Evaluating the scripts and thresholds of general practitioners to diagnose heart failure in old people

Klaartje Decaluwe, Department of General Practice, KU Leuven (KUL), Leuven, Belgium

Promotor: Bert Vaes Department of General Practice, KU Leuven (KUL), Leuven, Belgium and Institute of Health and Society, Université Catholique de Louvain (UCL), Brussels, Belgium

Co-promotoren: Jan Degryse Department of General Practice, KU Leuven (KUL), Leuven, Belgium and Institute of Health and Society, Université Catholique de Louvain (UCL), Brussels, Belgium

Master of Family Medicine

Masterproef Huisartsgeneeskunde

Evaluating the scripts and thresholds of general practitioners to diagnose heart failure in old people

Klaartje Decaluwe[1,*], Jan Degryse[1,2] and Bert Vaes[1,2]

[1]Department of General Practice, KU Leuven (KUL), Leuven, Belgium

[2]Institute of Health and Society, Université Catholique de Louvain (UCL), Brussels, Belgium

*Corresponding author: Klaartje Decaluwe
Kapucijnenvoer 33, blok j bus 7001, 3000 Leuven, Belgium
T: +32 (0)16 337 468 F: +32 (0)16 337 480
E-mail : klaartje.decaluwe@gmail.com

Number of words: 3796

Number of tables: 5

Number of figures: 1

Abstract

Background

Guidelines for the diagnosis of heart failure propose several algorithms, based on the combination of symptoms, signs and technical investigations. However, it is unclear whether these algorithms are incorporated in the diagnostic 'scripts' clinicians use in every day clinical practice. In this study, we evaluated the scripts and thresholds general practitioners use to diagnose heart failure in patients aged 75 and older.

Methods

Scripts and thresholds of 130 general practitioners in training and 63 experienced trainers were compared by using an online questionnaire based on the same principles as the script concordance test. Two major cases with an open question and nineteen minor cases with closed questions were presented. For the minor cases all respondents were asked to assign a diagnostic power to individual cues. Based on these powers a diagnostic threshold was calculated for each respondent for the two major cases.

Results

Trainers and trainees used the same scripts to diagnose heart failure in the two major cases. Only ~50 % of the participants used natriuretic peptides in their script even though they judged it as the most powerful cue to demonstrate or exclude heart failure. The power respondents gave to several cues differed significantly according to the context in which these cues were presented. In general, the average exclusive power of different cues was lower than the demonstrative power of cues. There was no difference in diagnostic threshold between trainers and trainees.

Conclusion

Young, unexperienced general practitioners use the same scripts as older, more experienced general practitioners to diagnose heart failure in old people. In general, technical investigations were less frequently queried compared to elements of the medical history and

clinical examination. The clinical context had a strong impact on the diagnostic power that was assigned to different cues.

Abstract

Achtergrond

Richtlijnen voor de diagnose van hartfalen stellen verschillende algoritmes voor op basis van het combineren van symptomen, klinische tekenen en technische onderzoeken. Het is echter onduidelijk of deze algoritmes ingebouwd zijn in de diagnostische scripts die dokters gebruiken in hun dagelijkse praktijk. In deze studie werden de scripts en drempels geëvalueerd die huisartsen gebruiken om hartfalen te diagnosticeren bij patiënten van 75 jaar en ouder.

Methode

De scripts en drempels van 130 huisartsen in opleiding en 63 ervaren huisarts-opleiders werden vergeleken door middel van een online bevraging op basis van de script concordance test. Twee grote casussen met een open bevraging en negentien kleine casussen met gesloten bevraging kwamen aan bod. Aan alle respondenten werd gevraagd om in de kleine casussen aan de individuele argumenten een diagnostische kracht toe te kennen. Met behulp van deze krachten werd de diagnostische drempel berekend voor elke respondent voor de twee grote casussen.

Resultaten

Huisarts-opleiders en huisartsen in opleiding gebruikten dezelfde scripts om hartfalen te diagnosticeren in de twee grote casussen. Slechts ±50% van de deelnemers gebruikten natriuretische peptiden in hun script alhoewel zij dit als het meest krachtige argument beoordelen om hartfalen uit te sluiten en aan te tonen. De kracht die de respondenten toekenden aan de verschillende argumenten verschilde significant naargelang de context waarin deze argumenten werden voorgesteld. In het algemeen was de gemiddelde uitsluitende kracht van de verschillende argumenten lager dan de aantonende kracht. Er was geen verschil in diagnostische drempel tussen opleiders en artsen in opleiding.

Conclusie

Jonge, onervaren huisartsen gebruiken dezelfde scripts als oudere, meer ervaren huisartsen om hartfalen te diagnosticeren bij ouderen. In het algemeen werd minder frequent geopteerd voor technische onderzoeken vergeleken met anamnese en klinisch onderzoek. De klinische context had een sterke impact op de diagnostische kracht die werd toegekend aan de verschillende argumenten.

What this paper adds

What is already known on this subject?

Previous studies attempted to evaluate the diagnostic reasoning of general practitioners when diagnosing heart failure. They concluded that there was no difference between young and more experienced clinicians and that clinical signs, symptoms and technical investigation were not used to the extent suggested in the guidelines.

What this study adds

This study was the first to investigate scripts and thresholds of general practitioners for the diagnosis of heart failure in old people. The clinician's judgment on the diagnostic power of individual clinical signs, symptoms and technical investigations was quantified in a direct way and overall individual diagnostic thresholds were calculated. The clinical context had a strong impact on the diagnostic power that was assigned to different cues. In general, the assigned exclusive power of different cues was lower than the assigned demonstrative power of cues.

Introduction

In our ageing society, heart failure is becoming more frequent. The prevalence increases with age from 0.8% between the age of 55 and 64, to 3% between 65 and 74, to 10% between 75 and 84 and up to 20% in people aged 85 and older (1,2). Heart failure has a poor prognosis: 30 – 40% of the patients diagnosed with heart failure die within the next year (1,3,4). Five years after the incidence of heart failure only 35% of the participants are still alive (1,5).

Heart failure is defined as the presence of three conditions: first, typical symptoms of heart failure such as dyspnoea, oedema or fatigue; second, typical signs of heart failure such as elevated jugular venous pressure, a third heart sound or a displaced apex beat; and third, the presence of structural heart disease or systolic or diastolic dysfunction of the heart (2,3,6). The several guidelines suggest different algorithms, using a combination of reported symptoms, clinical signs and technical investigations such as blood tests, chest X-ray, electrocardiography and echocardiography, to diagnose heart failure (2,3,6-10).

In young patients the diagnosis of heart failure is relatively straightforward. However, heart failure is more common in the old and the specificity of signs and symptoms decreases drastically in the old (11,12). Furthermore, the diagnosis of heart failure in older patients is often complicated by the presence of multiple comorbidities and the presence of polypharmacy. Two recent studies investigated the diagnostic value of signs and symptoms and additional technical examinations, like natriuretic peptides, for the diagnosis of new onset heart failure in old primary care patients and developed both a clinical decision rule to help clinicians (13,14). Also the MICE decision rule (male, myocardial infarction, crepitations and edemas) developed by Mant et al., based on a large meta-analysis of existing literature, could also be applied on patients aged 75 and older (10,15).

