Screening for Cluster Headache—Introduction of the SMARTED Scale

Heiko Pohl 1,*, Marco Joos 1, Maria S. Neumeier 1, Miranda Stattmann 1, Andreas R. Gantenbein 1,2 and Susanne Wegener 1

Abstract: Patients with cluster headache often report a long diagnostic delay. This study creates and validates a screening test that could help speed up the diagnostic process. We invited patients to enrol in this diagnostic case–control study if a trigeminal autonomic headache had been suspected or confirmed. Patients in whom the diagnosis of a cluster headache was not made were controls. First, all participants answered 22 diagnostic questions with “yes” or “no”. Next, we eliminated questions that did not distinguish well between the groups. Then, the variables entered a regression model with the headache diagnosis as the dependent variable. Finally, we combined the remaining variables into a diagnostic scale and tested its accuracy. Seventy-four patients participated, 45 of whom suffered from a cluster headache. The analyses identified five questions distinguishing cluster headache patients and controls. These addressed smoking, being awakened by the pain, restlessness during the attack, unilateral tearing, and duration of the attack (hence, the “SMARTED” scale). The area under the ROC curve was 0.938; sensitivity and specificity, the positive and negative predictive values were 98%, 65%, 81% and 94%, respectively. The SMARTED scale validly and accurately screens for cluster headache in patients suspected of a trigeminal autonomic headache.

Keywords: diagnostic accuracy; diagnostic uncertainty; headache diagnosis; headache service; primary headache

1. Introduction

Recognizing cluster headache (CH) is crucial because of specific therapeutic strategies and the tremendous disease burden [1–4]. However, surveys report protracted diagnoses [5,6]. Many doctors rarely encounter and perhaps do not expect a patient to suffer from CH [7,8]. More generally, non-neurologist physicians struggle to diagnose a headache [9]. Thus, people likely suffering from CH need to consult a specialist. Still, guidelines advise consulting general practitioners first, who then decide on a referral [10]. Although this practice may allot resources efficiently, it also introduces obstacles [11]. Some patients will never proceed to a neurologist.

Furthermore, the diagnostic process depends on the patient’s history, which the physician cannot validate. In particular, misunderstandings and imprecise information imperil diagnostic accuracy.

Despite these difficulties, patients likely to suffer from CH must be identified quickly. A screening test could help reach that objective with a positive result suggesting consulting an expert without detour. This approach resembles Red Flags estimating patients’ probability of a secondary headache [12]. An elevated risk suffices to justify additional diagnostic steps; certainty is not a prerequisite.

The development of such a test requires pieces of information that patients can reliably provide. In other words, physicians ought to infer the likely diagnosis from what patients know—not from what they would like to know. For example, unobtrusive symptoms
outhshined by the pain—such as Horner’s syndrome—could help make a headache diagnosis but be challenging to recall. Nevertheless, we hypothesize that it is possible to use a questionnaire to screen for CH.

Previously published questionnaires attempted to diagnose [13,14] or screen for CH [15–20]. While most studies validated the questionnaires in the patients treated at university headache centers, one study recruited individuals through a website [13], and one study included a general population sample [14]. While the former study resulted in a low sensitivity [13]; the latter could not diagnose cluster headache at all [14]. Thus, currently, there is no questionnaire suitable for screening the general population for CH available.

We suspect that it may be overly optimistic to assume that questionnaires can diagnose headache disorders, and developing a screening tool may be more realistic.

This study assesses the discriminatory power of several self-administered diagnostic questions and develops a screening test distinguishing CH from its mimics. In this first step, we include patients treated at a tertiary headache center. A subsequent project will assess the scale’s diagnostic accuracy in an unselected sample. Ultimately, we aim to develop a tool to prevent missed or protracted diagnoses.

2. Materials and Methods

2.1. Eligibility

Adult patients treated in the outpatient clinic of the department of neurology of the University Hospital of Zurich, Switzerland, from January 2012 to March 2022, were eligible to enroll if the referring physician or the treating neurologist had suspected them of suffering from a trigeminal autonomic cephalgia (TAC). We assumed that if a TAC is a differential diagnosis, the headache must resemble a TAC, at least to some degree (“mimics”).

2.2. Study Design

The authors of this study created 22 dichotomous diagnostic questions aiming to distinguish CH from phenotypically similar headaches based on their experience and assumptions. All questions could be answered with “yes” or “no”. We composed the original questionnaire in German; this article only lists English translations.

