BMJ Open

Age and gender differences in diagnostic decision-making of early heart failure: results of a mixed-methods interview-study using video vignettes

Gabriella Marx ,[1] Sarah Koens ,[2] Olaf von dem Knesebeck,[2] Martin Scherer[1]

To cite: Marx G, Koens S, von dem Knesebeck O, *et al*. Age and gender differences in diagnostic decision-making of early heart failure: results of a mixed-methods interview-study using video vignettes. *BMJ Open* 2022;**12**:e054025. doi:10.1136/bmjopen-2021-054025

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-054025).

GM and SK contributed equally.

Received 31 May 2021
Accepted 10 February 2022

[1]Department of General Practice and Primary Care, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2]Institute of Medical Sociology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Correspondence to
Gabriella Marx; g.marx@uke.de

ABSTRACT

Objectives Men and women at any age show similar symptoms and signs of heart failure (HF). Since early HF symptoms are ambiguous, doctors may overlook HF as possible cause and misinterpret the signs. The aim was to analyse differences in general practitioners' (GPs) HF diagnosis and diagnostic certainty by patient age and gender and to identify reasons for possible differences.

Design Factorial design with video vignettes presenting patients (played by professional actors) with early HF symptoms was used. Video vignettes differed regarding patients' gender (male/female), age (55 years/75 years) and migration background (no/yes: Turkish), while the dialogue was identical. GPs were asked about possible diagnoses and certainty of diagnoses (quantitative) and to narrate their thoughts on considered diagnoses (qualitative).

Setting General practices in northern Germany.

Participants 128 GPs stratified by gender and length of clinical experience (≤15 years or >15 years).

Results GPs considered HF more often in women than men (predicted probabilities with 95% CI: 0.83 (0.68 to 0.92) vs 0.63 (0.44 to 0.79), p=0.02), especially in older women compared with younger men (predicted probabilities with 95% CI: 0.89 (0.68 to 0.96) vs 0.52 (0.31 to 0.72), p=0,03). Suspected HF was not reasoned by the patient's gender and only seldom by the patient's age, but by reported symptoms. Diagnostic certainty of HF was higher in women than in men (predicted proportions with 95% CI: 0.48 (0.39 to 0.58) vs 0.36 (0.27 to 0.45), p=0.01), with highest certainty in older women and lowest in younger men (0.57 (0.45 to 0.69) vs 0.27 (0.17 to 0.37), p<0.01). GPs explained their certainty referring to both typical HF symptoms and their gut feeling.

Conclusion Despite an identical dialogue, the study showed differences by patients' gender and age in frequency and certainty of HF diagnosis. In order to avoid that GPs overlook or misinterpret early signs of HF, it is important to critically reflect diagnostic decisions and possible social influences.

Strengths and limitations of this study

- By choosing a factorial design with video vignettes, we adopted an efficient methodical tool to investigate medical decision-making, since potential influencing factors can be either controlled or minimised.
- The underlying script of the vignettes used in this study was developed in two steps including clinical experts and patients with heart failure followed by a nationwide pilot test with general practitioners.
- As participants were selected according to the stratification criteria gender and length of clinical experience, our study population is not representative for primary care doctors.
- Since we used filmed simulated encounters, we cannot rule out that doctors may have responded differently than in an encounter with real patients.

INTRODUCTION

Heart failure (HF) is a prevalent disease affecting about 64 million people worldwide in 2017.[1] Since HF prevalence increases with age and about two-thirds of patients with HF are at least 75 years old, HF will be of future relevance due to the demographic change.[2–4] It is known that prevalence of HF is equal among men and women with a lifetime risk of about 20% for both sexes at 40 years of age.[5–7] HF survival rate has generally improved during the past decades, but this was less the case for women and older people.[4 7–9]

To date, much knowledge exists on HF prevalence, incidence, comorbidities, mortality, aetiology, mechanisms, diagnosis and prognosis considering specifics on gender and age as shown in several recent reviews.[3 5 10–13] However, early HF in many cases remains unrecognised and untreated because of the often ambiguous signs and symptoms that may be misinterpreted. Most of these early symptoms, such as ankle swelling, reduced exercise tolerance, breathlessness or fatigue/tiredness, may be assessed as symptoms of other diseases as, for example, pulmonary or gastrointestinal diseases or psychological disorders. Moreover, 'particularly among older patients, symptoms may be considered normal ageing'.[10] Female patients

are especially affected by a delay in HF diagnosis and treatment, which is reasoned by provider perception of HF as 'man's syndrome' and a lack of awareness of gender-related issues.[14 15] It is known that younger men develop HF more often than younger women, whereas HF prevalence is higher in older women.[5] Both men and women show similar symptoms and signs,[11] but studies indicated that physicians may overlook HF as a possible cause in terms of misinterpreting the signs as Chronic obstructive pulmonary disease (COPD) or asthma in female patients.[11 15 16] Thus, women affected by HF need increased attention when presenting such symptoms.

