Risk of bias in studies investigating novel diagnostic biomarkers for heart failure with preserved ejection fraction. A systematic review
Henkens, Michiel T. H. M.; Remmelzwaal, Sharon; Robinson, Emma L.; van Ballegooijen, Adriana J.; Aizpurua, Arantxa Barandiaran; Verdonschot, Job A. J.; Raafs, Anne G.; Weerts, Jerremy; Hazebroek, Mark R.; Wijk, Sandra Sanders-van

Published in:
European Journal of Heart Failure

DOI:
10.1002/ejhf.1944

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Risk of bias in studies investigating novel diagnostic biomarkers for heart failure with preserved ejection fraction. A systematic review

Michiel T.H.M. Henkens1*, Sharon Remmelzwaal2, Emma L. Robinson1, Adriana J. van Ballegooijen3, Arantxa Barandiarán Aizpurua1, Job A.J. Verduynen1,3, Anne G. Raafs1, Jeremy Weerts1, Mark R. Hazebroek1, Sandra Sanders-van Wijk1, M. Louis Handoko4, Hester M. den Ruijter5, Carolyn S.P. Lam6,7,8, Rudolf A. de Boer8, Walter J. Paulus9,10, Vanessa P.M. van Empel1, Rein Vos11, Hans-Peter Brunner-La Rocca1, Joline W.J. Beulens2,12, and Stephane R.B. Heymans1,10,13

1Department of Cardiology, Maastricht University Medical Centre, Maastricht University, Maastricht, The Netherlands; 2Department of Epidemiology and Biostatistics, Amsterdam Cardiovascular Research Institute, Amsterdam UMC, Amsterdam, The Netherlands; 3Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, The Netherlands; 4Department of Cardiovascular Sciences Research Institute, Amsterdam UMC, Amsterdam, The Netherlands; 5Department of Physiology, Amsterdam Cardiovascular Sciences Research Institute, Amsterdam UMC, Amsterdam, The Netherlands; 6National Heart Centre Singapore, Singapore, Singapore; 7Duke-National University of Singapore, Singapore, Singapore; 8Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 9Department of Physiology, Amsterdam Cardiovascular Sciences Research Institute, Amsterdam UMC, Amsterdam, The Netherlands; 10Netherlands Heart Institute (CIN), Utrecht, The Netherlands; 11Department of Methodology and Statistics, Maastricht University, Maastricht, The Netherlands; 12Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; and 13Department of Cardiovascular Research, University of Leuven, Leuven, Belgium

Received 20 March 2020; revised 19 June 2020; accepted 20 June 2020

Aim

Diagnosing heart failure with preserved ejection fraction (HFrEF) in the non-acute setting remains challenging. Natriuretic peptides have limited value for this purpose, and a multitude of studies investigating novel diagnostic circulating biomarkers have not resulted in their implementation. This review aims to provide an overview of studies investigating novel circulating biomarkers for the diagnosis of HFrEF and determine their risk of bias (ROB).

Methods and results

A systematic literature search for studies investigating novel diagnostic HFrEF circulating biomarkers in humans was performed up until 21 April 2020. Those without diagnostic performance measures reported, or performed in an acute heart failure population were excluded, leading to a total of 28 studies. For each study, four reviewers determined the ROB within the QUADAS-2 domains: patient selection, index test, reference standard, and flow and timing. At least one domain with a high ROB was present in all studies. Use of case-control/two-gated designs, exclusion of difficult-to-diagnose patients, absence of a pre-specified cut-off value for the index test without the performance of external validation, the use of inappropriate reference standards and unclear timing of the index test and/or reference standard were the main bias determinants. Due to the high ROB and different patient populations, no meta-analysis was performed.
Introduction