We conducted a diagnostic case–control study with retrospective sampling to test the diagnostic questions. To that end, we divided the participants into two subgroups. The diagnosis of a cluster headache had been confirmed in patients in the first group (reference-standard positive); those in the second group had received another diagnosis (reference-standard negative).

We sent all eligible patients the list of questions, an information letter, and a consent form. If they did not respond to our invitation, we attempted to contact them once by phone and invited them to participate; all questionnaires were completed on paper. Patients were asked to answer the questions with either “yes” or “no”. If they had several headache types, we advised them to focus on the most painful one. If the answers differed between their attacks—e.g., an attack might not always be excruciating—we advised them to provide the most commonly applicable answer.

No number of participants was predefined prior to the study; the available data determined the sample size. In addition, we did not conduct a power calculation prior to the data collection. All data were collected specifically for this study from December 2021 to June 2022 and have not been published previously.

2.3. Data Analysis

First, we used the chi-squared test to analyze whether the diagnostic questions distinguished CH from mimics and calculated $p$ values and odds ratios. The null hypothesis is that the diagnostic questions do not distinguish between the two groups. Questions with $p$ values $> 0.20$ did not enter further analysis. We set the threshold $p$ value relatively high in order not to eliminate potentially helpful questions too early.
Next, we built a binary regression model and included all remaining diagnostic questions as independent variables. The dependent variable was the group allocation (CH vs. mimic); a backward elimination process excluded dependent variables in the model if the remaining variables resulted in a better prediction. Besides, the Hosmer and Lemeshow chi-square test and Nagelkerke pseudo-R square assessed model fit. We report \( p \) values and odds ratios for all questions in the model at the last step.

Finally, we combined the remaining questions into one diagnostic scale and computed a receiver operating characteristic (ROC) curve to compare the sensitivity and specificity of the scale using different cut-off values. To analyze if age influences whether patients are categorized correctly, we used the Mann–Whitney U test. In addition, we also calculated the positive and negative predictive value and the odds ratio.

Categorical variables are reported as frequencies and continuous variables as means and standard deviations.

We analyzed the data at the University Hospital Zurich using SPSS version 26 (IBM) [21] and set the significance level at 0.05. Missing data are indicated as not reported (n.r.).

3. Results
3.1. Descriptive Statistics

Of 216 contacted patients, the addresses of 14 patients changed since the last visit and were unknown. Of the remaining 202 patients, 85 enrolled in the study (85/202, 42%). Unfortunately, we had to exclude eleven of them as their precise headache diagnosis had remained unclear, e.g., because they had not shown up for follow-ups. Of the remaining 74 participants, 45 suffered from CH (45/74, 60%). The average ages of patients with CH and the controls were 49 ± 16 years and 46 ± 15 years, respectively. Of the patients with CH, 16 were female (16/45, 36%); in the control group, 16 were female (16/29, 55%).

Of the patients included as controls, 17 suffered from migraine and three had a TAC other than CH. Two patients each had occipital neuralgia, a posttraumatic headache, and a primary stabbing headache. Moreover, one patient each had trigeminal neuralgia, nummular headache, and tension-type headache.

First, we analyzed whether patients’ replies to each question distinguished patients with CH from the controls. Table 1 lists the results.