However, it is unclear whether these algorithms and clinical decision rules are present in the diagnostic ‘scripts’ clinicians use in every day clinical practice. The diagnostic process in every clinician is a result of scripts and thresholds. Scripts are networks of organized knowledge that are acquired and accumulated during clinical training and are refined with each clinical encounter. Clinicians mobilize these scripts to process information and progress towards a solution for difficult clinical problems (16-18). A diagnostic threshold is a certainty that has to be reached to accept or reject a diagnose. Scripts and thresholds are known concepts in clinical reasoning (19-23). To date, there are a few studies that investigated the clinical reasoning behind diagnosing heart failure in general practitioners (24-28). Moreover, previous studies did not quantify the importance of each cue in the diagnostic process. Furthermore, thresholds have not been evaluated in these studies.

Therefore, this study was conducted to investigate the scripts and thresholds general practitioners use to diagnose heart failure in patients aged 75 and older. Scripts and thresholds of unexperienced general practitioners in training and their experienced trainers were compared.

Methods

Participants

In this study early career physicians and older, more experienced doctors were compared. The early career physicians were students who finished their master degree in medical science and were in their second or third year of the specialty training program in family medicine. This group is supposed to have an up-to-date theoretical knowledge and background. The older, more experienced doctors where general practitioners responsible for the training of these trainees. They acquired experience and practical skills as a general practitioner for at least 5 years. In total, 436 trainees and their trainers were contacted by mail to complete an online questionnaire. At the beginning of the questionnaire each respondent registered their gender, if they were trainer or trainee, the years they already worked as a general practitioner and indicated how familiar they felt with the diagnosis of heart failure in clinical practice (Likert scale, 1 (low) to 10 (high)). The local ethical committee approved the study.

Questionnaire

An online questionnaire was composed, based on the principles of the script concordance test (18) with two major case vignettes with an open question and 19 minor case vignettes with closed questions (see appendix for example). All cases presented patients aged 75 and older with the possibility of having heart failure. In the two major case vignettes a video interview and text was used to sketch the case (the stimulus format). Next, every participant was offered the opportunity to see the information on 32 possible symptoms, signs and technical investigations (the response format). They were asked to click on the presented cues until they reached enough certainty about the presence or absence of heart failure. Afterwards they were asked to indicate the degree of certainty on their diagnosis (Likert scale 1 (low) to 10 (high)). Each click and the sequence of clicks for every respondent was

registered. The 32 cues were selected based on guidelines and previous epidemiological research (2,6,10,13-15,29). The first case was a patient with a high probability of heart failure, the second case a patient with a low probability of heart failure.

In the 19 minor cases the power the respondents gave to individual symptoms, signs and technical investigations was explored. For every case two cues were individually offered. For each cue the respondent was asked to indicate how the presence and the absence of the cue influenced the presence of heart failure in the presented case (visual scale with 'almost excluded' at one end and 'almost certain' at the other end, Likert scale 1 to 7). The cues were chosen based on the clinical decision rules from Kelder, Oudejans and the MICE rule for diagnosing heart failure (13,14,10,15). In total, 16 different cues were evaluated by each respondent. Most cues were offered multiple times in different cases. These 16 cues were also used in the two major cases.

Data analysis

All the data were converted to a numeric value and saved in an excel file. The frequency with which each cue was chosen was registered and compared between trainers and trainees (χ^2 test), for the two major cases separately. The rank order of each cue in the diagnostic process was compared between both groups with the Mann-Whitney U test.

For the 19 minor cases, the power of the presence and absence of each cue ranged between 1 and 7 and was converted to: -3 = heart failure almost excluded, -2 = heart failure most unlikely, -1 = heart failure probably unlikely, 0 = fifty percent chance that heart failure is present or absent, +1 = can be heart failure, +2 = probably heart failure and +3 = heart failure is almost certain. When the diagnostic power of a cue was scored two or more times in a different context the Friedman test or the Wilcoxon matched-pair signed-rank test was used to evaluate whether a significant difference existed. Subsequently, for each respondent the average diagnostic power for each cue was calculated. The average power for each cue was

then compared between trainers and trainees. The frequency each cue was used as an exclusive cue for heart failure was also registered.

Based on the average power of each cue for each individual respondent a diagnostic threshold for the two major cases was calculated. This threshold was the sum of the individual diagnostic powers of the cues each respondent used in his own diagnostic process. A distinction was made between the cues that were used to exclude heart failure and the cues used to demonstrate heart failure by each respondent. The difference between the sum of the exclusive powers and the sum of the demonstrating powers for each respondent was used as the individual diagnostic threshold. This was a 'proxy' threshold since a diagnostic power was not calculated for all the 32 possible cues in the two major cases, but a selection of 16 cues as described above. The Mann-Whitney U test was used to compare the diagnostic threshold between trainers and trainees. A linear regression analysis was performed to define the determinants of the diagnostic threshold. In order to avoid co-linearity, the correlation coefficients between all covariates were calculated. In case of co-linearity (r-value > 0.90), only one of the two covariates was considered in the multivariable model.

The statistical analyses were performed using SPSS 22.0 (SPSS Inc., Chicago, IL, USA).

Results

Participants

In total, 436 trainees and their trainers were invited. The response rate was 22.1% (n = 193) with 130 respondents (33.5%) in the trainees group and 63 respondents (14%) in the trainers group. The online questionnaire was answered by 28 men (21.5%) and 102 women (78.5%) in the trainees group and by 44 men (69.8%) and 19 women (30.2%) in the trainers group. Trainers had a mean clinical experience of 26.0±9.5 years and trainees had a maximum of 2 years clinical experience. Trainers gave themselves a median score of 7 [IQR 6 – 9] for experience with diagnosing heart failure, trainees gave themselves a median score of 4 [IQR 3 – 5] ($P < 0.001$).

Major cases

The first case, with a high probability of heart failure, was diagnosed positive for heart failure by 187 respondents. The remaining six (three trainees and three trainers, (χ^2, $P = 0.36$)) judged that it was not heart failure. Overall, the trainers were more certain of their diagnosis than the trainees (8 [IQR 6 – 9] vs 7 [IQR 6 – 8], $P = 0.016$). In order to reach their diagnosis trainees used a median of 12 [IQR 9 – 19] cues and the trainers 14 [10 – 19] cues ($P = 0.32$). Table 1 lists the frequency the cues were chosen with and the median rank order of each cue in the scripts that were applied. Only the item ‘dyspnea at rest’ was chosen significantly more by the trainees than by the trainers, although the trainers chose this item faster in their diagnostic process than the trainees. For three other cues (weight loss, vertigo and abdominal ultrasound) a trend ($P < 0.10$) was seen towards faster usage of these items by the trainers compared with the trainees. However, the first five items that were most frequently used did not differ between trainees and trainers: orthopnea (n = 104 (81%) and n = 51 (81%)), dyspnea at rest (n = 79 (61%) and n = 32 (51%)), history of edema (n = 72 (56%) and n = 44 (70%)), dyspnea on exertion (n = 66 (51%) and n = 29 (46%)) and coughing (n = 47 (36%)

and n = 28 (44%)). In total 41 (21.2%) respondents (31 trainees and 10 trainers (χ^2, P = 0.20) did not choose any technical investigation. Only 105 (54.4%) respondents wanted to know the level of the natriuretic peptides and 87 (45.0%) the results of the electrocardiography.