Heart failure with preserved ejection fraction (HFpEF) is a clinical syndrome that is associated with high mortality rates, poor quality of life and significant healthcare resource utilization.\(^1,2\) Currently, more than 5% of the elderly (>65 years of age) suffer from this debilitating syndrome.\(^1,2\) The prevalence is expected to rise even further in the upcoming years, due to the ageing population and the growing occurrence of other HFpEF risk factors.\(^1,2\)

Unfortunately, diagnosing HFpEF in the non-acute setting remains challenging. Natriuretic peptides (NPs) have limited diagnostic value for this purpose, which is mainly due to the high prevalence of conditions within this syndrome that can lead to higher \([\text{e.g., atrial fibrillation (AF), hypertension, pulmonary diseases, renal function disorders}]\) and lower \([\text{e.g., obesity}]\) circulating NP levels.\(^3–11\) Moreover, 18% to 30% of patients with haemodynamically proven HFpEF have NP levels below ‘diagnostic’ threshold.\(^12–14\)

The limited diagnostic accuracy of NPs, and the concept that other circulating biomarkers could help to diagnose this complex syndrome on a molecular level, has resulted in a multitude of studies investigating novel diagnostic HFpEF biomarkers.\(^3,11\) Remarkably, none of the suggested circulating biomarkers have been implemented in the HFpEF clinics. The heterogeneous and systemic nature of the syndrome could contribute to their lack of success,\(^11\) but a comprehensive overview of the literature on this topic is absent. We therefore aimed to provide an overview of studies investigating the diagnostic value of novel biomarkers for non-acute HFpEF and determine their risk of bias (ROB).

Methods

A systematic literature search—based on the PRISMA-DTA statement\(^15\)—of PubMed and EMBASE was performed to find diagnostic papers within the field of HFpEF from its inception until 21 April 2020. A broad search (online supplementary Appendix S1) was used for a set of systematic reviews and a meta-analysis for the (early) detection of left ventricular diastolic dysfunction (LVDD) and/or HFpEF. The search strategy and the protocol can be found on PROSPERO (CRD42018065018). Studies that reported the diagnostic value of novel circulating biomarkers for the detection of chronic HFpEF were included in this study.

Study selection

Four reviewers (SR, MLH, AB and JB) screened the titles and abstracts independently. Studies were included if they: (i) reported a diagnostic performance measure (e.g. area under the receiver operating curve, sensitivity, specificity, negative predictive value, positive predictive value) of a novel circulating biomarker for the diagnosis of HFpEF in humans as main or sub-analysis; and (ii) were written in English. Studies were excluded if they: (i) studied the diagnostic value of a biomarker in acute heart failure; (ii) only studied the diagnostic value of NPs; (iii) studied the diagnostic value within a rare patient population (e.g. beta thalassemia); or (iv) were a (systematic) review, meta-analysis, editorial, or conference abstract.

Data extraction

The following data were extracted for each study: publication details (first author, year of publication), study characteristics (patient population description, exclusion criteria), used reference standard, and the biomarker(s) studied (index test).

Risk of bias assessment

The methodological quality of the full-text articles was independently evaluated by four reviewers (SR, ER, RV, MH) by utilising the QUADAS-2 tool.\(^17\) This tool was used to determine the ROB within four domains: patient selection, index test, reference standard, and flow and timing. Based on the information provided in the included studies, the ROB was rated low, intermediate, or high for these domains separately.

For the reference standard domain the ROB was rated low if (exercise) right-sided heart catheterisation was used for the diagnosis of HFpEF, intermediate if signs/symptoms of heart failure with left ventricular ejection fraction \(\geq 40–50\)% and structural/functional abnormalities indicative of LVDD was used,\(^18–21\) and high for all other reference standards. Within the remaining domains the ROB was rated low, intermediate or high when respectively all, two, and one or none of the supporting questions (online supplementary Table S1; three pre-defined questions per domain) were answered in a positive manner. However, certain study characteristics—no avoidance of case-control/two-gated designs, or unclear/inappropriate timing for the index test and/or reference standard—would immediately lead to a high ROB for the respective domain. Inconsistencies in quality assessment between the four reviewers were resolved by discussion until consensus was reached.