<table>
<thead>
<tr>
<th>No.</th>
<th>Question</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>( p ) Value</th>
<th>Odds Ratio</th>
<th>n.r.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“My headaches are usually excruciating.”</td>
<td>0.91</td>
<td>0.41</td>
<td>0.001</td>
<td>7.24</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>“My headaches are always strictly unilateral.”</td>
<td>0.93</td>
<td>0.32</td>
<td>0.008</td>
<td>6.47</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>“My headaches always appear in one eye or close to it.”</td>
<td>0.93</td>
<td>0.31</td>
<td>0.009</td>
<td>6.30</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>“During the headache attack, the eye on the side of the pain reddens.”</td>
<td>0.73</td>
<td>0.680</td>
<td>0.001</td>
<td>5.81</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>“During the headache attack, the space between the margins of the eyelids (palpebral fissure) narrows.”</td>
<td>0.82</td>
<td>0.46</td>
<td>0.016</td>
<td>3.90</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>“During the headache attacks, my eye tears on the side of the pain.”</td>
<td>0.84</td>
<td>0.55</td>
<td>0.001</td>
<td>6.51</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>“My nose is already blocked at the beginning of a headache attack.”</td>
<td>0.61</td>
<td>0.75</td>
<td>0.004</td>
<td>4.77</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>“My nose already runs at the beginning of a headache attack.”</td>
<td>0.46</td>
<td>0.72</td>
<td>0.146</td>
<td>2.19</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>“During the headache attacks, I cannot lie or sit still.”</td>
<td>0.84</td>
<td>0.63</td>
<td>&lt;0.001</td>
<td>9.23</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>“During the headache attacks, I am always restless.”</td>
<td>0.93</td>
<td>0.38</td>
<td>0.002</td>
<td>8.15</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>“If I do not take medication, my headache attacks last less than three hours.”</td>
<td>0.71</td>
<td>0.74</td>
<td>&lt;0.001</td>
<td>6.91</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>“One headache attack rarely comes alone. When I have one headache attack, the next one often occurs on the same day or at the latest on the next day.”</td>
<td>0.96</td>
<td>0.14</td>
<td>0.202</td>
<td>3.44</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>No.</th>
<th>Question</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>p Value</th>
<th>Odds Ratio</th>
<th>n.r.</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>“My headache attacks usually occur around the same time of the day.”</td>
<td>0.71</td>
<td>0.64</td>
<td>0.004</td>
<td>4.43</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>“Very often do I get headache attacks during the night.”</td>
<td>0.64</td>
<td>0.78</td>
<td>0.001</td>
<td>6.13</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>“Headache attacks wake me up if I am asleep at their onset.”</td>
<td>0.93</td>
<td>0.39</td>
<td>0.001</td>
<td>9.06</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>“I have already wished to die during a headache attack.”</td>
<td>0.62</td>
<td>0.63</td>
<td>0.052</td>
<td>2.80</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>“I am afraid of getting further headache attacks.”</td>
<td>0.93</td>
<td>0.39</td>
<td>0.002</td>
<td>8.63</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>“At night, I hesitate to go to bed because I do not want to be woken up by a headache attack.”</td>
<td>0.48</td>
<td>0.71</td>
<td>0.141</td>
<td>2.28</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>“When my headache attacks appear, I avoid drinking alcohol because it could trigger attacks.”</td>
<td>0.71</td>
<td>0.46</td>
<td>0.193</td>
<td>2.14</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>“When my headache attacks appear, I usually do not leave home without painkillers so that I can treat myself quickly.”</td>
<td>0.91</td>
<td>0.18</td>
<td>0.304</td>
<td>2.12</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>“During the attacks, I lie down in a darkened room.”</td>
<td>0.64</td>
<td>0.35</td>
<td>&gt;0.999</td>
<td>1.09</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>“I am or used to be a smoker.”</td>
<td>0.80</td>
<td>0.64</td>
<td>&lt;0.001</td>
<td>7.00</td>
<td>2</td>
</tr>
</tbody>
</table>

3.2. Logistic Regression Results

Next, we included all questions into the binary regression model if the chi-squared test had reported a p value below 0.20. The group allocation (CH vs. others) was the dependent variable. In 19 steps, the stepwise backward elimination process excluded all variables except five (questions 6, 9, 11, 15 and 22). The Hosmer and Lemeshow test indicated a good model fit (p = 0.976); Nagelkerke pseudo-R-square was 0.759 after the last step. Overall, 35/36 CH patients (97%) and 19/22 controls (86%) were correctly classified. The data of 16 patients (16/74, 22%) were not included in the model, as they had not answered all relevant questions. Table 2 summarizes the remaining variables.

Table 2. Results of the binary logistic regression model; list of the dependent variables that were not removed during the backward elimination process; CI—confidence interval.

