GP’s judgement. This study gave some
Table 4. Absolute numbers of true positives, false negatives, true negatives, false positives, and test positives in an imaginary population of 1000 patients with an 11% prevalence of CHD

<table>
<thead>
<tr>
<th>Test strategy</th>
<th>True positives (95% CI)</th>
<th>False negatives (95% CI)</th>
<th>True negatives (95% CI)</th>
<th>False positives (95% CI)</th>
<th>Test positives (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unaided clinical judgement</td>
<td>91 (80 to 99)</td>
<td>19 (30 to 11)</td>
<td>543 (505 to 580)</td>
<td>347 (385 to 310)</td>
<td>443 (403 to 484)</td>
</tr>
<tr>
<td>Marburg Heart Score</td>
<td>101 (91 to 106)</td>
<td>9 (19 to 4)</td>
<td>539 (501 to 577)</td>
<td>351 (389 to 313)</td>
<td>457 (417 to 498)</td>
</tr>
<tr>
<td>Triage</td>
<td>90 (78 to 98)</td>
<td>20 (32 to 12)</td>
<td>644 (611 to 679)</td>
<td>244 (279 to 211)</td>
<td>339 (302 to 379)</td>
</tr>
<tr>
<td>Aided clinical judgement</td>
<td>100 (79 to 107)</td>
<td>10 (31 to 3)</td>
<td>595 (538 to 646)</td>
<td>295 (352 to 244)</td>
<td>382 (324 to 443)</td>
</tr>
</tbody>
</table>

CHD = coronary heart disease. For each test strategy the absolute numbers of the true positives among the 110 patients who have CHD, the absolute numbers of true negatives among the 890 patients who do not have CHD, and the absolute numbers of test positives including the respective CIs were calculated using the fractions of true positives, true negatives, test positives, and the respective CIs as obtained in this study. The absolute numbers of false negatives and false positives including the boundaries were calculated by completing the respective 2by2 tables. Please note that by using this approach the number of test positives is similar, but not identical to, the sum of true positives and false positives.

Funding
This study was funded by the Federal Ministry of Education and Research (BMBF — grant no. FKZ 01GK0701). The funding source had no involvement in the study.

Ethical approval
The study protocol was approved by the Ethics Committee of the Faculty of Medicine, Philipps University of Marburg, Germany. All participants gave informed consent before taking part.

Provenance
Freely submitted; externally peer reviewed.

Competing interests
The authors have declared no competing interests.

Acknowledgements
We thank all participating study physicians for their cooperation. We also would like to thank Juliette Rautenberg for providing English-language editing, improving the precision and fluency of the manuscript.

Discuss this article
Contribute and read comments about this article: bjgp.org/letters

insights into how different ways of using the MHS may affect the accuracy of the initial diagnostic approach to chest pain patients.

The performance of the MHS was at least equal to that of the physicians’ unaided clinical judgement. Considering the simplicity of the MHS, this result is remarkable. Evidence was found that using the MHS as a triage test can improve the specificity of the initial assessment. A disadvantage of this formalised use of the MHS as a triage test may be that the MHS considers only one of the relevant causes of chest pain. Integrating the MHS into clinical reasoning in an informal way probably better reflects how clinicians actually use such rules. This approach resulted in an improvement of both sensitivity and specificity. These differences were not significant. However, an adequately powered (power = 80%) study that aims to detect a difference of 8% in sensitivity between the ‘aided’ and ‘unaided clinical judgement’ using an unpaired design and assuming a prevalence of 11% would require about 2500 patients in each group.

Implications for practice
When comparing different test strategies one must consider the clinical situation and the respective consequences of different test results for patient management. In the early workup of the patient with chest pain, a negative test result would imply the need to consider reasons other than CHD for chest pain. Consequently, a false-negative result may lead to harm from a delayed or missed diagnosis. On the other hand, the consequence of a positive test result would be further diagnostic workup for CHD. Thus, a false-positive result may lead to anxiety and morbidity from additional testing. It can be assumed that the potential harm of the former outweighs that of the latter.16

In addition, one must also consider different perspectives. From the perspective of a single patient, a high probability of not having the disease given a negative test result, the negative predictive value (NPV), is highly desirable. What is high enough may depend on the patient’s preferences. However, all test strategies showed similar and reasonably high NPs ranging from 96.3% (unaided clinical judgement) to 98.7% (aided clinical judgement). There was also a trend indicating that the triage strategy results in a higher positive predictive value; that is, the probability of having the disease given a positive test result.

From the health system’s perspective, a high sensitivity and a high specificity are desirable. Compared with the unaided clinical judgement, using the triage strategy would result in a comparable absolute number of true positives and a significantly higher number of true negatives. As a result, a significantly lower number of patients without CHD would undergo unnecessary tests. This would be accompanied by a decreased number of test-positive patients who would need further diagnostic workup, indicating that this test strategy may also help reduce costs for further testing.

Clinicians should be aware that the MHS was developed and validated to assess patients presenting with chest pain. Studies conducted in various settings suggested that 7–20% of patients with myocardial ischaemia have no symptoms at all or present exclusively with atypical symptoms like dyspnoea or fatigue, and that this proportion is increased among older patients and patients with diabetes.17–20 It is not known how well the MHS would perform in these patients. Ebell stated that, besides accuracy and the potential to improve patient-related outcomes, useful clinical decision rules should be transparent, easy to use, and should have a good face validity.21 The ‘any three of five’ rule of the MHS could hardly be easier to use and fits well into the daily routine of busy clinicians. The criteria used in the rule have a high face validity, leading to a good acceptance of the rule.22 It can also be assumed that many GPs already use at least some of them to base their judgement on. However, the advantage of the MHS is the standardised combination of these criteria, leading to greater transparency.
REFERENCES