Conclusion

The majority of current diagnostic HFpEF biomarker studies have a high ROB, reducing the reproducibility and the potential for clinical care. Methodological well-designed studies with a uniform reference diagnosis are urgently needed to determine the incremental value of circulating biomarkers for the diagnosis of HFpEF.

Keywords

Heart failure with preserved ejection fraction • Diagnosis • Biomarker • Bias • QUADAS-2
<table>
<thead>
<tr>
<th>Study/country</th>
<th>Biomarkers</th>
<th>Cases (reference standard)</th>
<th>Controls</th>
<th>Cases/controls descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Study/country</td>
<td>Biomarkers</td>
<td>Cases (reference standard)</td>
<td>Controls</td>
<td>Cases/controls descriptives</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Baessler, 2012</td>
<td>GDF-15</td>
<td>LVDD with possible HF (n = 88)</td>
<td>No LVDD (n = 119)</td>
<td>Age (years): 50 ± 7/41 ± 12&lt;br&gt;Sex (% female): 55/73&lt;br&gt;NT-proBNP: 52 [29–96]/42 [25–66]&lt;br&gt;LVEF (%): 64 ± 9/64 ± 7&lt;sup&gt;a&lt;/sup&gt;&lt;br&gt;EF/E&lt;sub&gt;c&lt;/sub&gt;: 8 ± 3/5 ± 1&lt;sup&gt;b&lt;/sup&gt;&lt;br&gt;LAVI&lt;sup&gt;c&lt;/sup&gt;: 136 ± 32/102 ± 20</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mason, 2013</td>
<td>Copeptin; hsCRP; MR-proANP; MR-proADM</td>
<td>HFpEF (n = 57)</td>
<td>No HF (n = 308)</td>
<td>Age (years): 87 ± 6/84 ± 7&lt;br&gt;Sex (% female): 83/73&lt;br&gt;NT-proBNP: 1300 ± 160/764 ± 1280&lt;br&gt;LVEF (%): 68 ± 7/68 ± 7&lt;sup&gt;a&lt;/sup&gt;&lt;br&gt;EF/E&lt;sub&gt;c&lt;/sub&gt;: 12 ± 4/6 ± 1&lt;sup&gt;b&lt;/sup&gt;&lt;br&gt;LAVI&lt;sup&gt;c&lt;/sup&gt;: --&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>England</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang, 2013</td>
<td>sST2</td>
<td>HFrEF (n = 68)</td>
<td>No symptoms/signs HF (n = 39)</td>
<td>Age (years): 68 ± 10/60 ± 12&lt;br&gt;Sex (% female): 54/33&lt;br&gt;NT-proBNP: 262 ± 470/71 ± 53&lt;br&gt;LVEF (%): 68 ± 7/68 ± 7&lt;sup&gt;a&lt;/sup&gt;&lt;br&gt;EF/E&lt;sub&gt;c&lt;/sub&gt;: --&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jang, 2014</td>
<td>Angiogenin</td>
<td>HFrEF (n = 16)</td>
<td>Healthy controls (n = 16)</td>
<td>Age (years): 76 ± 4/68 ± 8&lt;br&gt;Sex (% female): 62/38&lt;br&gt;NT-proBNP: 3377 ± 2178–3995/55 [27–93]&lt;br&gt;LVEF (%): 55 ± 12/70 ± 4&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wong, 2015</td>
<td>Miscellaneous miRNAs</td>
<td>HFpEF (n = 30)</td>
<td>No history of CAD/HF (n = 30)</td>
<td>Age (years): 64 ± 9/66 ± 7&lt;br&gt;Sex (% female): --&lt;br&gt;NT-proBNP: 1712 (± 263)/86 (± 83)&lt;br&gt;LVEF (%): 59 ± 5/64 ± 4&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Singapore</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zordoky, 2013</td>