The second case, designed with a low probability of heart failure, was diagnosed negative for heart failure by 179 respondents. The remaining 14 (nine trainees and five trainers (χ^2, P = 0.80)) judged that it was heart failure. No difference in diagnostic certainty was found between trainers and trainees (median 7 [IQR 6 – 9] vs 7 [IQR 6 – 8], P = 0.33). Trainees used a median of 14 (IQR [9 – 20]) cues and trainers a median of 16 (IQR [9 – 23]) cues to reach their diagnosis (P = 0.38). Table 2 lists the frequency and the rank order of each item. Trainers chose the items 'weight loss' and 'chest X-ray' significantly more often than trainees. A trend (P < 0.10) in different frequencies between trainees and trainers was observed for three items: orthopnea, peripheral vessels and abdominal examination. The rank order of the different cues in the individual diagnostic scripts was not different between trainees and trainers. The first five items that were most frequently used did not differ between trainees and trainers and were the same as for case 1: orthopnea (n = 109 (84%) and n = 45 (71%)), dyspnea on exertion (n = 90 (70%) and n = 37 (62%)), dyspnea at rest (n = 83 (64%) and n = 43 (68%)), coughing (n = 60 (47%) and n = 38 (60%)) and history of edema (n = 50 (39%) and n = 19 (30%)). Also 'retrosternal pain' was chosen as often as 'history of oedema' by the trainers. In total 26 (13.5%) respondents (20 trainees and 6 trainers (χ^2, P = 0.26) did not choose any technical investigation. Only 106 (54.9%) respondents wanted to know the level of the natriuretic peptides and 93 (48.2%) the results of the electrocardiography.

Minor cases

The diagnostic power of 16 different cues was measured for each respondent by presenting 19 minor cases. The power respondents gave to several cues differed significantly

according to the context in which this cue was presented (Table 3). A difference in diagnostic power was thus found for both trainees and trainers for a normal lung auscultation, basal crackles, an increased jugular venous pressure, an abnormal chest X-ray and a regular pulse rate; for trainees alone for the presence of wheezing, orthopnea, a displaced apex beat, the presence of oedema, an abnormal electrocardiography and a decreased appetite; and for trainers alone for a normal level of NT-proBNP, the absence of dyspnea and an irregular pulse rate.

Table 4 shows the average diagnostic power of each cue. The average powers were compared between trainees and trainers and a significant difference was found for wheezing on lung auscultation, normal jugular venous pressure, presence of nocturnal paroxysmal dyspnea, normal apex beat, displaced apex beat, irregular pulse rate and decreased appetite. A normal level of NT-proBNP and the absence of dyspnea were scored as the strongest cues to exclude heart failure, while an increased level of NT-proBNP, the presence of orthopnea and the history of a myocardial infarction were chosen as the strongest cues to demonstrate heart failure. In general the average exclusive power of different cues was lower than the demonstrative power of cues that were presented.

Body mass index (BMI) and pulse rate were offered several times with different cut-off values: the diagnostic powers of a BMI of 28kg/m², 24.8kg/m², 19.5kg/m² and 18kg/m² were not different between trainees and trainers, and were on average scored as a neutral cue (average power = 0); the diagnostic powers of a pulse rate of 66/min and 109/min were scored different between trainees and trainers (0 [IQR 0 – 0] vs 0 [-1.0 – 0], $P = 0.001$ and 1.0 [0 – 1.0] vs 1.0 [1.0 – 2.0], $P < 0.001$, respectively).

Diagnostic threshold

Figure 1 shows the exclusive power and demonstrative power and the proxy diagnostic threshold for both major cases. For the first case the diagnostic threshold ranged between -

0.17 and 23.2, with a median threshold of 8.5 [IQR 5.8 – 11.6] for the trainees and 9.2 [IQR 6.0 – 11.7] for the trainers (P = 0.54). For the second case the diagnostic threshold ranged between -12.8 and 8.8, with a median threshold of 1.2 [IQR -1.2 – 2.9] for the trainees and 0.92 [IQR -2.5 – 4.0] for the trainers (P = 0.71). The correlation coefficient between the calculated diagnostic threshold and the diagnostic certainty every respondent had to score was 0.24 (P = 0.001) for the first case and -0.14 (P = 0.054) for the second case.

Table 5 shows the determinants of the diagnostic threshold for case 1 and 2. For the first case the number of cues used and the diagnostic certainty of the respondent were found to be independently correlated with the diagnostic threshold. For the second case male gender and the number of cues used were identified as independent predictors.

Discussion

This study, based on the same principles as the script concordance test, evaluated scripts and thresholds of general practitioners to diagnose heart failure in old people. Differences between early career physicians and experienced general practitioners were investigated. No differences were identified between trainees and trainers in the scripts that were used. In general, technical investigations were less frequently queried compared to elements of the medical history and clinical examination. The power respondents in both groups gave to several cues differed significantly according to the context in which this cue was presented. Furthermore, the assigned exclusive power of different cues was lower than the assigned demonstrative power of the cues that were presented. The proxy diagnostic threshold was the same for trainers and trainees, but the variation between individual participants was large.

The hypothesis that more experienced clinicians use clinical information in a more targeted way, use less cues to reach diagnostic certainty and have other scripts than less experienced clinicians could not be confirmed. Previous literature about scripts stated that scripts change when more experience is gained (16,17). A possible explanation for the discrepancy with the current findings could be that the trainers filled out the questionnaire as a teacher rather than a clinician, and thus clicked more cues than they would in real life. On the other hand, other studies that examined the diagnosis of a difficult clinical problem, e.g. heart failure or pulmonary embolism, support the conclusion that there is no difference between the diagnostic strategies used by medical students and more experienced clinicians, but the variation between individual participants is large (25,26,30). Possibly the expected evolution in scripts when gaining clinical experience from a theoretical point of view, is not represented in real life clinical practice for severe diagnoses. Future research using an incomplete format design could be elaborated to further investigate the difference in diagnostic scripts based on

clinical experience. In heart failure guidelines (2,3,6-9) a normal electrocardiography is incorporated as a good test to exclude heart failure. The current study however, showed respondents attributed little excluding power to this test and only half of the respondents had this test included in their script. The overall observation in the current study that technical investigations were little used, was also seen in previous research. Skånér et al. observed the same trend where no information about echocardiography was utilized by a third of the general practitioners in the presented situations and only 20% used the information about a normal electrocardiography as a cue to exclude heart failure (28). Furthermore, Skånér et al. found general practitioners estimated information on the presence of dyspnea, a history of myocardial infarction and the enlargement of the heart influenced their judgment on the presence of heart failure the most (25). In the current study the same trend was observed but the respondents also individually assessed the diagnostic power of each cue. A normal level of NT-proBNP and the absence of dyspnea were scored as the strongest cues to exclude heart failure, and an increased level of NT-proBNP, the presence of orthopnea and the history of a myocardial infarction were chosen as the strongest cues to demonstrate heart failure. The observation that NT-proBNP was present in only ~50% of the diagnostic strategies in spite of the assigned strong diagnostic power could be explained by the fact that these tests are not reimbursed in Belgium, although these tests were incorporated in the national guidelines of 2011. Skånér et al. did not examine NT-proBNP because it was not available at that time.

The observation that general practitioners gave different powers to the same cues in a different context is an addition to existing epidemiological data in which only one likelihood ratio for each cue is calculated. These findings could be explained by the fact that epidemiological powers are measurements whereas the current study concerns clinical judgments based on the integration of rich contextual information. Thus, in everyday clinical

practice general practitioners do not use one fixed diagnostic power for different cues. This has to be taken into account when estimating diagnostic thresholds.

