Evaluating the accuracy of a simple heuristic to identify serious causes of low back pain

Norbert Donner-Banzhoff\textsuperscript{a}, Tobias Roth\textsuperscript{a}, Andreas C Sönnichsen\textsuperscript{a}, Judith Luckmann\textsuperscript{b}, Corinna Leonhardt\textsuperscript{b}, Jean-F Chenot\textsuperscript{c}, Annette Becker\textsuperscript{a,c}, Stefan Keller\textsuperscript{b,d}, Frances Griffiths\textsuperscript{e} and Erika Baum\textsuperscript{a}


\textbf{Background.} Among patients presenting with low back pain (LBP), GPs have to identify those with serious, treatable conditions. However, excluding these conditions in every patient with LPB is time consuming and of low yield. We have suggested that identifying those patients where these serious conditions need to be considered can be made more efficient through asking patient if they feel their LBP is new or unfamiliar in some way.

\textbf{Objective.} To evaluate the diagnostic validity of a simple heuristic based on the patient’s view of the familiarity of LBP.

\textbf{Methods.} Cross-sectional diagnostic study with delayed-type reference standard, nested within a three-arm randomized trial of quality improvement for LBP. A total of 1378 patients presenting, with LBP, to one of 126 participating GPs were included. They were asked whether their LBP was familiar or not (index test). At 1 year, patients were interviewed with regard to relevant conditions that in hindsight might explain their LBP. Reviewers deciding on disease status (reference standard) were blinded to the results of the index test.

\textbf{Results.} Totally 1190 patients answered the index test question and were available for interview at 1 year. Only four of these had a serious cause of their LBP. Two of these were identified by the familiarity heuristic, resulting in low sensitivity.

\textbf{Conclusion.} The number of diseased patients was too small to obtain a reliable estimate of sensitivity. Low prevalence of serious disease in primary care poses difficulties for diagnostic research. In hindsight we would question whether an RCT-setting emphasizing non-specific LBP is suitable for this kind of research. At present, the familiarity heuristic cannot be recommended for patients presenting with LBP.

\textbf{Keywords.} Low back pain, medical history taking, early diagnosis, sensitivity and specificity.

\textbf{Introduction}

Current guidelines require GPs to identify serious conditions in patients presenting with low back pain (LBP). Symptoms known as ‘red flags’ should alert medical practitioners to conditions such as fracture, neoplasm, infection or inflammation, which require specific management. While LBP is one of the most frequently presented symptoms in primary care, GPs encounter the specific conditions mentioned above only in very rare cases. Most patients in general practice suffer from non-specific LBP.

To sift the common and benign from the rare and fatal is perhaps the greatest challenge of general practice. To meet this challenge, GPs use simple heuristics or mental shortcuts for their diagnostic and therapeutic
decision-making, e.g. rules of thumb\(^2\) and mindlines.\(^3\) In several areas of professional work, simple heuristics have been identified and found to perform no worse than complex decisional algorithms.\(^4\)

The objective of this study was to establish the diagnostic accuracy of a discrepancy heuristic\(^5\) to identify serious conditions underlying LBP. It consisted of asking patients with LBP, whether they had experienced the present kind of LBP before, i.e. whether it was familiar, or not. If familiar, we assumed there to be no serious cause (i.e. test negative). If the LBP was not familiar (i.e. test positive) the answer was taken as an indication for a potentially serious cause requiring at least closer consideration of specific ‘red flags’. If this discrepancy heuristic was accurate, ‘red flags’ would have to be checked only in selected cases, i.e. if the LBP was unfamiliar. This would make the diagnostic process more efficient.

A large cohort of patients with LBP within a randomized controlled trial provided the opportunity to evaluate this simple heuristic.

**Methods**

The analysis reported here is based on a cohort of patients with LPB recruited into a cluster-randomized controlled trial evaluating strategies to improve the quality of care.

Totally 126 primary care practices were allocated to one of three arms:

1. Intense guideline implementation by local opinion leaders, interactive CME-sessions and academic detailing.
2. As for group 1 plus training of practice nurses in motivational counselling strategies.
3. Printed version of the guideline was received by mail (control arm).

The study used the LBP-guideline published by the German Society of General Practice/Family Medicine. Participating practices were asked to consecutively recruit up to 20 patients aged 20 years or older presenting with LBP. Inclusion criteria were LBP on the day of recruitment irrespective of duration, novelty or previous history. Exclusion criteria were insufficient language skills, pregnancy and isolated thoracic pain.