<td>Miscellaneous metabolites</td>
<td>HFpEF (n = 24)</td>
<td>Healthy controls and patients at risk (n = 38)</td>
<td>Age (years): 68 ± 5/65 ± 7&lt;br&gt;Sex (% female): --&lt;br&gt;NT-proBNP: 1712 (± 263)/6727 (± 6290)&lt;br&gt;LVEF (%): 59 ± 5/25 ± 7&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Reference standard; <sup>b</sup> LVEF < 50%; <sup>c</sup> LVEF ≥ 50%.
<table>
<thead>
<tr>
<th>Study/country</th>
<th>Biomarkers</th>
<th>Cases (reference standard)</th>
<th>Controls</th>
<th>Cases/controls descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watson, 2015&lt;sup&gt;37&lt;/sup&gt; Ireland</td>
<td>Miscellaneous miRNAs</td>
<td>HFrEF (n = 75)</td>
<td>HFrEF &lt;50% (n = 75)</td>
<td>Age (years)</td>
</tr>
<tr>
<td>Sanders-van Wijk, 2015&lt;sup&gt;34&lt;/sup&gt; Switzerland and Germany</td>
<td>Cys-C; Hb; hsCRP; hsTnT; sST2</td>
<td>HFrEF (n = 112)</td>
<td>HFrEF ≤40% (n = 458)</td>
<td></td>
</tr>
<tr>
<td>Barroso, 2016&lt;sup&gt;35&lt;/sup&gt; Germany</td>
<td>IGFBP-7; IGF-1</td>
<td>HFrEF (n = 77)</td>
<td>No LVDD, LVEF &gt;50% (n = 55)</td>
<td></td>
</tr>
<tr>
<td>Liu, 2016&lt;sup&gt;36&lt;/sup&gt; China</td>
<td>sgp130; hsP2; CTSS; DPP4</td>
<td>HFrEF (n = 50)</td>
<td>No history of heart disease(s) (n = 50)</td>
<td></td>
</tr>
<tr>
<td>Polat, 2016&lt;sup&gt;37&lt;/sup&gt; Turkey</td>
<td>Gal-3</td>
<td>HFrEF (n = 44)</td>
<td>No systolic/diastolic dysfunction (n = 38)</td>
<td></td>
</tr>
<tr>
<td>Li, 2016&lt;sup&gt;38&lt;/sup&gt; China</td>
<td>Aq-Ca</td>
<td>HFrEF (n = 104)</td>
<td>No HFrEF (n = 701)</td>
<td></td>
</tr>
<tr>
<td>Berezin, 2016&lt;sup&gt;39&lt;/sup&gt; Ukraine</td>
<td>CD31+/annexin V+ EMPs to CD14+/CD309+ cell ratio</td>
<td>HFrEF (N = 79)</td>
<td>HFrEF ≤45% (n = 85)</td>
<td></td>
</tr>
<tr>
<td>Study/country</td>
<td>Biomarkers</td>
<td>Cases (reference standard)</td>
<td>Controls</td>
<td>Cases/controls descriptives</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
</tbody>
</table>
| Toma, 2017\(^{70}\) Canada | Miscellaneous proteins and transcripts | HFpEF (n = 21)  
  - Symptoms consistent with HF  
  - LVEF ≥50% | HFpEF ≤40% (n = 48)  
  - Symptoms consistent with HF  
  - LVEF ≥50% | Age (years)  
  Sex (% female)  
  NT-proBNP\(^{a}\) (pg/mL)  
  LVEF (%)

| Sinning, 2017\(^{41}\) Germany | GDF-15; sST2; CRP | HFpEF (n = 70)  
  - NYHA II–IV or treatment for HF  
  - LVEF ≥50%  
  - LVDD | HFpEF ≤50%, NYHA II–IV or treatment for HF (n = 38) | Age (years)  
  Sex (% female)  
  NT-proBNP\(^{a}\) (pg/mL)  
  LVEF (%)