In a low prevalence context, like general practice, diagnostic strategies are in general more focused on finding cues that exclude a diagnosis rather than cues that demonstrate the diagnosis. In the current study however, the diagnostic power of different cues to exclude heart failure was on average weighted less than the diagnostic power to demonstrate heart failure. Moreover, based on the number of cues that was used in both cases, it seemed that respondents in general kept looking for cues that demonstrated heart failure. This could possibly be explained by the fact that general practitioners are used to look at the patient in a holistic way and think less in terms of algorithms. Furthermore, clinicians are probably aware of the decreasing sensitivity and specificity of signs and symptoms for heart failure in the old. Moreover, the proposed clinical decision rules for the diagnosis of heart failure in the old are also more oriented towards summing up demonstrative cues rather than exclusive cues and thus pushing clinicians rather in the direction of demonstrating heart failure (13,14). On the other hand, the current findings also call for further implementation of existing strong excluders, like natriuretic peptides, in daily practice and further education of clinicians in using these technical investigations for the diagnosis of heart failure.

Strengths and limitations

This study is the first that evaluated scripts and thresholds of general practitioners to diagnose heart failure in old people. The strength of this study lies in the fact that each respondent was asked to score the diagnostic power of individual cues in a direct way compared to previous research that only deducted the weighing of each cue depending on the answers about the presence of heart failure. Furthermore, this study is the first that calculated a diagnostic threshold for the presence or the absence of heart failure.

A few limitations should be considered. First, the cases in the questionnaire were all fictive, ‘paper’ cases with the possibility of heart failure. It is possible that these cases were less representative for real life patients, but they all had a high ‘face validity’. Furthermore, it was also important for this study to examine the strategy of different doctors to a same stimulus rather than to examine their reactions in real situations. Second, the response rate of 22% is mediocre but this study was designed to compare diagnostic strategies between young and more experienced clinicians and was not designed to be representative for the Flemish general practitioner. Moreover, the sampling of experienced general practitioners may have been a potential problem, since general practitioners with a special interest in training and education were selected, probably making this group more homogeneous than a random sample of experienced general practitioners. Third, the diagnostic threshold was only a proxy threshold since not all 32 cues from the major cases were used, although the 16 most important cues in literature were included.

Conclusion

Young and unexperienced general practitioners used the same diagnostic reasoning and scripts as older, more experienced general practitioners to diagnose heart failure in old people. In general, technical investigations were less frequently queried compared to elements of the medical history and clinical examination. The power respondents in both groups gave to several cues differed significantly according to the context in which this cue was presented. Furthermore, the assigned exclusive power of different cues was lower than the assigned demonstrative power of the cues that were presented.

Acknowledgements

We like to thank all participating trainers and trainees for filling in the questionnaire. I'd like to thank my promoter, Bert Vaes, and Co-promoter, Jan Degryse, for the help with this manamathesis. Also many thanks to my colleagues in the office at Hoeilaart for the two interesting and instructive years as a trainee.

References

1. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. Heart. 2007; 93: 1137-1146.
2. Hoes AW, Voors AA, Rutten FH et al. Richtlijn hartfalen Nederlands Huisartsen Genootschap (NHG). Huisarts Wet 2010; 53: 368-89. www.nhg.org
3. National Clinical Guideline Centre (UK). Chronic Heart Failure. National clinical guideline for diagnosis and management in primary and secondary care. NICE clinical guidelines, No. 108. 2010.
4. Hobbs FD, Roalfe AK, Davis RC et al. Prognosis of all-cause heart failure and borderline left ventricular systolic dysfunction: 5 year mortality follow-up of the Echocardiographic Heart of England Screening Study (ECHOES). Eur Heart J. 2007; 28(9): 1128-34.
5. Lloyd-Jones DM, Larson MG, Leip EP et al. Lifetime risk for developing congestive heart failure: the Framingham Heart Study. Circulation. 2002; 106(24): 3068-72.
6. McMurray J, Adamopoulos S, Anker Stefan et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur Heart J 2012; 14: 803–869.
7. Yancy C W, Jessup M, Bozkurt B et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2013; 128: 240-327.
8. SIGN Scottish Intercollegiate Guidelines Network. Management of chronic heart failure. A national clinical guideline. 2007.
9. Malcolm A, Liu P, Demers C et al. Canadian Cardiovascular Society consensus conference recommendations on heart failure 2006: Diagnosis and management. Can J Cardiol. 2006; 22: 23–45.

10. Mant J, Doust J, Roalfe A et al. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. Health Technol Assess. 2009; 13: 1-207.
11. Fonseca C, Morais H, Mota T et al. The diagnosis of heart failure in primary care: value of symptoms and signs. Eur J Heart Fail 2004; 6: 795-2.
12. Olofsson M, Edebro D, Boman K. Are elderly patients with suspected HF misdiagnosed? A primary health care center study. Cardiology. 2007; 107: 226-32.
13. Kelder JC, Cramer MJ, van Wijngaarden J et al. The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. Clinical perspective. Circulation. 2011; 124: 2865-73.
14. Oudejans I, Mosterd A, Bloemen JA et al. Clinical evaluation of geriatric outpatients with suspected heart failure: value of symptoms, signs, and additional tests. Eur J Heart Fail. 2011; 13: 518-27.
15. Roalfe AK, Mant J, Doust JA et al. Development and initial validation of a simple clinical decision tool to predict the presence of heart failure in primary care: the MICE (Male, Infarction, Crepitations, Edema) rule. Eur J Heart Fail. 2012; 14: 1000-08.
16. Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. Acad. Med. 2000; 75: 182-90.
17. Charlin B, Boshuizen HP, Custers EJ et al. Scripts and clinical reasoning. Med Educ. 2007; 41: 1178-84.
18. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. BMC Med Inform Decis Mak. 2008; 8: 18-25.
19. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. N Engl J Med. 1975; 293: 229-34.

20. Pauker SG, Kassirer JP. The approach to clinical decision making. N Engl J Med. 1980; 302: 1109-17.
21. Van Puymbroeck H, Boeckx J, Blanckaert F et al. Klinische logica (deel2): Van diagnostisch landschap tot diagnose. Huisarts Nu. 2004; 33: 201-14.
22. Stolper CF, Rutten ALB, Dinant GJ. Hoe verloopt het diagnostisch denken van de ervaren huisarts? Huisarts Wet. 2005; 48: 993-7.
23. Hobus P. Expertise van huisartsen: praktijkervaring, kennis en diagnostische hypothesevorming. Universiteit van Maastricht.1994.
24. Hancock HC, Mason JM, Murphy JJ. Using the method of judgement analysis to address variations in diagnostic decision making. BMC Res Notes. 2012; 5: 139-45.
25. Skånér Y, Strender LE, Bring J. How do GPs use clinical information in their judgements of heart failure? A clinical judgement analysis study. Scand J Prim Health Care. 1998; 16: 95-100.
26. Skånér Y, Bring J, Ullman B et al. The use of clinical information in diagnosing chronic heart failure: a comparison between general practitioners, cardiologists, and students. J Clin Epidemiol. 2000; 53: 1081-8.
27. Skånér Y, Bring J, Ullman B et al. Heart failure diagnosis in primary health care: clinical characteristics of problematic patients. A clinical judgement analysis study. BMC Fam Pract. 2003; 4: 12-20.
28. Skånér Y, Backlund L, Montgomery H et al. General practitioners' reasoning when considering the diagnosis heart failure: a think-aloud study. BMC Fam Pract. 2005; 6: 4-14.
29. Vaes B, Boland B, Scavée C et al. Value of adding natriuretic peptides and electrocardiographic findings to assess the presence of cardiac dysfunction in patients ≥80 years of age. Am J Cardiol. 2013; 111: 1198-208.