In case of ambiguous but typical symptoms, general practitioners (GPs) are main contact for patients presenting with symptoms for the first time. Therefore, and since about two-thirds of patients with HF are treated by GPs, they play a central role in HF care.[6 10 17] The key role of GPs and GPs' decision-making in this context requires a need for diagnostic improvement and awareness of early signs. It was shown that prognosis depends on severity, age and gender of patients, which supports an early diagnosis followed by an early beginning of therapy,[18] but little is known about how GPs consider patients' gender and age during their decision-making processes. In previous analyses, we found that in patients' medical history, GPs have a greater focus on lifestyle aspects in men than in women which is not in line with current guidelines.[19]

To provide equal care to women and men with HF regardless of their age, insights in how primary care doctors interpret early HF symptoms and signs are needed. In this study, we therefore addressed the following questions: (1) Do GPs suspect HF in patients presenting with typical but ambiguous HF symptoms in general? (2) Are there differences in frequency and certainty of suspected HF diagnose according to patients' gender and age? (3) Are physicians' diagnostic decisions reasoned by gender and age of patients?

METHODS

Study design

To examine differences by age and gender of patients in primary care physicians' suspected diagnosis in patients with typical but ambiguous HF symptoms, we applied a factorial experimental design with clinically authentic video vignettes.[20 21] Using vignettes enables research which 'inherits the external validity strengths of survey research and the internal validity strengths of experimental methods.'[22] Controlled manipulations of factors of interest are possible, while other conditions are held constant.[23]

In the videos, professional actors played patients reporting typical New York Heart Association (NYHA) class II–III HF symptoms (eg, dyspnoea, fatigue, peripheral oedema) in a first non-urgent consultation with a primary care doctor. The script did not consider possible age or gender-related differences in presenting the type or the extent of symptoms, since the focus was rather on verbalised signs in general. Therefore, the actors chosen to play the role of the patients represent patients with possible early HF, lacking typical obvious physical signs of progressed HF (eg, shortness of breath at rest, persistent cough, meteorism). Box 1 gives an overview on the information given by the patients (please see online supplemental file 1 for the complete script). The script for the simulated encounter was developed in workshops (WS) with clinical experts (internists, GPs and cardiologists) and patients with HF. The dialogue was initially drafted by the researchers (GPs, sociologists, health services researchers) in accordance with international guidelines (content, structure and duration of typical anamnesis in Germany). In WS 1, the draft was discussed and consented from the clinical perspective. In WS 2, patients shared their own experiences with HF-related early consultations in small groups (four to five participants), each moderated by a non-clinical researcher. In the small groups, the patients' experiences were referred to the script and any differences were documented. The results were finally discussed with the whole group (n=12). Discussion showed that the script was mainly in accordance with the patients' experiences and patients found the dialogue typical and the doctor's questions realistic. However, patients reported that their blood pressure was not measured in the practice, wherefore in the final script the doctor asked for the current blood pressure. A second and a third point, examination of smoking and alcohol consuming, were deliberately omitted in the script, because the study aimed, among others, to explore, if doctors would address these issues as part of the medical history.

The dialogue was identical in all vignettes, while the patients, played by professional actors, differed in terms of gender (male/female), age (55 years/75 years) and Turkish migration background (no/yes). Combination of these patient characteristics resulted in eight video

Box 1 Information given by the patient in the vignette

Symptom-related information:
- Fatigue.
- Dyspnoea.
- Reduced exercise tolerance.
- Coughing.
- Pain in the side under exercise.
- Worsening sleeping disorder.
- Nocturia.
- Swollen legs.
- Slight weight gain.

Other information:
- Last physician consultation 5 years ago.
- Mother died a few months ago.
- No siblings.
- No known family illnesses.
- Cold a few weeks ago.
- No known pre-existing conditions (thyroid, blood sugar) or allergies.
- Normal to slightly increased blood pressure.

vignettes. Gender of patients in the video was represented by the name, which was shown in the beginning of the video. Additionally, the doctor addressed the person in the video as Mr Schmidt/Yildiz or Mrs Schmidt/Yildiz, respectively. Turkish migration background was indicated by appearance, by the name shown in the video (Mr/Mrs Schmidt or Mr/Mrs Yildiz) and a slight accent. Suitable actors were chosen in a casting conducted by an agency, and a professional filming team was assigned for the filming and editing of the vignettes. The videos were shot in a doctor's office, as one-shot videos and were filmed from the perspective of the doctor with view on the patient to create an authentic scenario. To minimise variation, the same recording of the doctor's voice was used for all videos. Moreover, all eight actors were instructed to portray a patient with HF in a similar way according to the first video, which was shown to the other actors in advance.