| Cui, 2018\(^{22}\) China | Gal-3; sST2 | HFpEF (n = 172)  
  - HFpEF ESC, 2016\(^{19}\) | Healthy controls (n = 52) | Age (years)  
  Sex (% female)  
  NT-proBNP\(^{a}\) (pg/mL)  
  LVEF (%)

| Nikolova, 2018\(^{33}\) America | cBIN1 | HFpEF (n = 52)  
  - History of fluid overload, prior HFH, or invasive evidence of elevated cardiac filling pressures  
  - LVEF ≥50% | Controls at risk (n = 52) | Age (years)  
  Sex (% female)  
  NT-proBNP\(^{a}\) (pg/mL)  
  LVEF (%)
<table>
<thead>
<tr>
<th>Study/country</th>
<th>Biomarkers</th>
<th>Cases (reference standard)</th>
<th>Controls</th>
<th>Cases/controls descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age (years)</td>
</tr>
<tr>
<td>Farinacci, 2019(^{44}) Germany</td>
<td>CECs</td>
<td>HFrEF (n = 27)</td>
<td>Healthy Controls (n = 10)</td>
<td><strong>69 ± 8.56 ± 3</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• NYHA I–III</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• HFH during last year</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cardiac functional/structural abnormalities suggestive for HFrEF or elevated NP levels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wong, 2019(^{45}) Singapore and New Zealand</td>
<td>Miscellaneous miRNAs</td>
<td>HFrEF (n = 179)</td>
<td>HFrEF ≤40% (n = 145)</td>
<td><strong>77 ± 9.70 ± 14</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Symptomatic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LVEF ≥50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi, 2019(^{46}) China</td>
<td>CTGF; TGF-(\beta)</td>
<td>DHF (n = 114)</td>
<td>No HF (n = 72)</td>
<td><strong>71 ± 11.69 ± 11</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Symptoms or signs of HF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LVEF ≥45% and normal LV size</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Structural heart disease such as LVH, left atrial enlargement, previous myocardial infarction and/or diastolic dysfunction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berezin, 2019(^{47}) Ukraine</td>
<td>CD31+/annexin V+/MVs; Gal-3; GDF-15</td>
<td>HFrEF (n = 178)</td>
<td>HFrEF/HFrEF (n = 210)</td>
<td><strong>55 ± 7.57 ± 7</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Previously treated primary diagnosis of HF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LVEF ≥50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fang, 2019(^{48}) China</td>
<td>RDW</td>
<td>HFrEF (n = 62)</td>
<td>I. No substantial cardiac dysfunction</td>
<td><strong>74 ± 9.67 ± 12</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Symptoms or signs of HF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LVEF ≥50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LAVI ≥34</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• NT-proBNP ≥400 ng/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>II. Possible HFrEF (n = 107)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2020 The Authors. European Journal of Heart Failure published by John Wiley & Sons Ltd on behalf of European Society of Cardiology.
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Study/country</th>
<th>Biomarkers</th>
<th>Cases (reference standard)</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Age (years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(n = 87)</td>
</tr>
<tr>
<td>Merino-Merino, 2020*</td>
<td>None</td>
<td>Non-reduced HF</td>
<td>64 ± 9/59 ± 10</td>
</tr>
<tr>
<td>Spain</td>
<td>Urate; CRP; TN; Fibrogen;</td>
<td>No non-reduced HF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gal-3; sST-2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only the number of subjects are shown for the validation cohort if multiple cohorts were used in one study; if multiple validation cohorts were used, only the cohort with the most included patients are shown. If only a sub-population in an article was used to determine the diagnostic value of a circulating biomarker, then only the information of this population is provided. To ensure readability, in some cases inclusion criteria were incorporated in the reference standard if they included LVEF, previous HF, symptoms/signs, or LVDD. More details about the study population—and the used exclusion criteria—can be found in online supplementary Table S4. If the mean and SD of one of the ‘Case controls descriptive’ were not directly provided in the article, a pooled mean and SD was calculated if possible. Descriptors are expressed as mean ± SD, median (IQR), mean (±SEM), or mean (95% CI).