At baseline data on demographics, LBP history, physical activity, general health status and functional status were collected by questionnaire and telephone interview. The written questionnaire included the question: ‘Is the LBP familiar to you?’ which could be answered ‘yes’ or ‘no’. This was the index test pertinent to the analysis presented here.

Trained study nurses contacted all study patients by phone to collect data for study end points, 12 months after entering the study. Pertinent to the analysis presented here, they asked highly sensitive filter questions related to relevant serious conditions (see Table 1) that might have caused LBP at the time of recruitment. If at least one of these were answered in the affirmative, diagnosis and/or complaints were recorded and patients were asked to be available for another interview to clarify the diagnosis. In the following telephone interview T. R. gathered details on health care utilization (e.g. hospital treatments, medication, present complaints and impairments). If necessary for establishing the diagnosis they were asked to give permission for T. R. to access their GP’s documentation for hospital discharge letters, investigations and diagnoses.

A reference committee consisting of two experienced GPs (E.B., N.D.B.) and a senior medical student (T. R.) reviewed the evidence collected for each patient. Based on this information, patients were judged to either have a relevant condition or not (delayed-type reference standard).\(^6\) All interviewers and the reference committee were kept blind to the answers to the index test question.

The heuristic being tested in this study is intended as a first filter of the GP’s diagnostic process. Therefore, of the established measures of diagnostic accuracy, sensitivity would be particularly important.

The study design was approved by the ethics review board of the University of Marburg, Faculty of Medicine. This report takes the criteria suggested by the STARD initiative\(^7\) into account where applicable.

**Results**

Of the 1378 patients recruited, 1353 answered the question with regard to the familiarity of their LBP (index test). Of these patients, 1190 were available for follow-up at 1 year. This group forms the basis of our analysis. Demographic and clinical information is presented in Table 2.

Totally 963 patients answered ‘no’ to all the filter questions at 1 year. We, therefore, classified these as

<table>
<thead>
<tr>
<th>Relevant serious conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herniated vertebral disc with impaired bladder, bowel or motor function</td>
</tr>
<tr>
<td>Rheumatoid arthritis, Reiter’s Syndrome and other specific inflammatory disease</td>
</tr>
<tr>
<td>Chronic inflammatory bowel disease or yersiniosis with pelvic joint involvement</td>
</tr>
<tr>
<td>Epidural abscess</td>
</tr>
<tr>
<td>Vertebral or rib fracture (osteoïporotic or other)</td>
</tr>
<tr>
<td>Primary or secondary spinal tumour</td>
</tr>
<tr>
<td>Gynaecological tumour</td>
</tr>
<tr>
<td>Colorectal tumour</td>
</tr>
<tr>
<td>Peptic ulcer</td>
</tr>
<tr>
<td>Ischaemic heart disease</td>
</tr>
</tbody>
</table>
not having a serious condition as a cause of their back pain. Among those who answered positively, 13 refused a further telephone interview or could not be reached. However, based on free text recorded at their 1 year follow-up interview, e.g. complaints and treatments, the reference committee was still able to classify them as having a serious condition as a cause of their back pain, or not (reference standard). Thus in total we had sufficient information to classify the outcome for 1189 of 1190 patients (see Fig. 1). There was only one patient for whom classification was not possible even after documentation from the GP had been reviewed. This was a female patient diagnosed with carcinoma of the sigmoid colon 10 months after recruitment into the study. We undertook a sensitivity analysis adding this patient to the group of patients with the disease (see below).

Only four patients were deemed to definitely suffer from one of the relevant diseases as a cause of their LBP. Two suffered from a herniated vertebral disk requiring operation, one was diagnosed with vertebral metastases of breast carcinoma and one had spinal nerve damage after operation for a relapse of ovarial carcinoma.

The familiarity heuristic identified two of these resulting in a sensitivity of 0.5 with a 95% CI of 0.01–0.99 (see Table 3). However, given the small number of diseased patients, any estimate of sensitivity or other measure of diagnostic efficacy would be imprecise. The cases missed by the heuristic had a herniated disc and vertebral metastases, respectively.

Sensitivity analysis, including as diseased the patient with carcinoma of the colon mentioned above, showed a sensitivity of 0.6, which did not change our main conclusion.

Discussion

In a delayed-type cross-sectional study with 1190 primary care patients a simple discrepancy heuristic could not be shown to accurately identify serious underlying causes of LBP.