30. Wigton RS, Hoellerich VL, Patil KD. How physicians use clinical information in diagnosing pulmonary embolism: an application of conjoint analysis. Med Decis Making. 1986; 6: 2-11.

Table 1. The frequency cues were used with and the median rank order of each cue in the scripts that were applied for case 1.

	Trainee			Trainer		
	Number of participants that clicked on the cue	Place of the cue		Number of participants that clicked on the cue	Place of the cue	
	n (%)	Range	Median [IQR]	n (%)	Range	Median [IQR]
History						
Orthopnoea	107 (82.3)	1 – 12	2 [1 – 3]	49 (77.8)	1 – 4	2 [2 – 3]
History of edema	106 (81.5)	1 – 22	4 [2 – 6]	53 (84.1)	1 – 13	3 [2 – 5]
Dyspnea at rest	88 (67.7)	1 – 16	2 [1 – 3]	33 (52.4)*	1 – 7	1 [1 – 2]$
Fatigue	73 (56.2)	1 – 21	6 [3 – 8]	37 (58.7)	2 – 17	6 [4 – 8]
Dyspnea on exertion	69 (53.1)	1 – 17	3 [2 – 4]	33 (52.4)	1 – 14	3 [2 – 4]
Coughing	67 (51.5)	2 – 18	5 [3 – 6]	39 (61.9)	1 – 16	4 [3 – 6]
Retrosternal pain	63 (48.5)	1 – 20	5 [4 – 7]	35 (55.6)	1 – 10	5 [3 – 7]
Nocturnal paroxysmal dyspnea	54 (41.5)	2 – 24	9 [5 – 13]	32 (50.8)	3 – 17	7 [5 – 11]
Palpitations	39 (30.0)	3 – 16	10 [8 – 13]	22 (34.9)	3 – 16	9 [6 – 11]

Fever	35 (26.9)	3 – 19	6 [5 – 7]	18 (28.6)	1 – 16	6 [5 – 7]
Syncope	29 (22.3)	3 – 13	11 [7 – 12]	13 (20.6)	4 – 14	8 [5 – 11]
Appetite	28 (21.5)	1 – 12	2 [1 – 5]	14 (22.2)	1 – 11	3 [1 – 6]
Weight loss	25 (19.2)	4 – 15	11 [9 – 13]	15 (23.8)	2 – 15	9 [5 – 10][£]
Vertigo	20 (15.4)	2 - 23	9 [9 – 10]	10 (15.9)	4 – 12	8 [5 – 10][£]
Clinical examination						
Lung auscultation	103 (79.2)	1 – 19	8 [6 – 12]	49 (77.8)	1 – 16	8 [6 – 11]
Edema clinical	97 (74.6)	1 – 22	10 [6 – 13]	46 (73.0)	2 – 19	11 [7 – 13]
Heart auscultation	90 (69.2)	1 – 18	8 [5 – 11]	42 (66.7)	1 – 15	9 [7 – 12]
Jugular venous pressure	89 (68.5)	1 – 20	9 [6 – 13]	38 (60.3)	3 – 17	9 [6 – 11]
Weight	81 (62.3)	2 – 27	12 [8 – 17]	46 (73.0)	4 – 22	13 [9 – 15]
Parameters (e.g. pulse)	71 (54.6)	1 – 21	11 [7 – 16]	32 (50.8)	5 – 18	11 [9 – 15]
Respiratory rate and saturation	55 (42.3)	4 – 23	13 [8 – 19]	32 (50.8)	3 – 20	12 [9 – 17]
Abdominal examination	43 (33.1)	4 – 28	18 [11 – 23]	19 (30.2)	10 – 23	17 [14 – 22]
Peripheral vessels	41 (31.5)	8 – 26	17 [12 – 21]	13 (20.6)	10 – 21	19 [14 – 21]
Apex beat	38 (29.2)	7 – 29	20 [11 – 24]	20 (31.7)	5 – 24	16 [10 – 23]

Neurological examination	26 (20.0)	11 – 30	24 [19 – 26]	13 (20.6)	11 – 25	21 [17 – 25]
Technical investigation						
Chest X-ray	72 (55.4)	3 – 36	15 [10 – 23]	37 (58.7)	3 – 32	14 [11 – 23]
NT-proBNP	67 (51.5)	1 – 33	14 [8 – 23]	38 (60.3)	1 – 29	14 [11 – 20]
Electrocardiography	54 (41.5)	1 – 28	16 [10 – 24]	33 (52.4)	2 – 26	15 [12 – 19]
Complete blood count and serum creatinine	39 (30.0)	7 – 35	17 [13 – 28]	26 (41.3)	9 – 31	19 [15 – 25]
Spirometry	30 (23.1)	9 – 37	24 [16 – 31]	20 (31.7)	1 – 33	21 [14 – 30]
Liver function tests	25 (19.2)	12 – 34	26 [20 – 30]	13 (20.6)	12 – 30	23 [18 – 29]
Abdominal echography	22 (16.9)	8 – 32	26 [18 – 28]	15 (23.8)	9 – 28	19 [17 – 26][£]

[*], χ^2 test, $P < 0.05$; [$], Mann-Whitney U test, $P < 0.05$; [£], Mann-Whitney U test, $P < 0.10$.

Table 2. The frequency cues were used with and the median rank order of each cue in the scripts that were applied for case 2.

	Trainee			Trainer		
	Number of participants that clicked on the cue	Place of the cue		Number of participants that clicked on the cue	Place of the cue	
	n (%)	Range	Median [IQR]	n (%)	Range	Median [IQR]
History						
Orthopnoea	107 (82.3)	1 – 19	2 [2 – 3]	46 (73.0)[µ]	1 – 6	2 [2 – 3]
History of edema	106 (81.5)	1 – 24	5 [3 – 9]	43 (68.3)	1 – 10	6 [4 – 9]
Dyspnea at rest	88 (67.7)	1 – 13	2 [1 – 3]	47 (74.6)	1 – 21	2 [1 – 2]
Fatigue	73 (56.2)	1 – 14	6 [5 – 8]	41 (65.1)	1 – 16	6 [5 – 8]
Dyspnea on exertion	69 (53.1)	1 – 20	3 [2 – 4]	44 (69.8)	1 – 22	3 [2 – 4]
Coughing	67 (51.5)	1 – 24	5 [4 – 6]	44 (69.8)	1 – 9	4 [3 – 5]
Retrosternal pain	63 (48.5)	2 – 23	6 [5 – 7]	35 (55.6)	2 – 8	5 [4 – 7]
Nocturnal paroxysmal dyspnea	54 (41.5)	2 – 27	9 [6 – 14]	32 (50.8)	4 – 14	10 [7 – 14]
Palpitations	39 (30.0)	2 – 26	9 [7 – 13]	26 (41.3)	4 – 13	10 [6 – 13]