A Co, albumin adjusted calcium; AF, atrial fibrillation; AS, American Society of Echocardiography; CAD, coronary artery disease; CMBN1, cardiac bridging integrator 1; CD, cluster of differentiation; CEC, circulating endothelial cell; CHF, chronic heart failure; CI, confidence interval; ODP, carboxy-terminal telopeptide of collagen type I; CCR, C-reactive protein, CTGE, connective tissue growth factor; CTP, cardiotropin-1; CTSS, cathepsin S; Cys-C, cystatin C; DHE, diastolic heart failure; DPP4, dipetidyl peptidase 4; E/æ', ratio of peak early mitral inflow velocity to early diastolic mitral annular velocity; EMP, endothelial cell-derived microparticle; ESC, European Society of Cardiology; Gal-3, galectin-3; GDF-1, growth differentiation factor-15; Hb, haemoglobin; HF, heart failure; Hf, heart failure hospitalisation; HfHF, heart failure with mid-range ejection fraction; HfHFH, heart failure with preserved ejection fraction; HfHFH, heart failure with reduced ejection fraction; HfP, heart failure preexisting hypertension; HfP, heart failure with reduced ejection fraction; HfP, heart failure with preserved ejection fraction; HL, left ventricle; LVEF, left ventricular ejection fraction; LVMI, left ventricular mass index (g/m²); LV, left ventricle; LVMi, left ventricular mass index (g/m²); miRNA, microRNA; MMP, matrix metalloproteinase; MR-proADM, mid-regional pro adrenomedullin; MR-proBNP, mid-regional proatrial natriuretic peptide; MV, microsirve; MIR, inner natriuretic peptide; MR-proBNP, N-terminal pro brain natriuretic peptide; NYHA, New York Heart Association; PICP, carboxy-terminal propeptide of procollagen type I; PINP, amino-terminal propeptide of procollagen type I; RED, red cell distribution width; SD, standard deviation; SEM, standard error of the mean; s1P13Q, soluble glycoprotein 130; sRAGE, soluble receptor for AGE; sRAGE, soluble receptor for AGE; TgF-β, transforming growth factor-β; TIMP, tissue inhibitor of matrix metalloproteinase; TR, trypsin-1; TRNP, TnT, troponin T; ULN, upper limit of normal.

* Of brain natriuretic peptide if stated.

** Not available.

Quality assessment

All papers had at least one domain within both the QUADAS2 (39%) showed a high ROB within all four domains (Supplementary Table S4). The ROB within the patient selection domain was high in 24 out of 28 studies (86%; Figure 7). This was mainly driven by the use of case-control/two-gated designs in all studies. This domain is a critical part of the QUADAS2 tool as it assesses the risk of selection bias caused by the fact that the index test and/or reference standard were chosen based on the results from a previous study. This was often the result of excluding difficult cases, such as patients with AF, obesity, and/or pulmonary hypertension. In all studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed in a laboratory without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values. In most studies, the index test was performed without pre-specification of cut-off values.
Risk of bias in studies investigating novel diagnostic HFpEF biomarkers

Figure 1 Percentage of studies with low, intermediate or high risk of bias within the four QUADAS-2 domains (patient selection, index test, reference standard, flow and timing) and the main reasons for a high risk of bias within these domains.

Objective: To assess the risk of bias in studies investigating novel diagnostic HFpEF biomarkers.

Methods: We conducted a systematic review of studies on novel HFpEF biomarkers published in English between 2000 and 2016. The risk of bias was assessed using the QUADAS-2 tool.