Sampling

A sample of doctors was drawn from a comprehensive list of internists and GPs in northern Germany. GPs were invited by post and informed about study details. The purpose of the study, diagnostic process of HF, however was not communicated. Instead, the aim was referred to as investigating decision-making processes in general. For the factorial design, the study population was stratified by doctors' gender and length of clinical experience as a board-certified internist or GP (≤15 years or >15 years) into four strata. To detect differences of 25% for the main effects (gender or age of patient) with a statistical power of 80% and a type I error of 0.05, a sample size of 128 physicians was needed. The final sample consisted of 64 female and 64 male doctors; half of them had ≤15 years of clinical experience and half of them had >15 years of clinical experience. Each of the eight videos was shown four times in each doctor stratum. Within the strata, the videos were allocated randomly.

Additional inclusion criterion was working in primary care at least 20 hours per week. GPs were invited per post and contacted per phone in case of no response. The invitation letter included a response form to request information on inclusion criteria as well as gender and length of clinical experience. Of eligible doctors, 50.4% participated. The recruitment process is presented in figure 1. To acknowledge participation and to partly offset loss of revenue, GPs were offered a stipend. All participating doctors signed informed consent.

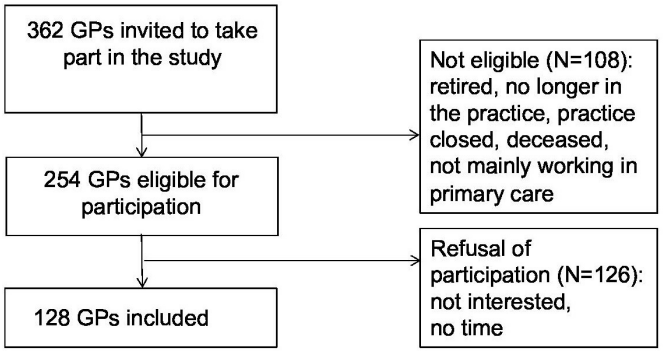

Figure 1 Flow chart. GPs, general practitioners.

Data collection

The script was pilot tested with GPs from different regions in Germany. The aim was to test whether the given information reflects a typical and realistic presentation of symptoms and doctor–patient conversation in general practice which could be confirmed. The pilot test did not show any need for change. After pilot testing, data collection was conducted by GM (sociologist, highly experienced interviewer), SK (healthcare scientist, trained interviewer) and a medical student (trained interviewer) as single face-to-face interview in the doctor's practice. The videos were only presented to the participating doctors in their own practice room directly after or during their usual practice hours. They were instructed to view the person in the video as a patient in their own practice. One video was shown to each doctor.

Each interview lasted about 30–40 min and consisted of a standardised and a qualitative part. After viewing the video, among others, GPs were asked about possible diagnoses (open ended, multiple answers possible) and certainty of suspected diagnoses (in per cent). The subsequent qualitative part was conducted following retrospective think aloud method[24] using narrative interview technique.[25] Since the time between watching the video and narrating the thoughts was extended by the quantitative interview part, we assumed that the 'inherent demands of narration' (Zugzwaenge des Erzaehlens) were able to close this gap.[26]

For the qualitative part, participants were initially requested to think at the beginning of the video and then narrate all thoughts they have had regarding the 'patient' and possible diagnoses while watching the video. A screenshot of the 'patient', with the indication of the name and age, was presented. The interviewer encouraged the narration by active listening (eg, *mhm* or nodding) and did not interrupt the story until the interviewee set the coda to signal the end (eg, *that's it*). The introduction was: 'Please think at the beginning of the presented case [point at screenshot]. Please tell me your thoughts you have had while watching the consultation from the beginning to the end of the shown video. Please go into all details, thoughts, ideas and aspects, you have had regarding the patient, the diagnosis and the possible therapy, also those you have discharged or you felt unsure. Please go into all details you remember. All details you consider as important are interesting for us.' The aim of this part of the interview was to uncover the underlying cognitive reasons in the decision-making processes. No specific questions regarding the impact or relevance of gender and age have been asked.

All interviews were audio-taped, transcribed verbatim and anonymised while transcription. Vignettes shown in the respective interview were identified by patient

characteristics, for example, female/old/migration history (FOMh), male/young/no migration history (MYnM). Transcripts were not returned to participants.

Data analysis

In this paper, the focus was on age and gender differences in HF diagnosis and diagnostic certainty. To analyse the quantitative part, all terms used for HF as suspected diagnoses in the free-text answers (eg, HF, cardiac failure, cardiac inefficiency) were coded as HF diagnosis. We did not prioritise according to the order of the suspected diagnoses mentioned and it was not relevant for us whether the suspected diagnosis of HF was mentioned first. Due to the balanced factorial design, differences by gender or age of patient are unconfounded (orthogonal to) with the other design factors (patients' migration history, doctors' gender and length of experience). Logistic regression analysis was computed to examine differences in frequency of suspected HF diagnosis by age and gender of patients. To adjust for physician characteristics, the model included migration background and age of the physicians. To compare frequencies by patient age and gender, predicted probabilities with 95% CIs for suspected HF diagnosis were computed. For the analysis of differences in physicians' certainty of HF diagnosis (in per cent), beta regression was computed and predicted proportions were calculated. To adjust for physician characteristics, the model included age and migration background of the physicians. To analyse differences in HF diagnosis and certainty of diagnosis in more detail, estimated marginal means for the interaction between age and gender were computed. Additionally, predicted probabilities of HF diagnosis and predicted proportions of diagnostic certainty, adjusted for age and migration background of the physicians, were calculated. Tukey-adjusted p values for pairwise comparison of four groups were calculated. P values of <0.05 were considered statistically significant. For the statistical analyses, the R statistical package was used,[27] including the packages emmeans and betareg.[28 29]