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>p Value</th>
<th>Coefficient</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>“During the headache attacks, my eye tears on the side of the pain.”</td>
<td>0.013</td>
<td>19.1</td>
<td>1.7–195.8</td>
</tr>
<tr>
<td>9</td>
<td>“During the headache attacks, I cannot lie or sit still.”</td>
<td>0.034</td>
<td>8.9</td>
<td>1.2–65.5</td>
</tr>
<tr>
<td>11</td>
<td>“If I do not take medication, my headache attacks last less than three hours.”</td>
<td>0.028</td>
<td>9.4</td>
<td>1.3–69.0</td>
</tr>
<tr>
<td>15</td>
<td>“Headache attacks wake me up if I am asleep at their onset.”</td>
<td>0.006</td>
<td>56.6</td>
<td>3.1–1028.0</td>
</tr>
<tr>
<td>22</td>
<td>“I am or used to be a smoker.”</td>
<td>0.046</td>
<td>10.5</td>
<td>1.0–105.4</td>
</tr>
</tbody>
</table>

3.3. Development and Testing of the Scale

We combined all five questions into a clinical scale and attributed equal weight to each one; summing the scores of the individual questions (alloting the value 1 to a “yes” and 0 to a “no”) calculates its total score.

Next, we plotted a ROC curve to estimate the ideal cut-off value associated with high sensitivity. The area under the curve was 0.938 (see Figure 1); the data of eight participants (8/74; 11%) did not enter this analysis due to missing data.

The sensitivity was 98%, and the specificity was 65% with a cut-off value of three (i.e., scores of three and above indicate a CH). The positive and negative predictive values were 81% and 94%, respectively. Patients with three or more points were significantly more likely to suffer from CH than those who scored less (p < 0.001, OR = 73.67, 95%-CI 8.64–628.05). Age had no statistical significant influence on the correctness of the classification (p = 0.485)—see Figure 2 for further details.
Proportion of correctly classified participants, 1

Figure 1. Receiver operating characteristic (ROC) curve of the scale obtained when predicting the group allocation (cluster headache vs controls). The scale comprises items 6, 9, 11, 15 and 22 as diagnostic questions; its area under the curve was 0.938.

Figure 2. Proportion of participants classified correctly by the SMARTED scale per age group. We set the cut-off value at three (i.e., scores of three and above indicate a cluster headache).

4. Discussion

4.1. Summary and Contributions

This study created a diagnostic scale with high sensitivity and reasonable specificity that helps identify CH patients whom a referring physician or neurologist considers a TAC a differential diagnosis.

The 22 analyzed diagnostic questions fell into two categories. The diagnostic criteria published in the International Headache Classification (ICHD-3) informed one group [22]. Examples are attack duration, cranial autonomic features, and restlessness. The second group addressed the burden of disease and triggers. Examples are fear of future attacks, suicidal ideation, and the effect of alcohol.

Most items had good discriminatory power (see Table 1). However, more questions in the first group distinguished CH and its mimics well. Thus, diagnostic criteria effectively separate headache disorders, even in a self-administered questionnaire.

All items with a p value below 0.20 entered the binary regression model—including question 11, which several participants had skipped. That item addresses the duration of
untreated attacks. Not answering may have been due to these patients not having untreated attacks. We retained the item as we aimed to develop a questionnaire for undiagnosed patients without specific treatment. However, with these patients in mind, we ought to specify that by “treatment”, we mean triptans, high-flow oxygen, and lidocaine nasal spray when using the scale in the future.

The regression model eliminated all but five diagnostic questions that independently distinguished between CH and its mimics. We combined them into one scale and attributed equal weight to each for simplicity. Thus, every question the patient answered with “yes” added one point to the total score. A ROC curve (see Figure 1) suggested excellent accuracy [23]. A cut-off value of three yields very high sensitivity and reasonable specificity.

Three items of the scale assess diagnostic criteria (restlessness, cranial autonomic symptoms, and attack duration). The two others are being woken up by the pain and smoking, which are prevalent features of CH attacks and the patients, respectively; they are rare in other types of headache [24–27]. Thus, these variables appear well-suited to distinguish CH from other types of pain, indicating strong face validity. We propose the acronym “SMARTED” as a mnemonic for the five items of the scale (smoking, awakening, restlessness, tearing, duration).

Our questionnaire is not the first of its kind. Other authors have pursued similar objectives in different languages [15–17,20]. The main methodologic difference lies in the composition of the control group. Most controls in previously published studies suffered from migraine and tension-type headache, which are not generally difficult to distinguish from CH. However, we took another approach by including controls in whom at least one doctor had included a TAC in the differential diagnosis.

Headache diagnoses lump together different phenotypes [22,28]. For instance, restlessness does not accompany CH attacks in all patients [29]. Thus, unlike psychometric scales that usually comprise multiple questions to reduce random error [30], diagnostic questionnaires for headache disorders must also include questions to accommodate individual variations of the disease. With a cut-off of three out of five, our screening tool allows for individual variation.