Results: Of the 45 included studies, 35% had a high risk of bias. The main reasons for a high risk of bias were unclear timing of the index and/or reference standard, no use of (exercise) right-sided heart catheterisation, no use of pre-specified cut-off values, no external validation, and use of case-control/two-gated designs.

Discussion: The overall high risk of bias might play an important role in the limited uptake of these biomarkers in the HFpEF clinics and calls for methodologically well-designed studies.

Patient selection
Most studies determined the diagnostic value of the biomarkers in cases with known HFpEF compared to (healthy) controls. During the early stages of novel biomarker discovery, these designs with contrasting populations can be useful to screen whether novel biomarkers might be of any interest for future analysis. Such studies may also reveal mechanistic insights into the syndrome. However, for diagnostic utility, these designs induce spectrum bias, which overestimates the diagnostic value of the investigated biomarker(s).

Index test
The use of optimal cut-off values for the index test without performing external validation within the majority of previous studies will have resulted in an overestimation of the diagnostic performance of the biomarkers examined. Moreover, a biomarker should have incremental value on top of easy to determine characteristics—e.g., age, sex and body mass index—to really yield potential for clinical use. While this was not part of the ROB assessment within this study, it will partially explain the lack of the implementation of novel diagnostic HFpEF biomarkers.

Reference standard
Test accuracy of a novel biomarker is based on the concept that every inconsistency between the index test and reference standard is due to an incorrect index test. Since different reference standards will significantly alter the prevalence of cases within the cohort of interest—as already shown within the field of LVDD—this will significantly affect the diagnostic value of the biomarker(s) studied. None of the included studies used (exercise) right-sided heart catheterisation—the real gold standard for HFpEF—as uniform reference standard. Studies validating the biomarker value against this gold standard are urgently needed.

Recognising the challenges of widespread implementation of gold standard invasive haemodynamic testing, we also examined the use of guideline-recommended reference standards that were published at the moment of publication for the diagnosis of heart failure with normal ejection fraction since 2007 or HFpEF since 2016, and found that most studies did not apply these. Also, these reference standards were not in line with the recently published H2FPEF or HFA-PEFF scores. Nonetheless, even

© 2020 The Authors. *European Journal of Heart Failure* published by John Wiley & Sons Ltd on behalf of European Society of Cardiology.
the recommended reference standards and risk scores differ significantly in included diagnostic criteria, used cut-off values and the role comorbidities play within these standards, highlighting the uncertainty of diagnosing HFpEF.

Flow and timing
Most studies did not provide (detailed) information regarding the timing of the index test and the reference standard. This lack of information is regrettable given that biomarker levels will likely change over time. Moreover, it is highly likely that diuretics are prescribed and/or dosage were changed in patients with signs of congestion. Diuretics will reduce filling pressure and very likely influence the concentration of the circulating biomarker measured. It has already been shown that diuretics affect the urinary proteome in rats, and the pleural protein concentration in patients with congestive heart failure. In the latter also an increase in total serum protein content after the administration of diuretics was observed. Therefore, it is highly desirable that the circulating biomarkers are measured at the same moment as the HFpEF diagnosis is made and before any intervention occurs.

Phenotype specific biomarkers
The question remains to which extent the absence of novel diagnostic HFpEF biomarkers is due to the real lack of diagnostic value of these biomarkers, vs. the heterogeneity of the syndrome itself. In contrast to HFpEF, heart failure with reduced ejection fraction, characterised by cardiomyocyte loss and ventricular dilatation, is diagnostically well-captured by natriuretic peptides that increase in response to wall stress and by troponins indicating cardiomyocyte injury. In the more heterogeneous HFpEF syndrome, biomarkers likely reflect less well the complex, mainly non-cardiac multi-organ nature of the syndrome. Therefore, biomarkers reflecting more general pathophysiological processes like inflammation (growth differentiation factor-15), fibrosis (soluble ST2, galectin-3), and metabolic dysfunction (insulin-like growth factor binding protein-7) could have potential; moreover, the search for one single biomarker may not be sufficient. An approach with multiple biomarkers in methodologically well-designed studies may be more appropriate and successful. One may postulate if it will ever be possible to find a single diagnostic test or panel of biomarkers with adequate diagnostic value for the entire syndrome, and perhaps the optimal approach may be to use specific biomarkers to diagnose distinct subtypes of HFpEF, which could eventually also lead to a more tailored therapy.