The qualitative parts were analysed using the first step of the documentary method (formulating interpretation) which focuses on the content of the narrations ('what was said'). For this, all interviews were thematically sequenced and then coded using deductive (derived from the quantitative part, for example, 'HF as suspected diagnosis') as well as inductive codes. Analysis was conducted by four researchers and discussed within the team in an iterative process to consent the codes. Findings were not cross-checked by participants. Coding was supported using Maxqda V.12.[30]

Finally, results of the quantitative part were triangulated with corresponding qualitative results to get in-depth knowledge of how and why GPs argued their decision on HF diagnosis regarding gender and age differences.

Patient and public involvement

Patients were involved in the process of the development of the dialogue for the vignettes. We conducted a WS with patients with HF, to discuss the dialogue. Patients were asked if they found the dialogue realistic and were encouraged to frankly make their own proposals. After the WS, small adjustments were needed. There was no further patient or public involvement.

RESULTS/FINDINGS

From June 2018 to February 2019, 128 primary care physicians (50% women) in practices in Hamburg (60.2%) and surrounding areas (39.8%) were interviewed face to face. Of the participants, 35.2% had a single practice, 63.3% worked in a joint practice and 1.6% in a medical care unit. The majority of the physicians were GPs (72.7%), 18.0% were internists and 9.4% were board-certified GPs and internists. In the interviews, 10.2% of the physicians reported having a migration background, meaning that they themselves or their mother and/or father were not born in Germany.

Frequency of suspected HF diagnosis

HF was named as a suspected diagnosis by 99 (77.3%) doctors. Predicted probabilities of suspected HF diagnosis by patient age and gender are presented in table 1. The results are adjusted for physicians' age and migration background. Doctors considered HF as a diagnosis significantly more often in women and in tendency in older patients (not significant).

Regarding the constellation of patient age and gender, doctors suspected HF most often in older women and least frequently in younger men (p=0.03, table 2).

Interestingly, despite significant differences in frequency of HF as suspected diagnosis, in the qualitative part, the patient's gender as decision criterion was either not mentioned at all or explicitly excluded as such:

> so the complaints she described simply, first of all regardless of gender, origin, age and so on and so on always think of heart failure, shortness of breath when

Table 1 Predicted probabilities of HF diagnosis by patients' age and gender with 95% CIs*

Gender			Age		
Male	Female	P value	55	75	P value
0.63 (0.44 to 0.79)	0.83 (0.68 to 0.92)	0.02	0.65 (0.47 to 0.80)	0.82 (0.66 to 0.92)	0.05

*Differences were analysed with logistic regression analysis with estimated marginal means, adjusted for age and migration background of physicians.
HF, heart failure.

Table 2 Predicted probabilities of HF diagnosis by patients' age and gender groups with 95% CIs*

Male, 55	Female, 55	Male, 75	Female, 75
0.52 (0.31 to 0.72)†	0.77 (0.54 to 0.90)	0.73 (0.49 to 0.88)	0.89 (0.68 to 0.96)‡

*Differences were analysed with logistic regression analysis with estimated marginal means, adjusted for age and migration background of physicians.
†Significantly different from female, 75 (p=0.03).
‡Significantly different from male, 55 (p=0.03).
HF, heart failure.

climbing stairs NYHA two to three fatigue and so on and so on nightly urination uh all that (1012_FOMh).

Instead, the focus as reason for considering HF was explicitly on typical HF symptoms. Most doctors mentioned the main symptoms dyspnoea (n=50), water retention (n=48) and nocturia (n=39). Frequency of these named symptoms did not differ by age or gender of the patients presented in the videos:

> Almost textbook-like all the symptoms associated with heart failure // mhm // edema nocturia exertional dyspnoea um, rapid exhaustion. (1007_FOnM)

> […] but I thought relatively quickly about the uh, diagnosis hm heart failure because he, spoke very quickly of breathing difficulties and especially of stress-related breathing difficulties, […] hmm then he spoke of his infection, which brought me to mm a relatively sudden beginning of a possible heart failure. (1039_MYnM)

> When the other syndromes- symptoms appeared, swollen legs, um nightly awakening, frequent urination the colleague asked everything then it was already clear that it was heading towards heart failure […]. (1051_FYMh)

The patient's age was mentioned as a reason for diagnosing HF in a few cases presenting older patients only (n=7):

> Then of course the age of the patient plays a role that certain diseases at this age have a certain frequency and others are rather unlikely, so it is unfortunately so that malignant diseases at this age also become more probable but also heart failure. (1079_MOnM)

> From the symptoms […] it looked like heart failure, uh, with the stress dyspnoea, uh, with the nocturia, uh, with the cough, uh, so it all fits together, uh, it often occurs at this age. (1053_FOnM)

Certainty of HF diagnosis

Table 3 shows predicted proportions of doctors' certainty of HF diagnosis by patient gender and age. Certainty of doctors' HF diagnosis was significantly higher in female patients than in male patients and in older patients compared with younger patients.