Fever	35 (26.9)	2 – 22	6 [5 – 7]	21 (33.3)	3 – 21	6 [5 – 6]
Syncope	29 (22.3)	6 – 14	11 [10 – 12]	12 (19.0)	6 – 11	11 [10 – 11]
Appetite	28 (21.5)	1 – 10	1 [1 – 1]	18 (28.6)	1 – 7	1 [1 – 1]
Weight loss	25 (19.2)	1 – 25	10 [7 – 12]	35 (55.6)*	1 – 19	8 [6 – 12]
Vertigo	20 (15.4)	1 – 28	9 [9 – 11]	14 (22.2)	4 – 23	9 [8 – 9]
Clinical examination						
Lung auscultation	105 (80.8)	1 – 26	9 [6 – 13]	49 (77.8)	2 – 16	9 [7 – 13]
Heart auscultation	98 (75.4)	1 – 27	9 [6 – 13]	44 (69.8)	1 – 15	10 [7 – 14]
Edema clinical	88 (67.7)	1 – 24	12 [8 – 18]	45 (71.4)	1 – 20	12 [8 – 17]
Peripheral vessels	73 (56.2)	2 – 26	14 [9 – 21]	27 (42.9)µ	2 – 23	16 [13 – 21]
Jugular venous pressure	71 (54.6)	1 – 25	12 [8 – 17]	37 (58.7)	1 – 28	12 [9 – 17]
Parameters (e.g. pulse)	67 (51.5)	3 – 21	12 [9 – 18]	28 (44.4)	6 – 18	13 [11 – 18]
Weight	65 (50.0)	1 – 30	15 [10 – 22]	37 (58.7)	3 – 23	15 [12 – 21]
Respiratory rate and saturation	63 (48.5)	3 – 25	15 [11 – 20]	35 (55.6)	5 - 21	15 [11 – 19]
Abdominal examination	40 (30.8)	1 – 30	21 [15 – 24]	27 (42.9)µ	4 – 25	19 [15 – 23]
Apex beat	35 (26.9)	7 – 31	24 [15 – 25]	22 (34.9)	2 – 26	19 [14 – 24]

Neurological examination	29 (22.3)	8 – 32	25 [17 – 26]	18 (28.6)	7 – 27	23 [17 – 25]
Technical investigation						
Chest X-ray	71 (54.6)	2 – 42	16 [11 – 25]	44 (69.8)*	1 – 32	16 [11 – 25]
NT-proBNP	71 (54.6)	2 – 38	15 [9 – 24]	35 (55.6)	1 – 29	19 [14 – 25]
Electrocardiography	58 (44.6)	5 – 35	19 [12 – 26]	35 (55.6)	6 – 27	18 [12 – 24]
Spirometry	54 (41.5)	4 – 43	18 [12 – 32]	34 (54.0)	3 – 33	21 [13 – 30]
Complete blood count and serum creatinine	49 (37.7)	6 – 40	20 [14 – 30]	27 (42.9)	3 – 31	19 [16 – 28]
Liver function tests	27 (20.8)	10 – 39	29 [22 – 30]	14 (22.2)	4 – 30	28 [24 – 29]
Abdominal echography	26 (20.0)	13 – 36	27 [20 – 28]	15 (23.8)	13 – 28	26 [21 – 27]

*, χ^2 test, $P < 0.05$; $^{\mu}$, χ^2 test, $P < 0.10$.

Table 3. Differences in diagnostic power of the cues in the 19 minor cases based on the clinical context.

	Total group	Trainee	Trainer
Lung auscultation			
Normal*	$P < 0.001$	$P = 0.003$	$P = 0.007$
Basal crackles$	$P < 0.001$	$P < 0.001$	$P = 0.001$
Wheezing$	$P < 0.001$	$P < 0.001$	$P = 0.074$
Dyspnea			
Absent$	$P = 0.059$	$P = 0.51$	$P = 0.020$
Present$	$P = 0.58$	$P = 0.78$	$P = 0.56$
NT-proBNP			
Normal*	$P = 0.002$	$P = 0.098$	$P = 0.007$
Increased*	$P = 0.054$	$P = 0.056$	$P = 0.17$
Heart murmur			
Absent$	$P = 0.51$	$P = 0.70$	$P = 0.13$
Present$	$P = 0.25$	$P = 0.40$	$P = 0.41$
Orthopnoea			

Absent*	*P* = 0.035	*P* = 0.14	*P* = 0.16
Present*	*P* < 0.001	*P* < 0.001	*P* = 0.25
Jugular venous pressure			
Normal*	*P* = 0.68	*P* = 0.43	*P* = 0.15
Increased*	*P* < 0.001	*P* < 0.001	*P* < 0.001
Nocturnal paroxysmal dyspnea			
Absent$	*P* = 0.62	*P* = 0.79	*P* = 0.23
Present$	*P* = 0.43	*P* = 0.96	*P* = 0.17
Apex beat			
Normal$	*P* = 0.80	*P* = 0.85	*P* = 0.65
Displaced$	*P* = 0.038	*P* = 0.041	*P* = 0.41
Edema			
Absent*	*P* = 0.20	*P* = 0.27	*P* = 0.32
Present*	*P* < 0.001	*P* < 0.001	*P* = 0.054
History myocardial infarct			
Absent$	P = 0.69	*P* = 0.70	*P* = 0.90

Present[$]	$P = 0.32$	$P = 0.58$	$P = 0.35$
Electrocardiography			
Normal[$]	$P = 0.024$	$P = 0.063$	$P = 0.20$
Abnormal[$]	$P < 0.001$	$P < 0.001$	$P = 0.69$
Chest X-ray (cardiothoracic index)			
Normal[$]	$P = 0.048$	$P = 0.070$	$P = 0.41$
Abnormal[$]	$P < 0.001$	$P < 0.001$	$P < 0.001$
Pulse rate			
Regular[$]	$P < 0.001$	$P < 0.001$	$P = 0.007$
Irregular[$]	$P = 0.17$	$P = 0.68$	$P = 0.002$
Appetite			
Normal[$]	$P = 0.58$	$P = 0.23$	$P = 0.32$
Decreased[$]	$P < 0.001$	$P < 0.001$	$P = 0.43$

[*], Difference in distribution of the given power between the cases (>2) with the Friedman test; [$], Difference in distribution of the given power between the cases (≤2) with the Wilcoxon matched-pair signed-rank test.

Table 4. Diagnostic power of the individual cues in the 19 minor cases.