Future perspectives
There is an urgent need for prospective studies to validate the diagnostic value of the HFA-PEFF score against gold standard invasive exercise haemodynamic testing in unselected symptomatic patients with suspected HFpEF. The inclusion of blood biomarker testing in such a study will enable the evaluation of the possible role of novel biomarkers in the HFA-PEFF algorithm on top of NPs and echocardiographic biomarkers. Possibilities that warrant investigation include implementation of biomarker testing in step 1 (pre-test assessment) or step 2 (diagnostic work-up) of the HFA-PEFF algorithm. Furthermore, promising novel biomarkers may be assessed as potential alternatives to NPs. NP levels should not be used as a selection criterion in these studies since 18% to 30% of patients with haemodynamically proven HFpEF have NP levels below ‘diagnostic’ threshold. Such studies will require close collaboration between basic scientists, clinicians, epidemiologists, industry, and (federal) sponsors.

Study limitations
Although all papers were reviewed and discussed by our interdisciplinary team until consensus was reached, the ROB classifications are based on the information provided in the studies, the pre-defined risk of bias criteria, as well as on the interpretation of the reviewers themselves. Therefore, it is possible that analysis of the studies by another group of reviewers results in another level of bias within certain domains of studies. However, we defined clear roles and results are rather uniform and unambiguous, making it highly unlikely that the main conclusion would differ significantly. Our review did not aim for a head to head comparison between these studies, and therefore should not be used for this purpose.

To the best of our knowledge, this review includes all current novel diagnostic circulating biomarker studies to detect chronic HFpEF. However, given the extent of the search performed, it cannot be completely excluded that studies were missed if diagnostic performance measures were not mentioned in the abstract. Additionally, the main aim of some studies was not to study the diagnostic value of circulating biomarkers to detect HFpEF, though since they studied the diagnostic value in sub-analysis, they were still included in this review to provide a complete overview of current circulating diagnostic HFpEF biomarker analysis.

Finally, since some studies included (previous) hospitalised patients and timing of the reference standard and the drawing of blood was often unclear, we may have unintentionally included acute HFpEF populations. Since this does not affect the main conclusion of this review, we decided not to exclude these studies.

Conclusion
The majority of current diagnostic HFpEF biomarker studies have a high ROB, reducing the reproducibility and the potential for clinical care. Methodological well-designed studies with a uniform reference diagnosis are urgently needed to determine the incremental value of circulating biomarkers for the diagnosis of HFpEF.

Supplementary Information
Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Search string for PubMed and EMBASE.
Figure S1. PRISMA flow diagram of study selection.

© 2020 The Authors. European Journal of Heart Failure published by John Wiley & Sons Ltd on behalf of European Society of Cardiology.
Table S1. Predefined questions that were used for the risk of bias assessment.

Table S2. Overview risk of bias within the QUADAS-2 domains.

Table S3. Main determinants of level of bias within the QUADAS-2 domains for the articles included in this review.

Table S4. Overview of the patient population and exclusion criteria of the diagnostic HFpEF circulating biomarker studies included in this review.

Funding

This work was supported by the European Union Commission’s Horizon 2020, and IMI2-CARDIATEAM [N°821508]. We acknowledge the support from the Netherlands Cardiovascular Research Initiative: an initiative with support of the Dutch Heart Foundation, CVON2015-RECONNECT, CVON2016-Early HFpEF and CVON 2017-ShePREDICTS. Additionally, J.W.v.B. is supported by a ZonMw VIDI grant. Conflict of interest: none declared.

References