However, care must be taken not to define the included phenotypes too loosely, i.e., allow for too much individual variation, as low specificity might be the consequence. Moreover, including too many patients with phenotypes completely unakin to CH could lead to overestimating the specificity. We included controls whose headache phenotypes overlap with CH to prevent that error. Therefore, our scale’s specificity might be considerably higher when used in the general population with more patients whose headache is easy to identify as non-CH. Because of the different compositions of the control groups, the specificity of our scale cannot be compared with other published questionnaires.

A score of three and above on the SMARTED scale implies high odds of a CH; it thus “rules in” the diagnosis. In other words, the scale screens headache patients for the disorder.

As a thorough history generally suffices to diagnose a CH [31], a screening test may seem dispensable. However, surveys indicate a lengthy diagnostic delay in many patients despite the consultation of several physicians [5,6]. This status quo is unsatisfactory because affected patients risk a considerable and potentially irreversible disease burden [1]. Thus, the diagnostic process requires optimization, and standardization may be the key. The present study is the first step in the direction of such optimization.

4.2. Future Work

The sample we used to create and validate the questionnaire is not the sample which we would eventually like to screen for cluster headache. We aim to screen an unselected sample of patients who have not yet consulted a physician. Thus, future prospective research must assess the scales’ accuracy—particularly its specificity—in an undiagnosed and unselected sample and in different age groups. The latter analysis seems important given that our calculations (see Figure 2) suggest a lower proportion of correctly classified patients in the age group between 50 and 60 years. It is important to note that a low
specificity would lead to high numbers of false positives, given the relatively low prevalence of the disorder [8].

Should the questionnaire perform well in an unselected sample, we must identify or create communication channels to contact persons to be screened. Moreover, the consequences of a positive test result need to be specified. Ideally, it enables simplified access to headache specialists.

4.3. Strengths and Limitations

A strength of this study is its methodology, which aimed to assess the discriminatory power of the diagnostic questions under challenging conditions, including a clinically relevant control group with clinical features overlapping those of CH.

There are some limitations. First, our sample was relatively small. As a result, the reported confidence intervals are wide (see Table 2). However, because of the low prevalence of CH, it is challenging to enroll large numbers of patients. In addition, the inclusion criteria for controls were somewhat restrictive, precluding a larger sample. However, as detailed above, we believe this method may be best suited to analyze the discriminatory power of the questions.

Second, the control group comprised many more females than the CH group. One reason for this finding is that headaches other than CH are generally more common in women than men [32]. We decided not to assess the discriminatory power of patients’ sex, as this could have introduced an additional selection bias.

Third, although almost half of the contacted patients participated in the study, we cannot exclude a unit non-response error. Hence, a prospective study is necessary to assess the quality of the SMARTED scale.

Fourth, not many participants suffered from a TAC other than cluster headache. Thus, conclusions about the accuracy of discrimination from these headaches are difficult to draw.

Lastly, our retrospective sampling, or “reversed-flow design”, may have resulted in a lower specificity than a sampling strategy that includes patients before the diagnosis is made would have [33].

5. Conclusions

The SMARTED scale is a valid and accurate tool for screening for CH in patients in whom a referring physician or a neurologist considers a TAC a differential diagnosis. However, more research is required to test its performance in different samples—especially in persons who have not consulted a headache expert yet. Ultimately, we hope to reduce the diagnostic delay, thus alleviating the disease burden for patients and their families.

Author Contributions: H.P. planned the study, analyzed the data, and wrote the manuscript. M.J. invited the participants to participate by mail and telephone. M.S.N., M.S., A.R.G. and S.W. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: HP was funded by the Werner Dessauer Stiftung.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the canton of Zurich, Switzerland. (BASEC 2021-00695) on 01/06/2021.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and analysed during the study and the German version of the diagnostic scale are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.
References


21. IBM Corp. IBM SPSS Statistics for Windows; Version 26.0; IBM Corp.: Armonk, NY, USA, 2019.


27. Barbanti, P.; Aurilia, C.; Dall’Armi, V.; Egeo, G.; Fofi, L.; Bonassi, S. The phenotype of migraine with unilateral cranial autonomic symptoms documents increased peripheral and central trigeminal sensitization. A case series of 737 patients. *Cephalalgia* 2016, 36, 1334–1340. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.