The main limitation of the study is the small number of patients with serious conditions that the ‘red flag’ symptoms are supposed to identify. Even if the sensitivity of the heuristic was high, with only four or five cases the estimate of its sensitivity would have been too imprecise to justify a clinical recommendation. However, it is possible that the sensitivity is sufficient, but this would only show in larger studies. Future studies addressing the issue of diagnostic effectiveness at the primary care level will need very large samples.

The main trial aimed to test the effect of interventions to improve the management of non-specific LBP. GPs may, therefore, have avoided recruiting patients

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Baseline characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (female%)</td>
<td>692 (58%)</td>
</tr>
<tr>
<td>Age (mean [standard deviation; range])</td>
<td>49 [13.3; 20–91]</td>
</tr>
<tr>
<td>Level of education (%)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>170 (16%)</td>
</tr>
<tr>
<td>Medium</td>
<td>452 (42%)</td>
</tr>
<tr>
<td>Low</td>
<td>439 (41%)</td>
</tr>
<tr>
<td>Living with partner (%)</td>
<td>828 (80%)</td>
</tr>
<tr>
<td>Applied for disability pension (%)</td>
<td>85 (8.5%)</td>
</tr>
<tr>
<td>Duration of back pain [years—median (range)]</td>
<td>16 (0–75)</td>
</tr>
</tbody>
</table>

Percentages are based on valid answers.

---

**Figure 1** Patient flow
with serious conditions so the proportion of four/five cases among 1211 patients with LBP may underestimate the true number of cases. Still the proportion of serious causes of LBP in the study population must be assumed to be very small, probably smaller than in previous reports on samples quite different from ours. Deyo and colleagues found a 0.66% prevalence of cancer and a 4% (respectively <1%) prevalence for osteoporotic or traumatic fracture, shown in a prospective cohort study of patients with LBP consulting in a walk-in clinic of a public health hospital. The prevalence of other causes of complicated LBP is usually cited with 0.01% for infection and 0.3% for inflammatory arthritis. It is possible that non-availability for follow-up in our study was associated with disease status; patients who were very ill may have refused consent for follow-up or were not available.

The index test question could be improved by including a reference to time. As an example, imagine an elderly patient suffering for 2 months from LBP caused by a (still undiagnosed) secondary neoplasm of the spine. He might have consulted his GP for his pain 2 weeks previously. Since he has experienced the symptom for several weeks and has consulted his GP, he may answer the familiarity question in the affirmative. However, if from a clinical point of view, this was the patient’s first episode of LBP, his pain would be regarded as new and in need of further investigation. This imaginary sequence could be clarified by taking a short oral history, which would also alleviate concerns that answers to written questions might differ from those given to questions asked in person. Patients and GPs would aim to detect change, which was either the onset of new LBP or a change in character of previous LBP.

We know of two more studies evaluating brief instruments in the assessment of LBP. Dunn and Croft showed ‘bothersomeness’ as measured by a single question to correlate with work absence and health care consultations for LBP 6 months later. Dionne et al. developed a 7-item algorithm to identify patients with a high risk of adverse occupational outcomes among LBP patients in primary care. Both instruments deal with the prognosis of non-specific LBP, whereas we have tried to improve diagnosis by separating non-specific LBP from specific serious causes.

GPs are expected to reliably identify rare instances of potentially serious conditions presenting as common symptoms. However, because of the low incidence, personal experience of diagnosing many diseases will always be limited. GPs, therefore, use rules that may involve assessment of context or may identify discrepancy from usual presentation or utilization behaviours. These rules operate at a very abstract level, transcending individual symptoms, findings or diseases. The more GPs provide continuous care beyond single disease episodes, the more they will have a frame of reference to identify discrepancies from the ‘usual’. Paying attention to discrepancies when deciding on further investigations is a good example of a simple rule exploiting the ecological structure of a working environment.

However, these and other simple rules or heuristics used by or recommended to GPs need checking for their accuracy. This study should inform the design of future studies in this area.

### Declaration

The study obtained ethical approval from the ‘Kommission für Ethik in der ärztlichen Forschung des Fachbereichs Medizin der Philipps-Universität Marburg’ (date 09.04.2002; study no. 35/02). Funding: Funding was provided by the Federal Ministry of Education and Research (BMBF) No. BMBF 01EM0113.

The authors deny any conflict of interest.

### References