Comparing the four groups of patients, diagnostic certainty was lowest in younger male patients, which was significantly different from older female patients, where certainty was highest (p<0.01). Diagnostic certainty also differed significantly by age in male patients (p=0.03), but there was no significant difference by age in female patients (table 4).

Qualitative analysis showed that doctors did not explicitly mention gender or age of the patients as influencing factor, which could be suspected due to quantitative data. Instead, they prioritise HF diagnosis mainly regarding symptoms. Especially in older patients, doctors focused solely on symptoms without any gender difference. However, only 13 doctors addressed this issue at all reasoning that patient's presented typical HF symptoms:

> Yeah, she actually describes everything for heart failure […] and that's just exertional dyspnoea, um nightly urinating um exhaustion, tiredness […] so this is clearly the prioritization for me. (1017_FOnM)

> The patient has described them [symptoms] fairly accurately […] so that makes the diagnosis of heart failure very likely. (1079_MOnM)

> I think she just said enough keywords that would just fit with the heart failure, well // mhm // that's why the diagnosis has just slipped into the foreground. (1016_FOnM)

Table 3 Predicted proportions of certainty of HF diagnosis by patient gender and age with 95% CIs*

Gender			Age		
Male	Female	P value	55 years	75 years	P value
0.36 (0.27 to 0.45)	0.48 (0.39 to 0.58)	0.01	0.33 (0.25 to 0.42)	0.51 (0.41 to 0.61)	<0.01

*Differences were examined by beta regression analysis with estimated marginal means, adjusted for age and migration background of physicians.
HF, heart failure.

Table 4 Predicted proportions with 95% CIs of certainty of HF diagnosis for gender and age groups*

Male, 55	Female, 55	Male, 75	Female, 75
0.27 (0.17 to 0.37)†	0.40 (0.28 to 0.52)	0.46 (0.33 to 0.59)‡	0.57 (0.45 to 0.69)§

*Differences were examined using beta regression analysis with estimated marginal means.
†Significantly different from male, 75 (p=0.03) and female, 75 (p<0.01).
‡Significantly different from male, 55 (p=0.03).
§Significantly different from male, 55 (p<0.01).
HF, heart failure.

Additionally, symptoms were further distinguished by severity and frequency:

> I would always concentrate more on the life-threatening or potentially threatening things in the first conversation […] so I would first clarify the possible potentially threatening things. (1017_FOnM)

In general, regarding diagnostic decision-making, doctors gave special attention to their gut feeling, their own professional experience, patients' presentation and patients' burden of suffering:

> These are just the things, where I say then you need your gut, the gut feeling must be be there that you then say you first look in this direction. (1031_FOnM)

> Of course, the age has to be recognised additionally, so you have to consider what happens most often in old age, and the complaints that he expresses then also fit with certain diseases, so [pause] yes actually that is, yes that is what you are so a bit, then somehow from the gut. (1038_MOMh)

GPs find HF diagnosis in younger patients more difficult, since other diseases are often also plausible:

> With a man you naturally think of, at the age of 55, everything uh, from uh from the hypothyroidism to uh to the male menopause, that can be anything. (1039_MYnM)

> […] what is the more likely diagnosis and I have to say that for the 55-year-old of course I think the CHD [coronary heart disease] is also a bit more likely than a heart failure if there is not a cardiomyopathy due to the infection somehow lies behind it. (1103_MYnM)

DISCUSSION

In the quantitative part of the interview, GPs suspected HF more often in female patients, especially in older women. They were also more certain about their suspected diagnosis in women regardless of patients' age, and in older patients in general, reasoned by the patient's symptoms. Diagnostic certainty varied in terms of male patients' age. GPs did not explicitly state age and gender of the patients as relevant factors for diagnostic decision-making in the qualitative part of the interview. Regarding the significant age and gender differences in diagnosis and certainty within the quantitative part, the qualitative analysis indicates that these are the result of unconscious decision-making processes. As it is commonly known from theories in the field of sociology and social psychology, interaction is influenced not only by verbal content but also by subtle and non-verbal communication (gestures). Social interaction, however, relies on the process of situative mutual and often unconscious interpretation of non-verbal gestures. Moreover, misinterpretation is possible, not always reflected and therefore cannot be verbalised.