	Total group	Trainee		Trainer	
	Average power, median (IQR)	Average power, median (IQR)	Cue used to exclude HF n (%)	Average power, median (IQR)	Cue used to exclude HF n (%)
Lung auscultation					
Normal	-0.5 (-1.0, -0.25)	-0.50 (-1.0, -0.20)	98 (75.4)	-0.75 (-1.0, -0.25)	54 (85.7)
Basal crepitation	1.5 (1.0, 2.0)	1.5 (1.5, 2.0)	1 (0.8)	1.5 (1.0, 2.0)	1 (1.6)
Wheezing	0 (-1.0, 0.50)	-0.50 (-1.0, 0.13)	72 (55.4)	0 (-0.50, 0.50)[£]	17 (27.0)
Dyspnea					
Absent	-1.0 (-1.5, -0.50)	-1.0 (-1.0, -0.50)	114 (87.7)	-1.0 (-1.5, -0.50)	58 (92.1)
Present	1.5 (1.0, 2.0)	1.5 (1.0, 2.0)	2 (1.5)	1.5 (1.0, 2.0)	0 (0)
NT-proBNP					
Normal	-1.3 (-2.3, -0.67)	-1.3 (-2.3, 0.67)	113 (86.9)	-1.3 (-2.7, -0.67)	53 (84.1)
Increased	2.0 (1.3, 2.7)	2.0 (1.3, 2.7)	0 (0)	2.0 (1.3, 2.3)	0 (0)
Heart murmur					
Absent	0 (-0.50, 0)	0 (-0.50, 0)	42 (32.3)	0 (-0.50, 0)	28 (44.4)

Present	1.0 (0.50, 1.5)	1.0 (0.50, 1.5)	2 (1.5)	1.0 (0.50, 1.5)	2 (3.2)
Orthopnea					
Absent	-0.67 (-1.0, -0.33)	-0.67 (-1.0, -0.33)	101 (77.7)	-0.67 (-1.0, -0.33)	50 (79.4)
Present	2.0 (1.7, 2.7)	2.0 (1.7, 2.7)	0 (0)	2.0 (1.7, 2.7)	0 (0)
Jugular venous pressure					
Normal	-0.33 (-1.0, 0)	-0.33 (-0.67, 0)	66 (50.8)	-0.67 (-1.0, 0)[£]	42 (66.7)
Increased	1.7 (1.3, 2.3)	1.7 (1.3, 2.1)	0 (0)	1.7 (1.0, 2.3)	0 (0)
Nocturnal paroxysmal dyspnea					
Absent	-0.5 (-1.0, 0)	-0.50 (-1.0, 0)	84 (64.6)	-0.50 (-1.0, 0)	43 (68.3)
Present	1.0 (1.0, 2.0)	1.0 (0.50, 1.6)	2 (1.5)	1.5 (1.0, 2.0)[£]	1 (1.6)
Apex beat					
Normal	0 (0, 0)	0 (0, 0)	19 (14.6)	0 (-0.50, 0)[£]	19 (30.2)
Displaced	1.0 (1.0, 1.5)	1.0 (1.0, 1.5)	1 (0.8)	1.0 (1.0, 1.5)[£]	0 (0)
Edema					
Absent	-0.75 (-1.0, 0)	-0.50 (-1.0, 0)	96 (73.8)	-0.75 (-1.0, 0)	47 (74.6)
Present	1.5 (1.3, 2.0)	1.8 (1.3, 2.0)	0 (0)	1.5 (1.3, 2.0)	0 (0)

History myocardial infarct					
Absent	-0.50 (-0.75, 0)	-0.50 (-0.63, 0)	71 (54.6)	0 (-1.0, 0)	29 (46.0)
Present	2.0 (1.0, 2.0)	2.0 (1.0, 2.0)	0 (0)	1.5 (1.0, 2.0)	0 (0)
Electrocardiography					
Normal	0 (-1, 0)	0 (-1.0, 0)	61 (46.9)	-0.50 (-1.0, 0)	34 (54.0)
Abnormal	0.50 (0, 1.0)	0.50 (0, 1.0)	4 (3.1)	0.50 (0, 1.0)	5 (7.9)
Chest X-ray (cardiothoracic index)					
Normal	-0.50 (-1.0, 0)	-0.50 (-1.0, 0)	78 (60.0)	-0.50 (-1.0, 0)	39 (61.9)
Abnormal	1.5 (1.0, 2.0)	1.5 (1.0, 2.0)	1 (0.8)	1.5 (1.0, 2.0)	0 (0)
Pulse rate					
Regular	0 (-0.50, 0)	0 (0, 0)	30 (23.1)	0 (-0.50, 0)	21 (33.3)
Irregular	1.0 (1.0, 1.8)	1.0 (1.0, 1.5)	6 (4.6)	1.5 (1.0, 2.0)[£]	0 (0)
Appetite					
Normal	0 (0, 0)	0 (0, 0)	15 (11.5)	0 (0, 0)	11 (17.5)
Decreased	0.50 (0, 1.0)	0.50 (0, 1.0)	10 (7.7)	1.0 (0, 1.0)[£]	3 (4.8)

[£], Significant difference in mean power assigned to the cues in the minor cases between trainees and trainers (Mann-Whitney U test, $P < 0.05$).

Table 5. Determinants of the diagnostic threshold (linear regression analysis).

	Case 1				Case 2			
	Bivariate		Multivariate		Bivariate		Multivariate	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Male gender	-1.1 (-2.4, 0.24)	0.11	-0.66 (-1.9, 0.62)	0.31	-1.3 (-2.3, -0.24)	0.017	-1.5 (-2.7, -0.29)	0.016
Trainer (versus trainee)	0.39 (-1.0, 1.8)	0.59	1.0 (-0.45, 2.5)	0.18	-0.53 (-1.6, 0.57)	0.34	0.32 (-1.1, 1.7)	0.66
Familiarity with heart failure	-0.061 (-0.34, 0.22)	0.67	-0.27 (-0.57, 0.036)	0.084	-0.13 (-0.35, 0.090)	0.24	-0.005 (-0.28, 0.27)	0.97
Number of cues used	0.31 (0.24, 0.37)	< 0.001	0.30 (0.23, 0.36)	< 0.001	-0.068 (-0.13, -0.010)	0.023	-0.071 (-0.13, -0.013)	0.017
Heart failure (yes or no)	4.2 (0.50, 7.0)	0.027	1.9 (-1.3, 5.0)	0.24	1.1 (-0.88, 3.1)	0.27	0.54 (-1.5, 2.5)	0.60
Diagnostic certainty	0.69 (0.30, 1.1)	0.001	0.77 (0.42, 1.1)	< 0.001	-0.24 (-0.49, 0.005)	0.054	-0.20 (-0.45, 0.055)	0.13

CI: confidence interval.

Figure 1. Threshold for the diagnosis of heart failure for trainers and trainees.