GPs described HF diagnosis as difficult since typical symptoms are non-specific and in particular in younger patients, other conditions can cause the same symptoms. Analysis shows a discrepancy between significant gender and age-related differences in diagnosis and diagnostic certainty on the one hand, and the missing explication of the relevance of age or gender in the qualitative part of the interview on the other hand. Since GPs were not asked to focus on age or gender of the patients when recapitulating their thoughts during watching the vignette, but to set their own priorities, it seems plausible that age and gender influence diagnosis and diagnostic decision-making on an unconscious level.

The application of a factorial design with video vignettes enabled us to overcome ethical and methodical issues of other approaches investigating medical decision-making.[22] Nevertheless, some limitations have to be considered, when interpreting our results. As participants were selected according to gender and length of clinical experience as stratification criteria, our study population is not representative for primary care doctors. According to the vignettes methodology, the doctor–patient interaction did not differ between age, gender or migration background of the doctor, but was identical in all vignettes as it was for patients. This implies the possibility that the identification with the doctor in the video may be lacking. However, the doctor was not shown in the video and the focus of the study was on presenting the patient's symptoms by eliminating as much influencing factors as possible. Due to the filmed simulated encounters, we cannot rule out that participating doctors may interact differently with real patients. However, it was shown that video vignettes are a valid way to assess doctors' practice[31] and can be generalised to real practice.[22] We took several steps to increase realism of the vignettes: (a) we developed the script together with clinical experts and patients with HF; (b) the videos were filmed by a professional filming

team with professional actors; (c) to enable the doctors to identify themselves with the depicted situation, the videos were filmed in a normal practice room and shown to the participants in their own practice room during or directly after their consultation hours; (d) the participants were told that the interview was no test but we were interested in their process of medical decision-making; and (e) we asked the doctors to view the patient as one of their own patients. The doctors were asked how typical the patient was compared with a patient in their practice and how realistic the dialogue was for a consultation in their practice. Of the physicians, 83.6% said the patient was very or rather typical and 84.4% considered the dialogue very or rather realistic.

Difficulties to attribute ambiguous symptoms for HF are a common problem in HF diagnosis in general practice.[17 32] In this study, physicians were not asked to give a final correct diagnosis, but they were asked what diagnoses they were considering based on the information given in the video, as we were interested in whether HF was mentioned as a suspected diagnosis. Suspected diagnoses are an important element in the diagnostic process since these hypotheses determine the further procedure.[33] Although the average of HF prevalence and incidence shows only a slight difference between men and women in Germany, GPs seem to be sensitised for HF in older women affecting the diagnosis and the certainty about it. This might be due to three reasons: first, patients with HF treated by GPs alone, instead of cardiologists, are often older and female[34 35]; second, the rise of HF prevalence in women at age of 80 years compared with men[36]; and third, a greater awareness of poor recognition of heart diseases in women over the last decades in general.[37] In the past, HF was classified as men-specific disease, assuming a hormone-related protection for premenopausal women,[37 38] which can be rebutted by more recent studies.[5 6 36] On this basis, it remains unclear why women are less often treated as recommended in the guidelines.[39 40]

Epidemiological studies show a higher rate of female patients with HF, especially those with preserved ejection fraction (HFpEF). The fact that both HFpEF symptoms and stage 2 HF show ambiguous symptoms related to other diseases may lead to less recognition of HF in women and in early HF stages.[10] A possibly existing gender awareness with a focus on older women in HF diagnosis may compensate this diagnostic difficulty. However, it needs to be recognised that female patients with HF tend to present a higher symptom burden including symptoms shown in the vignettes (eg, dyspnoea, oedema, fatigue).[12 15] It cannot be ruled out that the videos had shown more typical female patients with HF, but male and female experts and patients who were involved in the development of the script, and pretests did not query the dialogue. Especially in older patients, clinical symptoms were focused on diagnosis in our sample and doctors explained their decision and their certainty referring to both typical HF symptoms and their gut feeling. The latter is common in GPs' diagnostic decision-making and is known as a relevant third track of diagnostic reasoning besides medical decision-making and medical problem-solving.[41 42] The gut feeling might be a plausible reason for the differences between the decisions made in our sample and the participants following narrations and explanations of their thoughts while watching the vignette.

To conclude, gender and age awareness within HF diagnosis is mandatory to avoid overdiagnoses or underdiagnoses and, as a result, oversupply or undersupply on the basis of age and/or gender blindness. Stereotypes on (older and younger) age and gender are common in daily social interaction but may have a relevant impact on healthcare.[43 44] Therefore, GPs need to critically reflect their action and be aware of possible social influences. A special consideration should still be given to younger patients and younger men in particular. In daily clinical practice, GPs should question if their assumptions regarding the interpretation of symptoms and suspected diagnoses may be influenced by age and gender of the patients. This should be addressed in medical education and in continuing medical training for primary care physicians. Additionally, the findings could be taken into account in guideline development.

Acknowledgements The authors would like to thank the GPs for taking part in this study and giving insights in their thoughts, the clinical experts and patients for their support in developing the script for the video vignettes, the study team for discussing qualitative data, Colette Gras for supporting data collection and analysis, Daniel Lüdecke for statistical advice, and the Federal Ministry of Education and Research for funding the research.

Contributors OvdK was the principal investigator of the study, led the study design, supervised the research process, and is responsible for the overall content as guarantor. MS was co-investigator of the study and co-designed the study. OvdK, MS, GM and SK developed the script used in the video vignettes. GM and SK prepared the questionnaire and the interview guide, recruitment and data collection, and interviewed the participants, conducted data analysis and repeatedly discussed the findings of qualitative data. GM and SK wrote the first draft. All authors critically reviewed and approved the final version.

Funding This work was supported by the Federal Ministry of Education and Research (FKZ 01GL1711).

Disclaimer The views stated are those of the authors and not necessarily those of the funder.

Competing interests None declared.

Patient consent for publication Not required.

Ethical approval According to a statement of the Ethics Committee of the Hamburg Medical Association (processing number: PV5663), no ethics approval was needed to conduct the study.

Provenance and peer review Not commissioned; externally peer reviewed.

DATA SHARING STATEMENT No additional data are available.

ORCID iDs
Gabriella Marx http://orcid.org/0000-0001-6807-9518
Sarah Koens http://orcid.org/0000-0002-4390-2608

REFERENCES

1 Ponikowski P, Anker SD, AlHabib KF, *et al*. Heart failure: preventing disease and death worldwide: addressing heart failure. *ESC Heart Fail* 2014;1:4–25.
2 Gerber Y, Weston SA, Redfield MM, *et al*. A contemporary appraisal of the heart failure epidemic in Olmsted County, Minnesota, 2000 to 2010. *JAMA Intern Med* 2015;175:996.
3 Farré N, Vela E, Clèries M, *et al*. Real world heart failure epidemiology and outcome: a population-based analysis of 88,195 patients. *PLoS One* 2017;12:e0172745.
4 Groenewegen A, Rutten FH, Mosterd A, *et al*. Epidemiology of heart failure. *Eur J Heart Fail* 2020;22:1342–56.
5 Bozkurt B, Khalaf S. Heart failure in women. *Methodist Debakey Cardiovasc J* 2017;13:216–23.
6 Lloyd-Jones DM, Larson MG, Leip EP, *et al*. Lifetime risk for developing congestive heart failure: the Framingham heart study. *Circulation* 2002;106:3068–72.
7 AS G, Mozaffarian D, Roger VL, *et al*. Heart disease and stroke Statistics—2014 update: a report from the American heart association. *Circulation* 2014;129.
8 Levy D, Kenchaiah S, Larson MG, *et al*. Long-Term trends in the incidence of and survival with heart failure. *N Engl J Med* 2002;347:1397–402.
9 Roger VL, Weston SA, Redfield MM, *et al*. Trends in heart failure incidence and survival in a community-based population. *JAMA* 2004;292:344.
10 Peterson PN, Allen LA, Heidenreich PA, *et al*. The American heart association heart failure Summit, Bethesda, April 12, 2017: proceedings and calls to action. *Circ Heart Fail* 2018:11.
11 Mentzer G, Hsich EM. Heart failure with reduced ejection fraction in women. *Heart Fail Clin* 2019;15:19–27.
12 Savarese G, D'Amario D. Sex differences in heart failure. In: Kerkhof PLM, Miller VM, eds. *Sex- specific analysis of cardiovascular function*. Cham: Springer International Publishing, 2018: 529–44.
13 Garcia M, Mulvagh SL, Merz CNB, *et al*. Cardiovascular disease in women: clinical perspectives. *Circ Res* 2016;118:1273–93.
14 Marra AM, Salzano A, Arcopinto M, *et al*. The impact of gender in cardiovascular medicine: lessons from the gender/sex-issue in heart failure. *Monaldi Arch Chest Dis* 2018;88:988.
15 Eisenberg E, Di Palo KE, Piña IL. Sex differences in heart failure. *Clin Cardiol* 2018;41:211–6.
16 Lund LH, Mancini D. Heart failure in women. *Med Clin North Am* 2004;88:xii:1321–45.
17 Ponikowski P, Voors AA, Anker SD, *et al*. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2016;2016:2129–200.
18 Watson RD, Gibbs CR, Lip GY. Abc of heart failure. clinical features and complications. *BMJ* 2000;320:236–9.
19 Koens S, Marx G, Gras C, *et al*. Physicians' information seeking behavior in patients presenting with heart failure symptoms – Does gender of physician and patient matter? *Patient Educ Couns* 2020;103:2437–42.
20 von dem Knesebeck O, Gerstenberger E, Link C, *et al*. Differences in the diagnosis and management of type 2 diabetes in 3 countries (US, UK, and Germany): results from a factorial experiment. *Med Care* 2010;48:321–6.
21 Bönte M, von dem Knesebeck O, Siegrist J, *et al*. Women and men with coronary heart disease in three countries: are they treated differently? *Womens Health Issues* 2008;18:191–8.
22 Evans SC, Roberts MC, Keeley JW, *et al*. Vignette methodologies for studying clinicians' decision-making: validity, utility, and application in ICD-11 field studies. *Int J Clin Health Psychol* 2015;15:160–70.
23 Hillen MA, van Vliet LM, de Haes HCJM, *et al*. Developing and administering scripted video vignettes for experimental research of patient-provider communication. *Patient Educ Couns* 2013;91:295–309.
24 Ericsson KA. Protocol analysis and expert thought: concurrent verbalizations of thinking during experts' performance on representative tasks. In: Ericsson KA, Charness N, Feltovich PJ, *et al*, eds. *The Cambridge Handbook of expertise and expert performance*. Cambridge: Cambridge University Press, 2006: 223–42.
25 Flick U, ed. The Sage handbook of qualitative data collection. Los Angeles: Sage Reference, 2018.
26 Jovchelovich S, Bauer MW. Narrative interviewing. In: Bauer MW, Gaskell G, eds. *Qualitative researching with text, image and sound : a practical handbook*. London, UK, 2000: 57–74.
27 R Development Core Team. A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing 2019. Available: https://www.r-project.org [Accessed 28 May 2020].
28 Lenth R, Singmann H, Love P, *et al*. emmeans: estimated marginal means AKA least-squares means, 2020. Available: https://CRAN.R-project.org/package=emmeans
29 Zeileis A, Cribari-Neto F, Gruen B, *et al*. betareg: beta regression, 2020. Available: https://CRAN.R-project.org/package=betareg
30 VERBI Software. MAXQDA 2020 online manual, 2019. Available: maxqda.com/help-mx20/welcome [Accessed 6 May 2020].
31 Peabody JW, Luck J, Glassman P, *et al*. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715–22.
32 Smeets M, De Witte P, Peters S, *et al*. Think-aloud study about the diagnosis of chronic heart failure in Belgian general practice. *BMJ Open* 2019;9:e025922.
33 Faller H, Meng K, Neuderth S, *et al*. Judgement and decision-making. In: Faller H, Lang H, eds. *Medical psychology and sociology*. Berlin: Springer, 2019: 265–73.
34 Rutten FH, Grobbee DE, Hoes AW. Differences between general practitioners and cardiologists in diagnosis and management of heart failure: a survey in every-day practice. *Eur J Heart Fail* 2003;5:337–44.
35 Gracia Gutiérrez A, Poblador-Plou B, Prados-Torres A, *et al*. Sex differences in comorbidity, therapy, and health services' use of heart failure in Spain: evidence from real-world data. *Int J Environ Res Public Health* 2020;17:2136.
36 Kaduszkiewicz H, Gerste B, Eisele M, *et al*. Heart failure: epidemiology and care. In: Klauber J, Günster C, Gertse B, *et al*, eds. *Health services report 2013/14: depression*. Stuttgart: Schattauer, 2014: 209–29. https://www.wido.de/fileadmin/Dateien/Dokumente/Publikationen_Produkte/Buchreihen/Versorgungsreport/2013-2014/Kapitel%20mit%20Deckblatt/wido_vsr2013-2014_gesamt.pdf
37 Maas AHEM, Appelman YEA. Gender differences in coronary heart disease. *Neth Heart J* 2010;18:598–603.
38 Prescott E, Hippe M, Schnohr P, *et al*. Smoking and risk of myocardial infarction in women and men: longitudinal population study. *BMJ* 1998;316:1043–7.
39 Lenzen MJ, Rosengren A, Scholte op Reimer WJM, *et al*. Management of patients with heart failure in clinical practice: differences between men and women. *Heart* 2008;94:e10.
40 Yancy CW, Fonarow GC, Albert NM, *et al*. Influence of patient age and sex on delivery of guideline-recommended heart failure care in the outpatient cardiology practice setting: findings from improve HF. *Am Heart J* 2009;157:754–62.
41 Stolper E, van Bokhoven M, Houben P, *et al*. The diagnostic role of gut feelings in general practice. A focus group study of the concept and its determinants. *BMC Fam Pract* 2009;10:17.
42 Stolper E, Van de Wiel M, Van Royen P, *et al*. Gut Feelings as a Third Track in General Practitioners' Diagnostic Reasoning. *J Gen Intern Med* 2011;26:197–203.
43 Jackson SE, Hackett RA, Steptoe A. Associations between age discrimination and health and wellbeing: cross-sectional and prospective analysis of the English longitudinal study of ageing. *Lancet Public Health* 2019;4:e200–8.
44 Hamberg K. Gender bias in medicine. *Womens Health* 2008;4:237–43.