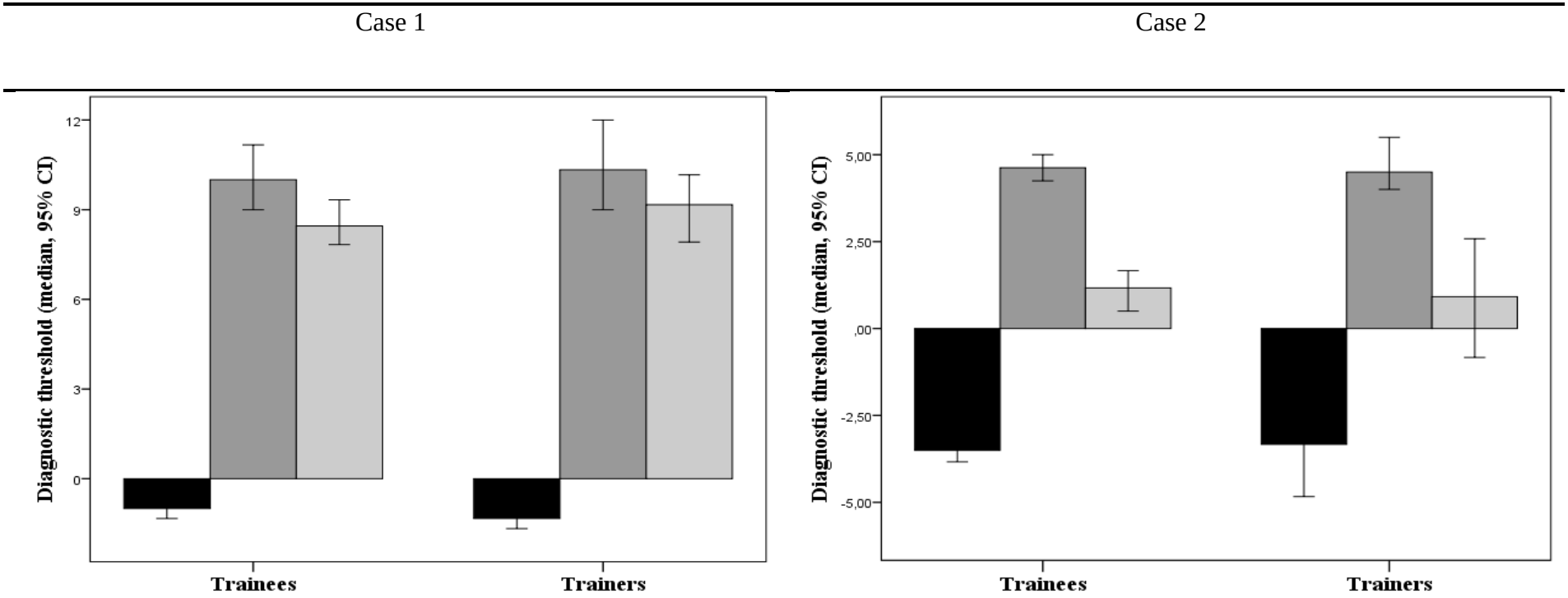
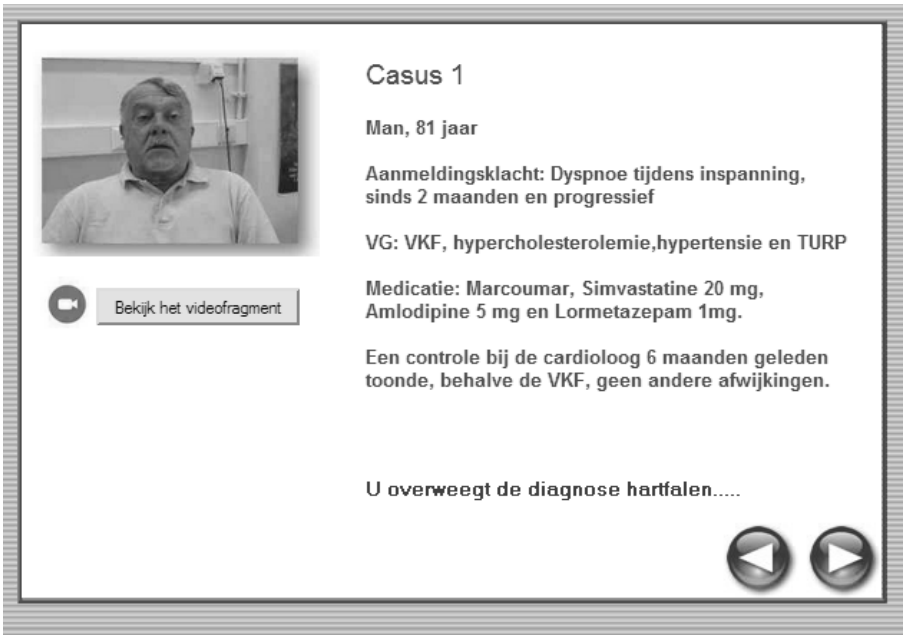


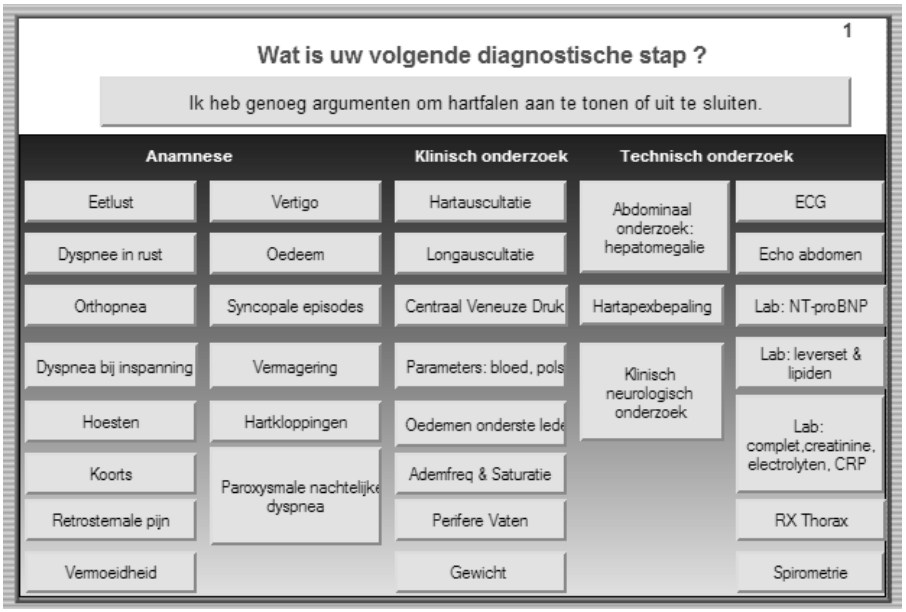
Figure legend.

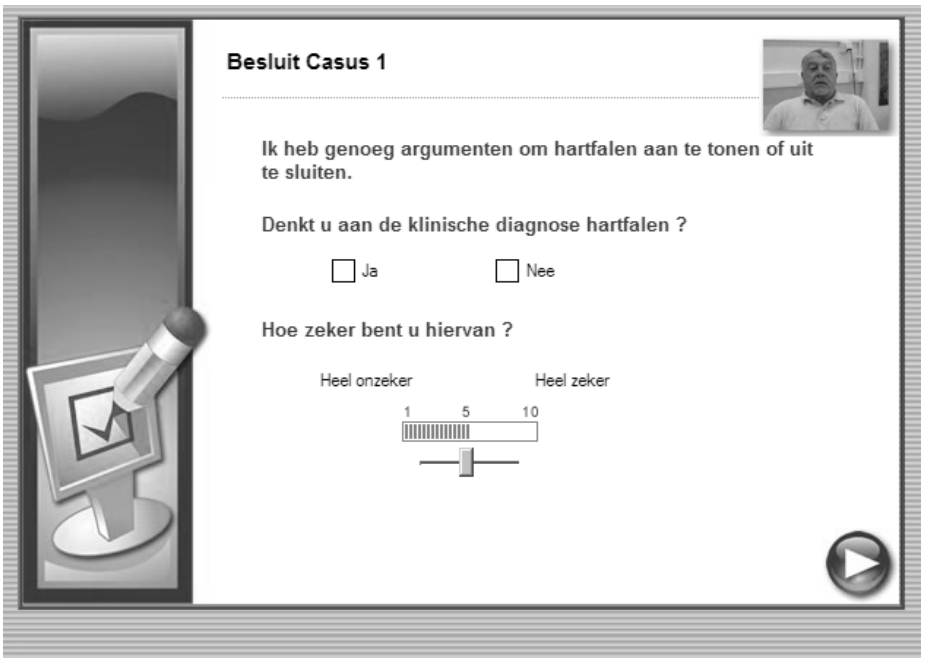
Exclusive power Demonstrating power Diagnostic threshold

Appendix

Example major case:







Example minor case:

