Technoleg Iechyd Cymru
Health Technology Wales

Evidence Appraisal Report

Natriuretic peptides to *rule-in* and *rule-out* a diagnosis of acute heart failure in adults in the emergency department setting

1. Purpose of the evidence appraisal report

This report aims to identify and summarise evidence that addresses the following question: what is the clinical and cost effectiveness of N-terminal pro B-type natriuretic peptide (NT proBNP) or B-type natriuretic peptide (BNP) as a diagnostic test to *rule-in* and *rule-out* acute heart failure (AHF) in adults in the emergency department (ED) setting, compared with NT proBNP, BNP, or standard care to *rule-out* AHF?

Evidence Appraisal Reports are based on rapid systematic literature searches, with the aim of identifying the best published clinical and economic evidence on health technologies. Researchers critically evaluate this evidence. The draft Evidence Appraisal Report is reviewed by experts and by Health Technology Wales multidisciplinary advisory groups before publication.

2. Health problem

Heart failure is a progressive medical condition in which the heart is unable to pump blood around the body properly (NICE 2014). As the heart is not able to supply enough oxygen to the body, heart failure often leads to shortness of breath (dyspnoea) and fatigue, and can also lead to fluid retention in the chest cavity, legs and ankles. AHF can present suddenly as a new diagnosis in people with no previous symptoms or as worsening of chronic heart failure (CHF). This can be due to damage to the heart muscle, a heart valve not functioning properly, an abnormal heart rhythm and other rarer causes (NICE 2014).

AHF can be challenging to diagnose, as symptoms and signs often overlap with other conditions, leading to both under- and over-recognition, and incorrect treatment (Mueller et al. 2019). Heart failure occurs predominantly in older people and its presentation can be complicated by multiple co morbidities (Mueller et al. 2019).

As of 2019, over 32,000 people in Wales were diagnosed with heart failure. AHF is the most common reason for emergency admission to hospital in England and Wales in adults older than 65 years, accounting for 5% of all emergency medical admissions (NHS Wales 2019). The average age of patients diagnosed with AHF in hospital in the UK is 78 years old (National Institute for Cardiovascular Outcomes Research 2020). Heart failure is a rising epidemic due to the ageing population (Magnussen & Blankenberg 2018).

3. Health technology

There is no single diagnostic test for heart failure, and diagnosis relies on clinical judgement based on a combination of history, physical examination and appropriate investigations (for example, electrocardiography [ECG], chest X-ray and blood tests) (NICE 2018). According to National Institute for Cardiovascular Outcomes Research (2020), echocardiography is the 'gold standard' diagnostic tool for heart failure, but there is variation in its use in hospitals in the UK.

Biomarker tests can assist with diagnosis of AHF. Biomarkers are substances measurable in the blood stream which can be used to diagnose and monitor disease. B-type natriuretic peptide (BNP) is a hormone released from the heart when the walls of the heart are stretched or there is pressure overload. N-terminal pro B-type natriuretic peptide (NT-proBNP) is a non-active prohormone that is released from the same molecule that produces BNP. BNP and NT-proBNP are classed as natriuretic peptides.

Testing of blood for the presence of natriuretic peptides facilitates the diagnosis of AHF because NT-proBNP and BNP levels are generally elevated in individuals with AHF. However, other factors are also associated with differing levels of these natriuretic peptides, which can complicate the interpretation of test results. For example, atrial fibrillation, age and renal failure are also associated with higher levels of NT-proBNP and BNP. On the other hand, unexpectedly low levels of natriuretic peptides can be detected in some people with decompensated end-stage heart failure, pulmonary oedema, or right-sided AHF, but this is also the case for people who are obese (Ponikowski et al. 2016).

Based on expert opinion, there is large variation across the UK for use of natriuretic peptides in the ED. National Institute for Health and Care Excellence (NICE) AHF guidelines recommend BNP or NT-proBNP thresholds to *rule-out* AHF, but do not make any recommendations for *ruling-in* AHF, even though they included *rule-in* and *rule-out* thresholds in their literature search protocol (NICE 2014) (see section 3.1.1 UK guidance [of this EAR] for more information). Expert opinion states that some EDs in Wales use NT-proBNP to *rule-out* AHF, and some large tertiary centres in the UK use it to *rule-in* AHF. Some health boards in Wales also currently use NT-proBNP within their CHF pathway.

Expert reviewers described that patients with natriuretic peptide levels below the *rule-out* threshold can be excluded for AHF. When different thresholds are applied for ruling-in and ruling-out AHF, there is an inherent 'grey zone' for patients that fall between the two cut-offs. Experts were of the opinion that patients with natriuretic peptide levels in the 'grey zone' need careful evaluation for a possible diagnosis of heart failure and alternate causes for natriuretic peptide elevation. This requires judgement from the clinician as not all of these patients will have AHF, and so there is the risk of over-investigation and excessive demand for echocardiography. For patients with levels above the *rule-in* threshold, expert reviewers believed it would be reasonable to treat them with intravenous diuretics empirically and then give an echocardiogram to confirm the diagnosis. This strategy allows for better targeting of investigations. Earlier confirmation of the diagnosis of heart failure may facilitate more rapid initiation of specific treatment, potentially shortening hospital stay and reducing mortality (Roberts et al. 2015).

We identified ten manufacturers of commercially available diagnostic NT-proBNP automated immunoassays, and six manufacturers of commercial BNP assays, some of which are analysed in the hospital laboratory and some using point-of-care (POC) devices (listed in Appendix 1). Expert reviewers mentioned that four health boards in Wales had access to the Roche Elecsys/Cobas NT-proBNP method and two health boards had access to the Abbott Alinity NT proBNP method. It is unclear whether BNP assays are being used in Wales. Expert reviewers were of the opinion that NT-proBNP is more stable than BNP and is therefore the most commonly used assay in the UK.

3.1 Guidance

3.1.1 UK guidance

HTW researchers did not identify any UK guidance providing recommendations regarding the use of NT-proBNP to *rule-in* AHF.

NICE guidelines recommend that in people presenting to acute settings with newly suspected AHF, a single serum measurement of BNP or NT-proBNP can be used to *rule-out* the diagnosis of heart failure if BNP is less than 100 nanograms per litre (ng/L) or NT-proBNP is less than 300 ng/L. If the levels are raised above these thresholds, then the recommendation is to perform transthoracic Doppler 2D echocardiography to establish the presence or absence of cardiac abnormalities (NICE 2014).

NICE AHF guidelines noted that cut-offs may vary for different age groups but did not give recommendations on this because studies "included patients with a very narrow age range and did not divide patients by the type of medication they were receiving" (NICE 2014).

3.1.2 International guidance

European Society of Cardiology (ESC) guidelines on AHF state that natriuretic peptides have a high sensitivity, and that normal levels of BNP and NT-proBNP in patients with suspected AHF mean a diagnosis is not likely to be appropriate (thresholds: BNP less than 100 ng/L and NT proBNP less than 300 ng/L). However, elevated levels of natriuretic peptides may not be sufficient to confirm the diagnosis of AHF, as they may also be associated with a wide variety of cardiac and non-cardiac causes (Ponikowski et al. 2016).

A Position Statement from ESC (Mueller et al. 2019), based on data from the PRIDE (Januzzi et al. 2005)) and ICON (Januzzi et al. 2006) studies, recommends:

- an NT-proBNP *rule-out* cut-off of 300 ng/L regardless of age; and *rule-in* cut-offs of:
- 450 ng/L for people younger than 50 years
- 900 ng/L for those 50 to 75 years
- 1,800 ng/L for people older than 75 years, or:
- a BNP *rule-out* cut-off of 100 ng/L and a *rule-in* cut-off of 400 ng/L, regardless of age.

The ESC Position Statement states that NT-proBNP concentration correlates more strongly with age and renal dysfunction than BNP, and so is the preferred age-dependent *rule-in* biomarker (Mueller et al. 2019).

When different thresholds are applied for *ruling-in* and *ruling-out* AHF, there is an inherent 'grey zone' for patients that fall between the two cut-offs (Table 1) (Mueller et al. 2019). In acute dyspnoea, 'grey zone' natriuretic values are present in 20% of patients and about 50% of these will have AHF. Patients with levels in the 'grey zone' need extra physician attention and ancillary testing. The ESC Position Statement notes that due to the strong correlation between renal dysfunction and age, no additional adjustment seems necessary for NT-proBNP once using age adjusted *rule-in* cut-offs. To optimise diagnostic accuracy in obese patients, the ESC Position Statement recommends lowering the established cut-off concentrations by up to 50% (Mueller et al. 2019).

Table 1. Natriuretic peptide cut-off levels for patients with dyspnoea in the acute setting (Mueller et al. 2019)

Age (years)	Heart failure unlikely	'Grey zone'	Heart failure likely
NT-proBNP (ng/L)			
< 50	300	300 to 450	More than 450
50 to 75	300	300 to 900	More than 900
> 75	300	300 to 1,800	More than 1,800
BNP (ng/L)			
All ages	Less than 100	100 to 400	More than 400

BNP: B-type natriuretic peptide; ng/L: nanograms per litre; NT-proBNP: N-terminal pro B-type natriuretic peptide

4. Evidence search methods

We searched for evidence that could be used to answer the review question: what is the clinical and cost effectiveness of NT proBNP or BNP as a diagnostic test to *rule-in* and *rule-out* AHF in adults in the ED setting, compared with NT-proBNP, BNP or standard care to *rule-out* AHF?

The criteria used to select evidence for the appraisal are outlined in Appendix 2. These criteria were developed following comments from the Health Technology Wales (HTW) Assessment Group and UK experts.

A systematic literature search was undertaken between 21 October 2020 and 24 November 2020, updated between 8 April 2021 and 15 April 2021, and then updated again on 24 August 2021. The search was adapted from the search performed by NICE (which looked at *rule-in* and *rule-out* cut-offs) and was restricted from the year 2014 until present to exclude studies identified by NICE AHF guidelines. However, some of the studies included in the NICE AHF guideline (Breidthardt et al. 2007, Moe et al. 2007) are included in this EAR, as they were referenced in the NICE guideline but the clinical outcomes were not reported in it. NICE states that this guideline was checked in December 2017 and no new evidence was found affecting the recommendations (NICE 2014).

As levels of NT-proBNP and BNP can increase with age, HTW included studies which reported age specific *rule-in* cut-off points. For completeness, we also included studies that did not consider age as a variable affecting NT-proBNP *rule-in* and *rule-out* values. We searched for evidence on BNP and NT-proBNP use in children but did not identify any studies for diagnosing AHF in the ED for this population.

The diagnostic values and cut-off thresholds for use of natriuretic peptides in people with suspected heart failure in the outpatient setting are usually associated with different sensitivity and specificity values, and so we only included studies with patients presenting to the ED (some of the studies included in the secondary evidence also looked at other acute settings), most of which had acute dyspnoea.

Appendix 3 (i.e. the PRISMA diagram) summarises the selection of articles for inclusion in the review.

5. Clinical effectiveness

HTW researchers reviewed the evidence in the NICE AHF guideline (NICE 2014) and reported outcomes in this EAR if we did not identify them in more recent evidence. Since publication of the NICE AHF guideline in 2014, there have been a number of published studies assessing the diagnostic accuracy and clinical outcomes of NT-proBNP and BNP testing, and the outcomes from these studies were also reported in this EAR.

Our rapid review identified two additional meta-analyses, published since NICE AHF guidelines, investigating the diagnostic accuracy of NT-proBNP and BNP (Hill et al. 2014, Roberts et al. 2015). The meta-analysis by Hill et al. (2014) looked at the age-specific *rule-in* and age-independent *rule-out* cut-offs for NT-proBNP, recommended in the ESC Position Statement and reported in Table 1 of this EAR (Mueller et al. 2019). They also assessed the impact of age on BNP levels: they reported different cut-off points based upon age, each using different reasoning and criteria to select the cut-off points (Hill et al. 2014). The meta-analysis by Roberts et al. (2015) did not take age into consideration as a factor affecting NT-proBNP or BNP levels. This meta-analysis was conducted as part of the development of the NICE AHF guideline (NICE 2014). The quality of the meta-analyses, using Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS 2), are summarised in Table 7; and the strengths, limitations and generalisability of studies are outlined this EAR under Section '5.3 Certainty and Quality of the Evidence.'

Because this is a rapid review, and well-conducted sources of secondary evidence exist that cover our review question, we have focused on these as sources of evidence. For completeness, we have also included primary studies published since the last search date of the reviews used, which was March 2015 (Ibrahim et al. 2017, Januzzi et al. 2018, Kozhuharov et al. 2019). However, we do not believe these fundamentally alter the findings of the secondary evidence sources and only inform the diagnostic accuracy of NT-proBNP. We have therefore reported these studies within Appendix 4 and provided only brief commentary in the main text on how they impact on the findings of the secondary evidence.

HTW researchers included two observational studies investigating the diagnostic accuracy of the 'grey area' created when using a *rule-in/rule-out* combination cut-off for NT-proBNP measurements (Bombelli et al. 2015, Darche et al. 2017). These primary studies were not included in the NICE AHF guideline or the two meta-analyses. We did not identify any studies investigating this 'grey area' for BNP measurements.

HTW researchers identified one meta-analysis reporting clinical outcomes from using NT proBNP/BNP measurements for AHF (Lam et al. 2010). We identified this review from a recent health technology assessment (Ontario Health 2021). Two additional randomised controlled trials (RCTs) (Breidthardt et al. 2007, Moe et al. 2007), which are referenced but clinical outcomes not reported on in the NICE AHF guidelines, are also included in this EAR, as they reported on outcomes not reported in the meta-analyses by Lam et al. (2010).

5.1 Diagnostic accuracy

5.1.1 NT-proBNP

All of the secondary evidence we identified reported the diagnostic test accuracy of NT-proBNP using separate *rule-in* and *rule-out* thresholds. We only identified two observational studies using a *rule-in* and *rule-out* test in combination (Bombelli et al. 2015, Darche et al. 2017)

5.1.1.1 Secondary evidence

Four of the studies in the meta-analysis by Hill et al. (2014) used a *rule-out* cut-off of 300 ng/L and *rule-in* cut-offs of 450 ng/L (less than 50 years old); 900 ng/L (50 to 75 years old); and 1,800 ng/L (older than 75 years). Using these thresholds, the overall *rule-in* estimate for sensitivity was 88%

(95% confidence interval [CI]: 84% to 91%), for specificity was 73% (95% CI: 64% to 82%), for the positive likelihood ratio (LR) was 3.53 (95% CI: 2.41 to 5.19), and negative LR was 0.18 (95% CI: 0.13 to 0.29). (Table 2). Out of these four studies, the meta-analysis only included the *rule-out* data for one of them (Januzzi et al. 2006): sensitivity was 99% and specificity was 60%.

Two studies in the meta-analysis by Hill et al. (2014) used a *rule-out* cut-off of 300 ng/L and a *rule-in* cut-off of 450 ng/L (less than 50 years) and 900 ng/L (older than 50 years). For one of these studies, the overall *rule-in* estimate for sensitivity was 96.2% (95% CI: 89% to 99%), for specificity was 80.95% (95% CI: 58% to 94%), for the positive LR was 5.05 (95% CI: 2.08 to 12), and for the negative LR was 0.046 (95% CI: 0.015 to 0.058). The *rule-out* sensitivity was 100% (95% CI: 95% to 100%, specificity was 43% (95% CI: 21% to 65%), positive LR was 1.75 (95% CI: 1.2 to 2.53) and negative LR was 0 (Shaikh & Ahmad 2011). For the other study in the meta-analysis, the overall *rule-in* estimate for sensitivity was 90% and for specificity was 85%, and overall *rule-out* sensitivity was 99% and specificity 60% (Januzzi et al. 2005, Table 2). Pooled results of these studies were not reported in the meta-analysis; the results that were pooled included studies with *rule-out* thresholds other than 300 ng/L (Hill et al. 2014).

Two studies in the meta-analysis used a different *rule-out* threshold of 300 ng/L: (Chenevier-Gobeaux et al. 2008) examined people aged 85 years and older, using a *rule-out* cut-off of 1,750 ng/L and a *rule-in* cut-off of 6,000 ng/L. They reported a sensitivity of 85% and specificity of 59%, and sensitivity of 53%, for *rule-out* and *rule-in*, respectively. Berdagué et al. (2006) examined people aged 70 years and older, using a *rule-out* cut-off of 1,200 ng/L and a *rule-in* cut-off of 4,500 ng/L. They reported a sensitivity of 97% and specificity of 55%, and sensitivity of 64% and specificity of 86%, for *rule-out* and *rule-in*, respectively.

Table 2. Effect of age-optimised decision points on diagnostic performance of NT-proBNP to *rule-out* or *rule-in* AHF (Hill et al. 2014)

Reference in systematic review	Reference test	Index assay and analyser	*Rule-in* cut-off (ng/L)	*Rule-out* cut-off (ng/L)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ (95% CI)	LR- (95% CI)
Januzzi et al. (2006) n = 1,256	Clinical adjudication	Elecsys proBNP assay (Roche)	< 50 years (450) 50 to 75 years (900) > 75 years (1,800)		90 (88 to 92)	84 (81 to 87)	5.63 (4.63 to 6.84)	0.12 (0.10 to 0.15)
				All ages (300)	99	60	NR	NR
Liteplo et al. (2009) n = 94	Clinical adjudication	Elecsys proBNP assay (Roche)	< 50 years (450) 50 to 75 years (900) > 75 years (1,800)		85 (71 to 93)	63 (50 to 75)	2.29 (1.58 to 3.33)	0.24 (0.11 to 0.51)
Martinez-Rumayor et al. (2010) n = 599	Clinical adjudication	Elecsys proBNP assay (Roche)	< 50 years (450) 50 to 75 years (900) > 75 years (1,800)		90 (83 to 94)	86% (82 to 90)	6.43 (NR)	0.12 (NR)
Robaei et al. (2011) n = 68	Clinical adjudication	Elecsys proBNP assay (Roche)	< 50 years (450) 50 to 75 years (900) > 75 years (1,800)		81 (63 to 92)	66 (51 to 79)	2.38 (1.50 to 3.79)	0.29 (0.13 to 0.65)
Overall estimate n = 2,017	Clinical adjudication	Elecsys proBNP assay (Roche)	< 50 years (450) 50 to 75 years (900) > 75 years (1,800)		88 (84 to 91)	73 (64 to 82)	3.53 (2.41 to 5.19)	0.18 (0.13 to 0.29)
Shaikh & Ahmad (2011) n = 100	Clinical adjudication	Elecsys proBNP assay (Roche)	< 50 years (450)		100 (79 to 100)	33 (4 to 77)	1.49 (0.85 to 2.6)	0
			> 50 years (900)		97 (89 to 99)	87 (59 to 98)	7.26 (1.99 to 26)	0.036 (0.009 to 0.048)
			All patients overall (900)		96 (89 to 99)	81 (58 to 94)	5.05 (2.08 to 12)	0.046 (0.015 to 0.058)
				All ages overall (300)	100 (95 to 100)	43 (21 to 65)	1.75 (1.2 to 2.53)	0

Reference in systematic review	Reference test	Index assay and analyser	*Rule-in* cut-off (ng/L)	*Rule-out* cut-off (ng/L)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ (95% CI)	LR- (95% CI)
Januzzi et al. (2005) n = 599	Clinical adjudication	Elecsys proBNP assay (Roche)	All patients overall (900)		90 (CI NR)	85 (CI NR)	NR	NR
			< 50 years (450)		93 (CI NR)	95 (CI NR)	NR	NR
			≥ 50 years (900)		91 (CI NR)	80 (CI NR)	NR	NR
				All ages overall (300)	99 (CI NR)	68 (CI NR)	NR	NR
Chenevier-Gobeaux et al. (2008) n = 210	Clinical adjudication	Elecsys proBNP assay (Roche)		≥ 85 years (1,750)	85 (CI NR)	59 (CI NR)	NR	NR
			≥ 85 years (6,000)		53 (CI NR)	NR	NR	NR
Berdagué et al. (2006) n = 256	Clinical adjudication	Elecsys proBNP assay (Roche)		≥ 70 years (1,200)	97 (CI NR)	55 (CI NR)	2.77 (CI NR)	0.05 (CI NR)
			≥ 70 years (4,500)		64 (CI NR)	86 (CI NR)	4.57 (CI NR)	0.42 (CI NR)

CI: confidence interval; ED: emergency department; LR+: positive likelihood ratio; LR-: negative likelihood ratio; n: number of participants; ng/L: nanograms/litre; NR: not reported

Roberts et al. (2015) conducted a meta-analysis which didn't consider age as a factor affecting natriuretic peptide levels. When evaluating the sensitivity and specificity of NT-proBNP, the meta-analysis by Roberts et al. (2015) divided the thresholds into three age-independent groups (≤ 300, 300 to 1,800 [all studies using a threshold within these values], and ≥ 1,800 ng/L). At a cut-off of ≤ 300 ng/L, the pooled sensitivity and specificity was 99% and 43%, respectively. Furthermore, the PPV and NPV was 64% and 98%, respectively. When the threshold was 300 to 1,800 ng/L, the pooled sensitivity and specificity was 90% and 76%, respectively. Additionally, the PPV and the NPV value was 80% and 88%, respectively. Only three studies applied a threshold of ≥ 1,800 ng/L, and thus the factors were not pooled. The ranges of sensitivities with this threshold were between 60% to 87%, and the specificities were between 72% and 95%. At the lowest threshold (< 300 ng/L) sensitivity was consistently high. As the threshold increased, sensitivity decreased and specificity increased yet remained quite variable (Table 3).

Table 3. Diagnostic accuracy for cut-off thresholds for NT-proBNP (Roberts et al. 2015)

NP threshold (ng/L)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Heterogeneity: I^2 (%)
300	99 (97 to 100)	43 (26 to 62)	64 (57 to 73)	98 (89 to 100)	94
300 to 1,800 (all studies using thresholds within these values)	90 (86 to 93)	76 (69 to 82)	80 (74 to 84)	88 (82 to 92)	97
1,800	67 (60 to 73) to 87 (81 to 92)*	72 (63 to 80) to 95 (91 to 98)*	80 (73 to 86) to 94 (89 to 97)*	71 (65 to 76) to 82 (73 to 89)*	NR
*A sensitivity or specificity range is given where there were insufficient number of studies to conduct diagnostic meta-analysis and generate a pooled sensitivity and specificity value					
BNP: B-type natriuretic peptide; LR-: negative likelihood ratio; CI: confidence interval; LR+: positive likelihood ratio; n: number of participants in study; ng/L: nanograms per litre; NP: natriuretic peptide; NPV: negative predicitve value; NT-proBNP: NR: not reported; PPV: positive predictive value					

5.1.1.2 Primary evidence

Diagnostic accuracy

Because this is a rapid review, and well-conducted sources of secondary evidence exist that cover our review question, we have focussed on these as sources of evidence. For completeness, we have also included primary studies published since the last search date of the reviews used (March 2015). However, we do not believe these fundamentally alter the findings of the secondary evidence source and only inform the diagnostic accuracy of NT-proBNP. We identified three observational studies reporting on the diagnostic accuracy of NT-proBNP for AHF in the ED (Ibrahim et al. 2017, Januzzi et al. 2018, Kozhuharov et al. 2019). These studies and findings are described in detail in Appendix 4, and so we have only provided a brief commentary here on how they relate to the findings of the secondary evidence.

All three of the studies used the age-related cut-off NT-proBNP thresholds described in Table 1. Sensitivity for age-stratified cut-off points ranged from between 85.7% to 100% for 450 ng/L (less than 50 years); 79.3% to 91% for 900 ng/L (50 to 75 years old); and 75.9% to 87% for 1,800 ng/ml (older than 75 years). Specificity for age-stratified cut-off points ranged from between 76% to 91% for 450 ng/L; 75% to 84.0% for 900 ng/L; and 75.0% to 81% for 1,800 ng/ml. Positive predictive values (PPVs) for age-stratified cut-off points ranged from between 43% to 60% for 450 ng/L; 58% to 79% for 900 ng/L; and 62.0% to 90% for 1,800 ng/L. The sensitivity and negative predictive values (NPVs) for the *rule-out* cut-off of 300 ng/ml were 93.9% to 98%, and 96% to 98.0%, respectively. These findings support the secondary evidence that natriuretic peptide biomarker

levels usually have higher sensitivity than specificity, with sensitivity decreasing and specificity increasing as the natriuretic peptide threshold increases. This suggests that natriuretic peptides may be more useful for ruling-out than ruling-in AHF in the ED setting.

'Grey area'

Darche et al. (2017) conducted an observational study combining *rule-in* and *rule-out* thresholds for the diagnosis of AHF in 312 patients presenting to a German ED with acute dyspnoea. The study aimed to present the accuracy for the patient population as a whole and on a series of subsets of patients, including those who fall in a 'grey zone', where the relationship between level of NT-proBNP and AHF is less certain. The thresholds and 'grey zone' used in this study are in line with those described in Table 1, and only 14 out of 312 participants met this criteria and could be used in this analysis.

The sensitivity of NT-proBNP was 89.21% and 71.43% for the whole population and the 'grey zone' population, respectively. Specificity was 75.72% for the whole population compared to 82.61% for those in the 'grey zone' (Table 4). The study also explored whether presence of factors that may also influence levels of NT-proBNP influenced accuracy. They report that there are limited differences in accuracy for patients with severe kidney failure, advanced age, and obesity.

Table 4. Diagnostic performance of NT-proBNP for ruling-in and ruling-out AHF for all patients presenting to the ED and for those in the 'grey zone' (Darche et al. 2017)

Cohort	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC (95% CI)
All patients (n = 312)	89.21	75.72	74.70	89.70	0.892 (0.852 to 0.924)
NT-proBNP 'grey zone' (n = 14)	71.43	82.61	55.55	90.48	0.731 (0.601 to 0.838)

AUC: area under the curve; MR-proANP: mid-regional pro-A-type natriuretic peptide; NPV: negative predictive value; NT-proBNP: N-terminal pro B-type natriuretic peptide; PPV: positive predictive value

A study by Bombelli et al. (2015) compared the NT-proBNP cut-off thresholds of 300 ng/L for *rule-out* and 1,800 ng/L for *rule-in* to other thresholds in 895 patients aged 80 years and older in an Italian ED, using the Elecsys proBNP assay (Roche).

The study reported a sensitivity of 99%, and NPV 93% for the lower value of 300 ng/L; and specificity of 58% and PPV 63% for the higher value of 1,800 ng/L. The use of both thresholds resulted in a 'grey area' of 31.4% (n = 281), of whom 82% did not suffer heart failure. When the lower cut-off was changed from 300 ng/L to 980 ng/L, sensitivity slightly decreased to 95% and NPV to 91%. The use of a higher upper cut-off of 5,340 ng/L, instead of 1,800 ng/L, resulted in an increase in specificity to 85% and PPV to 76%. The use of these new thresholds also determined an increase in the number of patients falling in the diagnostic uncertainty area (42.4%).

In the attempt to reduce this 'grey area' and to maintain an acceptable overall diagnostic performance, the study proposed other threshold values, set at 1,470 ng/L (sensitivity and NPV decreased to 90% and 87%, respectively) and 4,200 ng/L (specificity and PPV increased to 80% and 72%, respectively). When compared to the 300 and 1,800 ng/L cut-offs, they obtained a relevant reduction in the number of patients in the diagnostic uncertainty area (27.4% [n=245]) (Bombelli et al. 2015).

5.1.2 BNP

5.1.2.1 Secondary evidence

For BNP, eight articles examined the relationship between age and BNP in the meta-analysis by Hill et al. (2014). Of these, four papers reported different cut-off points based upon age, each using different reasoning and criteria to select the cut-off points. The diagnostic accuracy data of these studies can be found in Table 5. In all cases, increasing age was associated with an increase in BNP concentration, but the association of age with the diagnostic performance of the test was not clear in the papers (Hill et al. 2014).

Table 5. Effect of age on diagnostic performance of BNP (Hill et al. 2014)

Reference in systematic review	Assay	Age (years)	cut-off (ng/L)	Sensitivity (%)	Specificity (%)	LR+	LR-
Maisel	Triage	18 to 69	100	86.3	81.6	4.69	0.17
			200	76.9	90.9	8.45	0.25
			300	68.8	93.8	11.10	0.33
			400	59.5	94.7	11.23	0.43
		70 to 105	100	93.6	53.3	2.00	0.12
			200	84.8	72.0	3.03	0.21
			300	75.3	77.0	3.27	0.32
			400	65.1	83.1	3.85	0.42
Ray	Triage	Greater than 65	250	73	91	8.11	0.30
Rogers	iSTAT	Greater than 75	100	94	41	1.59	0.15
			184	91	66	2.68	0.14
Chenevier-Gobeaux	Triage	65 to 84	270	73	83	4.29	0.33
		Greater than 85	250	85	64	2.36	0.23
			290	80	69	2.59	0.29
			380	70	73	2.59	0.41
			400	67	75	2.68	0.44
			500	60	79	2.86	0.51
Confidence intervals not reported in meta-analysis by Hill							
BNP: B-type natriuretic peptide; LR- : negative likelihood ratio; LR+ : positive likelihood ratio; ng/L: nanograms per litre;							

When a diagnostic meta-analysis was conducted by (Roberts et al. 2015), the pooled sensitivity, specificity, PPV and NPV of BNP at a threshold of ≤ 100 ng/L were 95% (95% CI: 93% to 96%), 63% (95% CI: 52% to 73%), 67% (95% CI: 63% to 75%), and 94% (95% CI: 90% to 96%), respectively. At a BNP level of 100 to 500 ng/L, the pooled sensitivity, specificity, PPV and NPV values were 85% (95% CI: 81% to 88%), 86% (95% CI: 79% to 91%), 85% (95% CI: 78% to 90%) and 86% (95% CI: 82% to 89%), respectively. As only four study cohorts reported data for BNP at a threshold of ≥ 500 ng/L, diagnostic meta-analysis was not performed and the reported sensitivity from the study cohorts ranged from 35% to 83% and the paired specificity from 78% to 100%. At the lowest threshold (< 100 ng/L), sensitivity was consistently high whereas specificity varied widely across all studies. The pattern of BNP mirrors that of NT-proBNP at different thresholds: that is, decreasing sensitivity and increasing specificity with increasing threshold (Table 6).

The meta-analysis suggested no statistically significant difference between NT-proBNP and BNP tests at the *rule-out* thresholds of ≤ 300 ng/L and ≤ 100 ng/L, respectively ($p > 0.05$).

When converted to absolute patient numbers, the use of BNP rather than NTproBNP in an acute care setting potentially increased the false-negative test results by between eight and 31 more people per 1,000 people. Sensitivity was similar and specificity only modest for both natriuretic peptides in the low- to intermediate-ranges of measured values. The study concluded that it is important that for values measured above the *rule-out* thresholds, the information is correlated with clinical and imaging assessment to confirm a diagnosis of heart failure and to exclude non-cardiac causes of an increased natriuretic peptide level.

Table 6. Diagnostic accuracy for cut-off thresholds for BNP (Roberts et al. 2015)

NP threshold (ng/L)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Heterogeneity: I^2 (%)
100	95 (93 to 96)	63 (52 to 73)	67 (63 to 75)	94 (90 to 96)	98
100 to 500 (all studies using thresholds within these values)	85 (81 to 88)	86 (79 to 91)	85 (78 to 90)	86 (82 to 89)	97
500	35 (17-56) to 83 (69-92)*	78 (56-93) to 100 (91-100)*	89 (75-96) to 100 (63-100)*	55 (69-80) to 69 (48-84)*	NR
*A sensitivity or specificity range is given where there were insufficient number of studies to conduct diagnostic meta-analysis and generate a pooled sensitivity and specificity value					
BNP: B-type natriuretic peptide; LR-: negative likelihood ratio; LR+: positive likelihood ratio; n: number of participants in study; ng/L: nanograms per litre; NP: natriuretic peptide; NPV: negative predicitve value; NT-proBNP: NR: not reported; PPV: positive predictive value					

5.2 Clinical outcomes

HTW researchers identified one meta-analysis by Lam et al. (2010), which was not included in the NICE AHF guideline, that investigated the effect of BNP or NT-proBNP testing on clinical outcomes in adults presenting to the ED with acute dyspnoea. The review included 2,513 patients from five RCTs (two out of five studies tested NT-proBNP, and three out of five studies tested BNP) and all comparator groups in the studies received usual care with no natriuretic peptide testing (see Table 7 for details of studies included in the meta-analysis). The natriuretic peptide cut-off thresholds were not reported in the meta-analysis, and neither was whether the cut-offs were used to *rule-in* or *rule-out* heart failure.

An RCT by (Moe et al. 2007) investigated how NT-proBNP testing affects the management of 500 patients presenting with dyspnoea to seven Canadian EDs. *Rule-in*/*rule-out* thresholds were based on manufacturer-supplied data, and later on those recommended in the ESC Position Statement and Table 1 of this EAR (Mueller et al. 2019)

Breidthardt et al. (2007) performed a follow-up analysis of a previously completed RCT involving 452 patients who presented to the ED with acute dyspnoea. Patients were randomly assigned to a group where BNP measurements were taken or to a group with standard assessment. The *rule-out* threshold was 100 ng/L and the *rule-in* threshold was 500 ng/L. The majority of patients with BNP concentrations between 100 and 500 ng/L were considered to have mild to moderate heart failure, but the protocol recommended clinical judgment and possible further diagnostic testing.

The study characteristics of each of these studies are described in more detail in Table 7 and Table 8.

5.2.1 Length of hospital stay

The meta-analysis by Lam et al. (2010) reported that overall, there was a decrease in the number of days at the hospital for the group who received natriuretic peptides tests (-1.22 days [95% CI: -2.31 to -0.14]) compared to those who did not. The meta-analysis also showed a reduction in the number of days that the intervention group stayed in the critical care unit (-0.56 days [95% CI: -1.06 to -0.05]).

5.2.2 Mortality

The meta-analysis by Lam et al. (2010) found that in-hospital mortality estimates varied across trials but that natriuretic peptide testing to diagnose heart failure in people presenting to the ED with acute dyspnoea does not significantly reduce mortality (OR: 0.96 [95% CI: 0.65 to 1.41; p value: 0.83]).

5.2.3 Hospital admission rate

Admission rates to the hospital from the ED appeared to decrease in the tested group compared with the control group in the meta-analysis, although this finding was not statistically significant (OR: 0.82 [95% CI: 0.67 to 1.01; p value: 0.06]) (Lam et al. 2010).

5.2.4 30-day rehospitalisation rate

The pooled estimate in the meta-analysis showed no significant difference in the 30-day re admission rate in those whose natriuretic peptide levels were tested versus the control group (OR: 0.88 [95% CI: 0.64 to 1.20; p value 0.41]) (Lam et al. 2010).

5.2.5 Length of ED visit

The RCT by Moe et al. (2007) found that the median duration of the initial ED visit reduced by 21%: 5.6 hours in the NT-proBNP group and 6.3 hours in the usual care group (p = 0.0309).

5.2.6 60-day rehospitalisation rate

The RCT by Moe et al. (2007) found a significant reduction in the number of patients re-hospitalised by 60 days (13% versus 20%; p = 0.0463) was observed in those in the NT-proBNP group.

5.2.7 Morbidity (days spent in hospital at 360 days)

The study by Breidthardt et al. (2007) found that number of days spent in-hospital at 360 days was significantly lower in the BNP group (median: 12 days [interquartile range 2 to 28 days]) compared with the control group (median: 16 days [7 to 32 days; p = 0.025]).

5.2.8 Quality of life

We did not identify any studies reporting how patient quality-of-life is affected by NT-proBNP and BNP measurement in the ED.

5.3 Certainty and quality of the evidence

This section highlights any gaps, uncertainties, or issues with the reliability and/or generalisability of the evidence. The majority of evidence comes from existing published meta-analyses (Hill et al. 2014, Lam et al. 2010, Roberts et al. 2015), but some additional observational studies were also reported. These are well-conducted, but the authors highlight some issues with the certainty of the evidence they identified:

- The majority of the primary and secondary evidence we identified did not identify the diagnostic test accuracy of NT-proBNP or BNP using *rule-in* and *rule-out* tests in combination; instead of working out an overall accuracy, studies generally reported on single thresholds used for both *rule-in* and *rule-out*.
- No UK-specific data were available, so findings may not be generalisable to the Welsh NHS. The average age of patients diagnosed with AHF in hospital in the UK is 78 years, which is older than the mean age of some of the studies, which could particularly affect specificity. The majority of studies in the meta-analyses and primary studies reported in this EAR looked at the Elecsys proBNP assay (Roche) and BNP Triage POC assay (Biosite). Expert reviewer opinion suggests that this may be reflective of NT-proBNP assay-use in Wales (BNP assay-use in NHS Wales is unclear). The reference standard in most of the studies was clinical diagnosis based on symptoms and signs. This is comparable to real life according to expert review. Use of echocardiography to diagnose AHF in the primary and secondary studies was variable, which also reflects current practice, according to expert opinion.
- The majority of studies identified are observational (prospective cohort): results from the observational studies are likely to be of lower certainty than those from RCTs.
- Hill et al. (2014) noted that there was considerable heterogeneity across the results of the studies they identified and pooled, which reduces the precision of the pooled analysis they report. This applied to evidence on both natriuretic peptides. In most studies in Roberts et al. (2015) the reference standard was taken as a retrospective synthesis of clinical and imaging data, but was heterogeneous between studies, which may account partly for variability in reported specificity. The majority of the primary studies reported in this EAR and in the meta-analyses combined a subjective assessment of a patient with some combination of data points as a criterion standard for the diagnosis of AHF, which creates heterogeneity across the criterion standard diagnosis of AHF.
- Many of the studies in the review by Roberts et al. (2015) were performed within the ED setting, but they also included studies in other acute settings if the natriuretic peptide was being used to differentiate heart failure from other diagnoses. As diagnostic thresholds vary depending on clinical characteristics, this may limit applicability in some settings.
- Whilst Hill et al. (2014) factored age and renal function into their analysis of diagnostic accuracy, Roberts et al. (2015) did not. Given that background natriuretic peptide levels change with age and renal function, this could affect the generalisability of the results across age groups.
- Expert reviewer opinion states that in existing studies, the prior history of heart failure biases the adjudication of AHF, distorting the natriuretic peptide threshold performance.
- Primary studies that reported clinical outcomes in people who received natriuretic peptide testing versus those who did not used differing thresholds to *rule-in* or *rule-out* heart failure. It is unclear whether all of these are relevant or recommended, which may reduce the generalisability of results from these studies. In addition, it is not reported whether *rule-in* and *or rule-out* natriuretic peptide thresholds were used in the meta-analysis by (Lam et al. 2010).

With regards to the quality of patient selection for Hill et al. (2014), the majority of studies examining either BNP or NT-proBNP had a high risk of bias, or there was not enough information presented in the published manuscripts to make an assessment. Most of the studies had low or unclear risk of bias with respect to the index test, reference standard used, and flow and timing. Both Hill et al. (2014) and Roberts et al. (2015) used QUADAS 2 to assess risk of bias. Lam et al. (2010) assessed the quality of evidence as moderate (using Grading of Recommendations, Assessment, Development and Evaluations [GRADE]).

Table 7. Clinical characteristics of studies in meta-analyses

<table>
<tr><th>Systematic review</th><th>Included studies</th><th>Study design (number of studies)</th><th>Population</th><th>Index test</th><th>Reference test</th><th>Quality (QUADAS-II)</th></tr>
<tr><td colspan="7">Age-specific biomarker thresholds</td></tr>
<tr><td>Hill et al. (2014)</td><td>NT-proBNP
5 studies reported relevant NT-proBNP data

<table><tr><th>Natriuretic peptide threshold</th><th>Number of studies (number of participants)</th></tr><tr><td>Age-independent rule-out threshold (300 ng/L),
and
age-adjusted rule-in threshold of 450 ng/L (< 50 years old), 900 ng/L (50 to 75 years old); 1,800 ng/L (older than 75 years)</td><td>4 (2,017)</td></tr><tr><td>Age-independent rule-out threshold (300 ng/L),
and
age-adjusted rule-in threshold of 450 ng/L (< 50 years old) or 900 ng/L (> 50 years old)</td><td>1 (100)</td></tr></table>
BNP
See Table 5 in EAR for available results of BNP studies. Number of participants for BNP assays were not reported in the meta-analysis

Search period: 1989 to June 2012</td><td>NT-proBNP
Prospective cohort (3)
Prospective RCT (1)
Cross-sectional (1)

BNP
Prospective cohort (2)
Cross-sectional (1)
NR (1)

EDs in France, USA, Norway</td><td>NT-proBNP
Mean age ranged from 61 to 74 years
Prevalence of AHF ranged from 40% to 79%

BNP
Mean age ranged from 64 to 83 years
Prevalence of AHF ranged from 44% to 50%</td><td>NT-proBNP
All used Elecsys proBNP assay (Roche). Analysed using Elecsys 1010 analyser, or Modular Analytics E170 analyser (Roche)

BNP
BNP Triage POC: 3 (Biosite)
iSTAT (Abbott): 1</td><td>All used Clinical adjudication</td><td>With respect to patient selection, a majority of studies examining either BNP or NT-proBNP had a high risk of bias, or there was not enough information presented in the published manuscripts to make an assessment. Most of the studies had low or unclear risk of bias with respect to the reference standard used, and flow and timing.</td></tr>
<tr><td colspan="7">Biomarker thresholds not considering age</td></tr>
</table>

<table>
<tr><th>Systematic review</th><th>Included studies</th><th>Study design (number of studies)</th><th>Population</th><th>Index test</th><th>Reference test</th><th>Quality (QUADAS-II)</th></tr>
<tr><td>Roberts et al. (2015)</td><td>NT-proBNP
<table><tr><th>Natriuretic peptide threshold (ng/L)</th><th>Number of studies (number of participants)</th><th>Number of cases</th></tr><tr><td>≤300</td><td>10 (n = 3,349)</td><td>1,695</td></tr><tr><td>300 to 1,800</td><td>13 (n = 3,223)</td><td>1,652</td></tr><tr><td>≥1,800</td><td>3 (n = 840)</td><td>444</td></tr></table>BNP
<table><tr><th>Natriuretic peptide threshold (ng/L)</th><th>Number of studies (number of participants)</th><th>Number of cases</th></tr><tr><td>≤100</td><td>19 (n = 6,950)</td><td>3,049</td></tr><tr><td>100 to 500</td><td>20 (n = 4,543)</td><td>2,160</td></tr><tr><td>≥500</td><td>4 (n = 283)</td><td>145</td></tr></table>Search period: Up to January 2014</td><td>NT-proBNP
Prospective cohort (14)
Retrospective cohort (1)
Cross-sectional (3)

Setting
ED (15)
Cardiology and pulmonary admissions (1)
Prehospital emergency (2)

BNP
Prospective cohort (17)
Cross sectional (6)
Retrospective cohort (3)

Setting
ED (20)
ICU (2)
Inpatients (2)
Acute admissions (2)</td><td>NT-proBNP
Mean age ranged between 59.1 years and 81 years
Prevalence of HF ranged between 23% and 82%

BNP
Mean age ranged between 56 years and 84.3 years
Prevalence of HF ranged between 31% and 71%</td><td>NT-proBNP
NT-proBNP Elecsys (Roche): 17

NT-proBNP Dimension Dade (Dade-Behring): 1

BNP
BNP Triage (Biosite): 19

BNP Abbott Architecht (Abbot Diagnostics): 3

BNP Access (Beckman Coulter): 2

BNP inhouse assay: 2</td><td>NT-proBNP
Retrospective review: 11
Final hospital diagnosis: 7

BNP
Retrospective review: 22
Final hospital diagnosis: 4</td><td>NT-proBNP
Overall
High ROB: 9
Low ROB: 9

BNP
Overall
High ROB: 16
Low ROB: 10</td></tr>
<tr><td colspan="7">Clinical outcomes</td></tr>
<tr><td>Lam et al. (2010)</td><td>2,513 participants

Search period: January 1996 to July 2010</td><td>5 RCTs in 5 countries (Switzerland, Australia, Canada, the Netherlands, USA)</td><td>Patients presenting to ED with acute dyspnoea
Marginally more men in the studies, and average patient age was older than</td><td>NT-proBNP
1 study used test by Roche and 1 used test by Hoffman-La Roche

BNP test
2 studies used Rapid Triage</td><td>Standard care (routine testing, clinical examination or both [no natriuretic peptide testing])</td><td>Low risk of bias using the Cochrane Collaboration risk of bias tool

GRADE: moderate</td></tr>
</table>

Systematic review	Included studies	Study design (number of studies)	Population	Index test	Reference test	Quality (QUADAS-II)
			70 years for 3 studies but younger in 2 studies/	(Biosite) and 1 study used Abbott AxSYM MEIA automate immunoassay		

BNP: B-type natriuretic peptide ED: emergency department; GRADE: Grading of Recommendations, Assessment, Development and Evaluations; HF: heart failure; ICU: intensive care unit; ng/L: nanograms/litre; NR: not reported; NT-proBNP: N-terminal pro B-type natriuretic peptide; RCT: randomised controlled trial; ROB: risk of bias

Table 8. Study characteristics of primary studies

Study	Outcome(s)	Study type	Number of patients	Patient characteristics	Index test/reference standard
Bombelli et al. (2015)	Diagnostic accuracy	Natriuretic peptide: NT-proBNP diagnostic accuracy Thresholds for those aged 80 years and older: 300 ng/L 980 ng/L 1,470 ng/L 1,800 ng/L 4,200 ng/L 5,340 ng/L Setting: One Italian ED	N = 895 Inclusion criteria: aged 80 years and older Exclusion criteria: NR	Mean age: 86 years old Number of males/females): 368/527	Index test: Serum NTproBNP using NT-proBNP assay (analysed by the Elecsys 2010 analyser [Roche]). Reference test: authors using the data from the hospital database rather than a direct clinical assessment of the subjects.
Darche et al. (2017)	Diagnostic accuracy	Location: ED in Germany Thresholds: 450 ng/L for under 50 years of age, and > 900 ng/L for over 50 years of age to guide *rule-in*, and 300 ng/L to guide *rule-out*	N = 312	Number of males/females): 184/118	Index test: Stratus CS Acute Care NT-proBNP assay (Siemens) Reference test: Clinical adjudication
Moe et al. (2007)	ED/hospital duration, rehospitalisation, mortality	Cohort data from RCT 7 EDs in Canada Supported by Roche diagnostics Thresholds: Based initially on manufacturer-supplied data, and later on the PRIDE study: thresholds of 450 ng/L for under 50 years of age, and > 900 ng/L for over 50 years of age to guide *rule-in*, and 300 ng/L to guide *rule-out*	N = 500 Exclusion criteria: advanced renal failure (serum creatinine >250 µmol/L), acute myocardial infarction, malignant disorders, and dyspnea from clinically overt origins, including pneumothorax and chest wall trauma	People presenting to ED with acute dyspnoea Cohort settings: N = 500 Mean age = 70 years M = 52%	Index test: NTproBNP Roche Reference standard: Retrospective review by cardiologists
Breidthardt et al. (2007)	Clinical outcomes	Thresholds: *rule-out*: 100 ng/L, and *rule-in*: 500 ng/L	N = 452 ED in Switzerland	Age (mean): 71 years Male: 58%	Index test: Biosite Triage to measure BNP

Study	Outcome(s)	Study type	Number of patients	Patient characteristics	Index test/reference standard
					Reference test: Clinical adjudication
Ibrahim et al. (2017): ICON-APAC New Zealand cohort	Diagnostic accuracy	Natriuretic peptide: NTproBNP diagnostic accuracy Threshold: 300 ng/L (all ages) 450 ng/L (< 50 year olds) 900 ng/L (50 to 75 years old) 1800 ng/L (> 75 years old Setting: One ED in New Zealand	N = 500 Inclusion Criteria: Breathlessness as primary complaint in ED Exclusion Criteria: Age under 21, shortness of breath related to trauma, current renal haemodialysis	Mean age: 73 years (63 to 81) Number of males/Females: 288/212	Index test: Serum NTproBNP using Elecsys NT-proBNP immunoassay (analysed using the Elecsys Cobas e411 immunoanalyser) (Roche) Reference standard: Two clinicians (an ED specialist and a cardiologist) independently reviewed all available information on each patient and attributed breathlessness to heart failure, pneumonia, or another cause. Time between index test and reference standard: NR
Ibrahim et al. (2017): ICON-APAC Singapore cohort	Diagnostic accuracy	Natriuretic peptide: NTproBNP diagnostic accuracy Threshold: 300 ng/L (all ages) 450 ng/L (< 50 year olds) 900 ng/L (50 to 75 years old) 1,800 ng/L (> 75 years old Setting: One ED in Singapore	N = 60 Inclusion Criteria: Breathlessness as primary complaint in ED Exclusion Criteria: Age under 21, shortness of breath related to trauma, current renal haemodialysis	Mean age: 56 years (46 to 65 years) Number of males/females (n): 396/210	Index test: Serum NTproBNP using Elecsys NT-proBNP immunoassay (analysed using the Elecsys Cobas e411 immunoanalyser) (Roche) Reference standard: Two clinicians (an ED specialist and a cardiologist) independently reviewed all available information on each patient and attributed breathlessness to heart failure, pneumonia, or another cause. Time between index test and reference standard: NR

Study	Outcome(s)	Study type	Number of patients	Patient characteristics	Index test/reference standard
Januzzi et al. (2018): ICON-RELOADED study	Diagnostic accuracy	Natriuretic peptide: NTproBNP diagnostic accuracy Threshold/s: 300ng/L (all ages) 450 ng/L (< 50 year olds) 900 ng/L (50 to 75 years old) 1,800 ng/L (> 75 years old Study type: pooled cohort Setting: Multiple (19) EDs in USA and Canada	N= 1,461 Inclusion Criteria: Dyspnoeic patients presenting to ED Exclusion Criteria: < 22 years old, Renal insufficiency requiring dialysis or known estimated glomerular filtration rate <15 mL/min/1.73 m2 prior to enrolment	Mean age: 56.4 years (+/- 15.7) Number of male/females: 743/718	Index test: Serum NTproBNP using Elecsys proBNP II (analysed on the Cobase 601 analyser [Roche Diagnostics]). Reference standard: Clinical events adjudication committee, blinded to NT-proBNP results, independently reviewed and adjudicated diagnosis of acute HF. Time between index test and reference standard: NR
Kozhuharov et al. (2019): BASEL V study	Diagnostic accuracy	Natriuretic peptide: NTproBNP diagnostic accuracy Threshold: 300 ng/L (all ages) 450 ng/L (< 50 year olds) 900 ng/L (50 to 75 years old) 1,800 ng/L (> 75 years old Setting: Two EDs in Switzerland	N= 2,053 Inclusion Criteria: Acute dyspnoea as primary complaint in ED Exclusion Criteria: Patients with terminal kidney failure on chronic dialysis	Median age: 78	Index test: Serum NT-proBNP (manufacturer NR) Reference standard: Final hospital diagnosis confirmed by two cardiologists by using medical records and investigation results (e.g. X-ray, echocardiogram) Time between index test and reference standard: NR

BNP: B-type natriuretic peptide ED: emergency department; GRADE: Grading of Recommendations, Assessment, Development and Evaluations; HF: heart failure; ICU: intensive care unit; N: total number of participants; ng/L: nanograms/litre; NR: not reported; NT-proBNP: N-terminal pro B-type natriuretic peptide; RCT: randomised controlled trial; ROB: risk of bias

5.4 Ongoing trials

We identified three ongoing systematic reviews investigating the diagnostic performance and clinical utility of age-stratified thresholds of NT-proBNP to confirm a diagnosis of AHF in an acute care setting:

- Mohamed Subhan Anwar, Kuan Ken Lee, Dr Anoop Shah, Alan Japp, James Januzzi, Nicholas Mills. The diagnostic thresholds of N-terminal pro-B-type natriuretic peptide (NTproBNP) in acute heart failure: a systematic review and collaborative individual patient data meta-analysis. PROSPERO 2019 CRD42019159407. Anticipated publication date: August 2021. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019159407
- Anna Maw. Diagnostic accuracy of lung ultrasound and brain natriuretic peptide in patients with suspected acute decompensated heart failure: a comparative accuracy systematic review and meta-analysis protocol. PROSPERO 2019 CRD42019126783. Anticipated publication date: March 2020. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019126783
- Immaculate Nevis, Yuan Zhang, Aroma Akhund, David Wells, Melissa Walter. Natriuretic peptide testing for the diagnosis of heart failure. PROSPERO 2020 CRD42020148036 Anticipated completion date: 31 July 2020. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020148036

6. Cost effectiveness

6.1 Health economic literature review

To address the question of whether the use of both *rule-in* and *rule-out* thresholds of BNP and NT proBNP as diagnostic tests to *rule-in* and *rule-out* acute heart failure (AHF) is cost effective compared with BNP or NT-proBNP to solely *rule-out* AHF, or clinical examination/standard investigations only, we updated a literature search undertaken for NICE CG187 (NICE 2014). We screened the titles and abstracts of records identified in the update search and 78 potentially relevant health economic studies were identified. An additional four full texts supplied by the manufacturer were considered. 81 studies were excluded. One study was included, and is briefly summarised below and in Table 9 (Siebert et al. 2021).

NICE CG187 included six health economic studies for the review question 'In people with suspected (or under investigation for) acute heart failure, is the addition of natriuretic peptides to the standard initial investigations (using ECG, chest x-ray and blood tests) more accurate compared to standard initial investigations, clinical judgement and each other?' The question for this EAR focuses on *rule-in* and *rule-out* thresholds of BNP and NT-proBNP. Therefore, while NICE CG187 included six health economic studies, only five were included in this EAR. The studies are summarised in Table 9. Two studies compare the use of NT-proBNP *rule-in* and *rule-out* thresholds with conventional diagnostic assessment (Moe et al. 2007, Rutten et al. 2008). Three within trial cost-effectiveness studies, each based on reports at different follow-up periods of the BASEL randomised controlled trial, compare the use of BNP *rule-in* and *rule-out* thresholds in addition to conventional diagnostic assessment with conventional diagnostic assessment alone (AHTA 2007, Breidthardt et al. 2007, Mueller et al. 2006).

Original health economic modelling was also undertaken for NICE CG187. This entailed a cost-effectiveness model on standard or specialist heart failure management, with and without natriuretic peptide testing. The sensitivity and specificity used in the model were taken from diagnostic meta-analyses of sensitivity and specificity data using thresholds ≤100pg/ml for BNP

and 300-1800pg/ml for NT-proBNP. However, none of the strategies compared in the model involved the simultaneous application of *rule-in* and *rule-out* thresholds of natriuretic peptides, and so the model undertaken for CG187 was excluded from this EAR.

All six health economic studies included in this Evidence Appraisal Report were assessed as partially applicable with potentially serious limitations. Three studies compared an NT-proBNP strategy with conventional diagnostic assessment and three studies compared a BNP strategy with conventional diagnostic assessment. These are briefly summarised here and limitations are discussed in Table 9. One US cost-effectiveness analysis and one Dutch cost-effectiveness analysis found that an NT-proBNP-based strategy with *rule-in* and *rule-out* thresholds was cheaper with fewer serious adverse events or fewer deaths, respectively, compared with clinical assessment alone (Rutten et al. 2008, Siebert et al. 2021). In addition, one Canadian within-trial cost consequence analysis reported that the NT-proBNP strategy was cheaper than conventional diagnostic assessment, while clinical outcomes were mixed (Moe et al. 2007). Studies in Australia and Switzerland which compared a BNP strategy with *rule-in* and *rule-out* thresholds with conventional diagnostic assessment also found the BNP strategy to be dominant (cheaper with lower mortality (AHTA 2007, Mueller et al. 2019). However, a subsequent analysis of data from a later follow-up of the same trial as Mueller et al. (2019), concluded that BNP was less costly than diagnostic assessment alone but found no difference in long-term mortality (Breidthardt et al. 2007). No studies compared a BNP or NT-proBNP strategy using both *rule-in* and *rule-out* with a strategy using *rule-out* thresholds only.

Table 9. Summary of included health economic studies: Siebert et al. (2021), Moe et al. (2007), Rutten et al. (2008), AHTA (2007), Mueller et al. (2006) and Breidthardt et al. (2007)

Study details	Study population and design	Data sources	Results	Quality assessment
NT-proBNP				
Author and year: Siebert et al. (2021) **Country:** USA **Type of economic analysis:** Cost-effectiveness analysis (health outcome: serious adverse events avoided) **Perspective:** US Medicare perspective **Currency:** US dollars **Price year:** 2019 US dollars **Time horizon:** 6 months **Discounting:** NA **Potential conflict of interest:** Industry funded (Roche Diagnostics)	**Population** Dyspnoeic patients presenting to an emergency department Cohort settings: NR **Intervention** Clinical assessment supported by NT-proBNP NT-proBNP test result, which could fall into 1 of 3 categories: positive, negative, or inconclusive. Positive and negative diagnoses were respectively determined by *rule-in* (>450, >900, and >1,800 pg/ml for patients aged <50, 50–75, and >75) and *rule-out* thresholds (<300 pg/ml). In the grey zone, diagnosis was based on clinical assessment. **Comparator** Clinical assessment alone **Study design** Decision tree model, incorporating echocardiograms, intensive care admissions, cardiology ward admissions, hospitalisations after the index ED admission, late or repeated hospitalisation, readmission to the ED and	**Source of baseline and effectiveness data:** AHF prevalence in dyspnoeic patients presenting to ED, as well as sensitivity and specificity of NT-proBNP with *rule-in* and *rule-out* thresholds, were obtained from ICONRELOADED. Diagnostic accuracy of clinical assessment alone was not available from this study. Therefore, a statistical prediction model was fitted to individual patient data from the study, using logistic regression with variables related to patients' demographics, medical history, and clinical symptoms to estimate the accuracy of clinical assessment alone. Probabilities of clinical events conditional on diagnostic category were based on US participant data from ICON-RELOADED and in some cases the PRIDE study (see differences in management assumptions between strategies in 'limitations'). **Source of resource use and cost data:**	**Base case results** **Costs** NT-proBNP: $20,247 Clinical assessment alone: $22,584 NT-proBNP saves $2,337 **Serious adverse events per person** NT-proBNP: 1.31 Clinical assessment alone: 1.39 NT-proBNP reduced serious adverse events by 5.8% **ICER (health outcome: serious adverse events)** Deterministic: NT-proBNP dominates clinical assessment alone **Sensitivity analysis** Probabilistic sensitivity analysis: NT-proBNP dominates in 83% of simulations Deterministic sensitivity analyses: The greatest changes in clinical benefit resulted from changes in specificity of diagnostic strategies, probabilities of readmissions related to negative diagnoses, and number of readmissions with true negative diagnoses. Parameters with the greatest impact on costs were: specificity of the interventions, hospitalisation (probabilities of readmission after false diagnosis and	**Applicability** The study is partially applicable as it considers the US Medicare perspective **Limitations** This study has potentially serious limitations. • The study does not include all relevant comparators (BNP). • Grey zone diagnostic accuracy assumed equal to overall diagnostic accuracy of clinical assessment alone • Differences in management assumptions and resource use between strategies: o 100% of false positives in clinical assessment alone arm are hospitalised after emergency department admission, compared with 80% in NT-proBNP arm. o 41% of true negatives are hospitalised after emergency department admission in NT-proBNP arm, compared with 55% in clinical assessment alone arm o 14% of true negatives experience late or repeated hospitalisation

Study details	Study population and design	Data sources	Results	Quality assessment
	urgent follow-up visits.	Resource utilization conditional on diagnostic category were based on US participant data from ICON-RELOADED. Resource items: ED stay, hospitalization, and urgent follow-up Cost of ED care based on the Nationwide Emergency Department Sample database. Cost of hospitalization from the Hospital Inpatient National Statistics database. Urgent care visits cost from the Physicians' Fee & Coding Guide.	proportion of rehospitalisations taking place directly from ED after true-negative diagnoses with clinical assessment). No tornado diagram bars cross the line to negative serious adverse event results or to positive costs for the NT-proBNP-supported strategy, indicating the robustness of results.	in NT-proBNP arm compared with 19% in clinical assessment alone arm o 8% of false positives receive echocardiogram during emergency department visit in NT-proBNP arm while only 6% receive echocardiogram in clinical assessment alone arm. • No long-term consequences of false negative and false positive results were considered • NT-proBNP cost not included
Author and year: Moe et al. (2007) **Country:** Canada **Type of economic analysis:** Cost consequence analysis **Perspective:** Canadian Health System (Payer) **Currency:** US dollars **Price year:**	**Population** People presenting to ED with acute dyspnoea Cohort settings: N = 500 Mean age = 70 years M = 52% **Intervention** Conventional diagnostic assessment supplemented by NT-proBNP result. Information to interpret NT-proBNP results based initially on manufacturer-supplied data, and later on the PRIDE study (Januzzi et al. 2005). PRIDE used thresholds of 450pg/ml for under 50 years of age, and	**Source of baseline and effectiveness data:** Within-trial (IMPROVE-CHF study). **Source of resource use and cost data:** Resource items include cost of NT-proBNP test at initial emergency department visit, initial and subsequent emergency department visits, hospitalisations, physician fees and outpatient services. Resource use data was obtained from hospital charts and outpatient service resource use was estimated based on 60-day	**Base case results** **Costs** Median per patient cost of care at 60-days Conventional diagnostic assessment (n=254): $6129 NT-proBNP (n=246): $5180 NT-proBNP saves $949 (p=0.0232) **Health outcomes (NT-proBNP vs. conventional diagnostic assessment)** 60-day all-cause mortality RR: 1.22 (p=0.58) In-hospital mortality RR: 1.89 (p=0.19) Hospitalisation (initial) RR: 0.98 (p=0.83) 60-day re-hospitalisation RR: 0.66 (p=0.046)	**Applicability** The study is partially applicable as it considers the Canadian health system, and has potentially serious limitations. **Limitations** NT-proBNP thresholds not adjusted for gender, renal function or obesity. Short follow-up unlikely to reflect all differences in costs and outcomes A specific patient management strategy was not stated. Unclear whether thresholds used to interpret NT-proBNP results

Study details	Study population and design	Data sources	Results	Quality assessment
2005 US dollars **Follow-up:** 60 days **Discounting:** NA **Potential conflict of interest:** Industry funded (Roche Diagnostics)	>900pg/ml for over 50 years of age to guide rule-in, and 300pg/ml to guide *rule-out* **Comparator** Conventional diagnostic assessment **Study design** Within-trial cost consequence analysis of Improved Management of Patients With Congestive Heart Failure (IMPROVE-CHF) RCT. Analysis of averaged individual level resource use with local unit costs applied	telephone interviews with patients and chart records. Hospital costs were based on budget records of hospitals using a national database for financial and statistical information. Physician fees and outpatient services were estimated using average reimbursement fees from Ontario and Quebec. The total cost (including operating and capital costs) of NT-proBNP in the emergency department was estimated as $37.	Median days hospitalised (initial): NT-proBNP one fewer (p=0.302) Median duration of ED visit (hours): NT-proBNP 0.7 fewer (p=0.031) Median days hospitalised in ICU: NT-proBNP 0.5 higher (p=0.723) **ICER** N/A **Sensitivity analysis** None **Subgroup analysis** A subgroup analysis was conducted including only people with an intermediate likelihood (20%-80%) of heart failure based on emergency department physician assessment (n=105 and n=114 for NT-proBNP and usual care respectively). For this subgroup, NT-proBNP saves $1,496 (p=0.1264). Duration of ED visit is significantly lower in the NT-proBNP group for this subgroup of patients.	changed during the study, as it states that interpretation was initially guided by manufacturer-supplied data and later on PRIDE (Januzzi et al. 2005). The study states that hospitalised patients received a further NT-proBNP test 72 hours after admission to guide subsequent management. Resource items specify that the initial NT-proBNP test is included, but follow up tests are not listed. Outpatient service resource use was based on patient telephone interviews.
Author and year: Rutten et al. (2008) **Country:** Netherlands **Type of economic analysis:** Cost-effectiveness analysis (health outcome: 30-day mortality)	**Population** People presenting to emergency department with acute dyspnoea Cohort settings: N = 477 Mean age = 59 years M = 54% **Interventions** Conventional diagnostic assessment supplemented by NT-proBNP result as follows: Rule out: 93pg/ml for males and 144pg/ml for females	**Source of baseline and effectiveness data:** Within-trial analysis of a single-centre RCT, included within clinical section of this review as part of Lam et al meta-analysis. Thirty day outcome assessed by review of electronic hospital records. If no records present, patients were contacted. If unsuccessful, GPs were contacted.	**Costs** NT-proBNP: $4,984 Conventional diagnostic assessment: $6,352 Incremental: $1,364 (95% CI Saves $246 to $3,215 more) **Health outcomes (NT-proBNP vs. conventional diagnostic assessment)** Median days to discharge: -2.0 (p=0.04) Initial hospitalisation RR: 0.00 (p=0.26)	**Applicability** The study is partially applicable as it considers the Dutch health system, and has potentially serious limitations. **Limitations** NT-proBNP thresholds not adjusted for age, renal function or obesity. Short follow-up unlikely to reflect all differences in costs and outcomes.

Study details	Study population and design	Data sources	Results	Quality assessment
Perspective: Dutch (Payer) **Currency:** US dollars **Price year:** 2005 US dollars **Follow-up:** 30 days (mortality, readmission) **Costs-** NR **Discounting:** NA **Potential conflict of interest**: None **Source of funding:** Erasmus Medical College Medical Research Advisory Committee	Rule in: 1,017pg/ml **Comparator** Conventional diagnostic assessment **Study design** Within-trial cost effectiveness of a single-centre RCT. Analysis of averaged individual level resource use with local unit costs applied	**Source of resource use and cost data:** Resource items include hospital days on general ward and/or intensive or coronary care unit, cardiopulmonary investigations (chest x-ray, electrocardiography echocardiography, chest computed tomography, coronary angiography, right-sided heart catheterisation, pulmonary function tests, bronchoscopy, myocardial perfusion, scintigraphy, bicycle ergometry), and the total cost of NT-proBNP measurement ($34) Hospital admission costs based on national university hospital prices. Diagnostic investigation costs were based on the charge to health insurance companies.	Median duration of initial hospitalisation: -0.3 (p=0.48) Intensive care admission RR: 1.02 (p=0.92) In-hospital mortality RR: 0.95 (p=0.89) 30-day all-cause mortality RR: 0.86 (p=0.26) 30-day readmission RR: 0.60 (p=0.18) **ICER (health outcome: 30-day mortality)** NT-proBNP dominates (fewer deaths and less costly than) conventional diagnostic assessment. This was based on cost-effectiveness plane figure showing results of bootstrap sampling; ICER value not otherwise reported or calculable. Probability that NT-proBNP is cost effective not reported, however 95% confidence intervals (indicated on figure) cross into the 'higher mortality' south-west quadrant **Sensitivity analysis** None **Subgroup analysis** A post hoc sub-group analysis indicated that the effect on costs is largest in patients with cardiac dyspnoea compared with noncardiac dyspnoea (mean saving of $2,627 compared with non-cardiac dyspnoea patients $150)	A specific patient management strategy was not stated Twenty people were transferred to another hospital. Only time to discharge and costs incurred at Erasmus MC were included in analysis. Follow up was not specified for costs but it appears that only the costs of initial hospitalisation have been included. Re-hospitalisations costs do not appear to have been considered (though there was no significant difference between groups for this outcome). Costs (follow-up NR) were plotted against 30 day mortality for the cost effectiveness plane. 30 day outcome was assessed by review of electronic health records or be contacting patient by telephone if no records were found. The study acknowledges that generalisability may be limited as 62% of people in study were first reviewed by a GP.

Study details	Study population and design	Data sources	Results	Quality assessment
BNP				
Author and year: AHTA (2007) **Country:** Australia **Type of economic analysis:** Cost-effectiveness analysis (health outcome = 30-day mortality) **Perspective:** Australian payer **Currency:** Australian dollars Price year: 2005 **Time horizon:** 30 days **Discounting:** NA **Potential conflict of interest:** None	**Population** Patients presenting to ED with acute dyspnoea Cohort settings: N = 452 Mean age = 71 years M = 58% **Intervention** Conventional assessment supplemented and guided by B-type natriuretic peptide testing as follows: *Rule-out:* 100pg/ml *Rule-in*: 500pg/ml **Comparator** Conventional diagnostic assessment **Study design** A within-trial cost effectiveness analysis of the BASEL RCT. Analysis of averaged individual level resource use, with Australian unit costs applied	**Source of baseline and effectiveness data:** Mortality rate was acquired from the 30-day published findings of the BASEL trial **Source of resource use and cost data:** Resource use data from BASEL trial. Cost items: Emergency care and admitted patient care including cardiopulmonary investigations, outpatient care, and BNP test Australian Refined Diagnosis Related Group cost estimates were used for hospital charges for heart failure and alternative diagnoses (Department of Health and Aging 2006). The unit cost of BNP testing was obtained through local laboratory benchmarking data (AUS$50.59 for a batch run of 10)	**Base case results** **Costs** **Total cost per patient, mean:** BNP: AUS$3,756 Conventional: AUS$4,094 Difference: BNP saves AUS$338 **Health outcomes** **30-d all-cause mortality:** BNP: 0.10 Conventional: 0.12 Difference: BNP 0.026 lower(95% CI -0.083 to -0.032) **Time to discharge, median (IQR):** BNP: 8 (1–16) Conventional: 11 (5–18) Difference: BNP 3 lower *P* = 0.001 **Initial hospitalization rate:** BNP: 0.75 Conventional: 0.85 Difference: BNP 0.10 lower *P* = 0.008 **30-d readmission rate:** BNP: 0.12 Conventional: 0.10 Difference: BNP 0.02 higher *P = 0.63* **ICER (health outcome 30-day mortality):** BNP in addition to conventional diagnostic assessment dominates (less costly and lower	**Applicability** The study is partially applicable as it considers the Australian payer perspective. **Limitations** The study has potentially serious limitations: BNP thresholds not adjusted for gender, age, renal function or obesity Short follow-up unlikely to reflect all differences in costs and outcomes

Study details	Study population and design	Data sources	Results	Quality assessment
			mortality) conventional diagnostic assessment alone **Analysis of uncertainty:** PSA using bootstrap sampling BNP cheaper and mortality lower: 78.8% BNP cheaper but mortality higher: 18.8% BNP more expensive and mortality lower: 1.9% BNP more expensive and mortality higher: 0.5% At 30 d, the primary cost saving element is the patient admission rate (initial plus re-admission) In a sensitivity analysis where the episode cost admission was increased by $231 (the cost of an echocardiogram) in the BNP arm, BNP saved $139 per patient. Threshold analysis: Cost of BNP test would need to be $389 for no incremental cost savings.	
Author and year: Mueller et al. (2006) **Country:** Switzerland **Type of economic analysis:** Cost-effectiveness analysis (health outcome: all-cause mortality)	**Population** Cohort settings: N = 452 Mean age = 71 years M = 58% **Intervention** Conventional assessment supplemented and guided by B-type natriuretic peptide testing as follows: *Rule-out*: 100pg/ml *Rule-in*: 500pg/ml	**Source of baseline and effectiveness data:** 180 day follow-up of the BASEL trial **Source of resource use and cost data:** Resource use from BASEL RCT. Total cost of treatment included all hospitalizations after the initial presentation to the ED. Emergency and admitted	**Base case results** **Costs** **Total cost per patient, mean (SD):** BNP: $7,930 (8,805) Conventional: $10,053 (10,176) Difference: BNP saves $2,123 *P* = 0.004 **Health outcomes** 180-d all-cause mortality: BNP: 20% Conventional: 23%	**Applicability** The study is partially applicable as it considers the Swiss payer perspective **Limitations** This study has potentially serious limitations: BNP thresholds not adjusted for gender, age, renal function or obesity

Study details	Study population and design	Data sources	Results	Quality assessment
Perspective: Swiss payer **Currency:** US dollars Price year: 2003 Time horizon: 180 days **Discounting:** NA **Potential conflict of interest:** None **Source of funding:** Swiss National Science Foundation, Swiss Heart Foundation, Novartis Foundation, Krokus Foundation, University of Basel	**Comparator** Conventional diagnostic assessment **Study design** A within trial cost effectiveness analysis of a single-blinded RCT trial. Analysis of mortality and averaged individual level resource use with local (Swiss) unit costs applied	patient care (except medication for non-cardiac and non-pulmonary conditions) including cardiopulmonary investigations, outpatient care, intensive care and BNP test Costs of cardiovascular and pulmonary medication calculated according to standard rates in Switzerland in 2003. For the cost of BNP testing ($47) charges were standardised according to the actual rates for patients with general insurance who were living in Basel, Switzerland. Expenses for hospital care were primarily determined by the intensity of care and the length of stay.	Total days in hospital, median (IQR): BNP: 10 (2–24) Conventional: 14 (6–27) *P* = 0.005 Days in hospital for dyspnea, median (IQR): BNP: 9 (1–20) Conventional: 13 (6–24) *P* = 0.003 Initial hospitalisation rate: BNP: 0.75 Conventional: 0.85 Difference: -0.10 *P* = 0.008 Admission rate to ICU: BNP: 0.15 Conventional: 0.24 Difference: -0.09 *P = 0.01* **ICER (health outcome all-cause mortality at 180 days):** BNP dominates conventional diagnostic assessment alone (less costly, lower all-cause mortality) **Analysis of uncertainty:** PSA using bootstrap sampling found BNP to be dominant in 80.6% of the simulations, less costly/higher mortality in 19.3% of the simulations, more costly/lower mortality in 0.04% of the simulations, and more costly/higher mortality in 0.02% of the simulations Sensitivity analyses were performed for changes in the duration of the initial hospitalization, cost of BNP	Not all relevant costs were included (study did not include costs of non-cardiac and non-pulmonary medications)

Study details	Study population and design	Data sources	Results	Quality assessment
			testing, time in intensive care, cost of long-term medication, and re-hospitalisation days. Results were robust, with exception of changes in rehospitalisation days. Subgroup analysis showed that the 180-day cost-reduction benefit of BNP testing was enhanced in patients with a history of coronary artery disease (p=0.005) and those with pulmonary disease (p=0.01)	
Author and year: Breidthardt et al. (2007) **Country:** Switzerland **Type of economic analysis:** CEA (health outcome = all-cause mortality) **Perspective:** Swiss (Payer) **Currency:** US dollars **Price year:** 2003 **Time horizon:** Health outcomes: 720 days; Costs: 360 days **Discounting:** None	**Population** Patients presenting to ED with acute dyspnoea Cohort settings: N = 452 Mean age = 71 years M = 58% **Intervention** Conventional assessment supplemented and guided by B-type natriuretic peptide testing (*Rule-in*: 500pg/ml; *rule-out* 100pg/ml) **Comparator** Conventional diagnostic assessment **Study design** A within trial cost effectiveness analysis of a single-blinded RCT trial. Analysis of mortality and averaged individual level resource use with local (Swiss)	**Source of baseline and effectiveness data:** 720 day published findings of the BASEL trial. Study included in clinical review for this evidence appraisal report. **Source of resource use and cost data:** Resource use (360 days) from BASEL RCT. Total cost of treatment included all hospitalisations after the initial presentation to the ED. Emergency and admitted patient care (except medication for non-cardiac and non-pulmonary conditions) including cardiopulmonary investigations, outpatient care, intensive care and BNP test. Costs of cardiovascular and pulmonary medication calculated according to standard rates in Switzerland in	**Base case results** **Costs** **Total cost per patient, mean** BNP: $10,144 Conventional: $12,748 Difference: BNP saves $2,604 *P* = 0.008 **Health outcomes** 720-d all-cause mortality: BNP: 37% Conventional: 36% Total days in-hospital, median (IQR): BNP: 12 (2–28) Conventional: 16 (7–32) *P* = 0.025 Days in-hospital for dyspnoea, median (IQR): BNP: 11 (2–23) Conventional: 14 (6–26) *P* = 0.009 **ICER (health outcome all-cause mortality):** NR- BNP is less costly than conventional diagnostic assessment alone; no difference in long-term	**Applicability** The study is partially applicable as it considers the Swiss payer perspective **Limitations** Potentially serious limitations: Natriuretic peptide thresholds not adjusted for gender, age, renal function or obesity Not all relevant costs were included Unable to calculate ICER (not reported in study) due to different follow-up time for costs and outcomes. PSA diagram is based on 360 days follow-up, though 360 day mortality data used not reported elsewhere in publication. Unclear whether 720 day data were used for

Study details	Study population and design	Data sources	Results	Quality assessment
Potential conflict of interest: None **Source of funding:** Swiss National Science Foundation, Swiss Heart Foundation, Novartis Foundation, Krokus Foundation, University of Basel.	unit costs applied **NB This is a later analysis of the same RCT and cohort as Mueller 2006 but using costs and outcomes from a longer follow-up**	2003. For the cost of BNP testing ($47) charges were standardised according to the actual rates for patients with general insurance who were living in Basel, Switzerland. Expenses for hospital care were primarily determined by the intensity of care and the length of stay.	mortality, and fewer days spent in-hospital. **Analysis of uncertainty:** PSA using bootstrap sampling found BNP to be dominant in 39.5% of the simulations, less costly/higher mortality in 59.1% of the simulations, more costly/lower mortality in 0.5% of the simulations, and more costly/higher mortality in 0.9% of the simulations. The reduction in initial mortality observed in frail elderly patients was no longer evident at 720 days. The reduction in days hospitalized was the major driver for a significant reduction in total treatment cost at 360-days	mortality and 360 day data used for costs.

Abbreviations
BNP: B-type natriuretic peptide; ED: emergency department; EQ-5D: EuroQol five-dimensions questionnaire; GP: general practitioner; ICER: incremental cost-effectiveness ratio; ICU: intensive care unit; IMPROVE-CHF (RCT): Improved Management of Patients With Congestive Heart failure; NA: not applicable; NR: not reported; NT-proBNP: N-terminal pro-B-type natriuretic peptide; PRIDE (RCT): N-terminal Pro-BNP Investigation of Dyspnea in the Emergency Department study; QALY: quality-adjusted life year; RCT: randomised controlled trial; RR: risk ratio

6.2 Manufacturer-submitted cost-utility analysis

No studies were identified in the health economic literature review in section 6.1 which compared a strategy using BNP or NT-proBNP with both *rule-in* and *rule-out* with a strategy using *rule-out* thresholds only. This is addressed in an unpublished cost-utility analysis which was submitted by the manufacturer to HTW for the purpose of this appraisal. This cost-utility analysis from the UK NHS perspective is described in Table 10.

The model aimed to update the health economic model undertaken for CG187, which was excluded from this EAR as it did not include any strategies involving the simultaneous application of *rule-in* and *rule-out* thresholds of natriuretic peptides. In the updated model, the BNP strategy was excluded. It instead compared three strategies:

1. Clinical decision alone
2. NT-proBNP using a single *rule-out* threshold
3. NT-proBNP using *rule-in*/*rule-out* thresholds

The manufacturer cited several reasons for excluding BNP from the model, including that NT-proBNP is used rather than BNP in current practice in Wales and that the NICE CHF guideline excluded BNP (due to its higher sensitivity in the CHF population and because of BNP's known interaction with sacubitril/valsartan, which limits the utility of BNP as a test to monitor treatment effectiveness).

Table 10. Summary of manufacturer-submitted cost-utility analysis

Study details	Study population and design	Data sources	Results	Quality assessment
Year: Unpublished (2021) **Country:** UK Type of economic analysis: Cost-utility analysis **Perspective:** NHS and PSS **Currency:** UK pounds **Price year:** 2020 Time horizon: Lifetime Discounting (costs/QALYs): 3.5%/3.5% **Potential conflict of interest:** Model developed by manufacturer of Elecsys NT-proBNP device	**Population** People presenting to emergency departments with suspected acute heart failure Cohort settings: Median age = 80 years M = NR **Interventions** NT-proBNP RI/RO: Clinical assessment supported by NT-proBNP NT-proBNP test result, which could fall into 1 of 3 categories: positive, negative, or inconclusive. Positive and negative diagnoses were respectively determined by rule-in (>450, >900, and >1,800 pg/ml for patients aged <50, 50–75, and >75) and *rule-out* thresholds (<300 pg/ml). In the grey zone, diagnosis was based on clinical assessment. **Comparators** NT-proBNP RO only Clinical assessment alone (CD) **Study design** Decision tree covering the initial diagnosis and the short term events that occur within the hospital stay, Two decision tree structures presented:	**Source of baseline and effectiveness data:** AHF prevalence in dyspnoeic patients presenting to ED, as well as sensitivity and specificity of NT-proBNP with rule-in and *rule-out* thresholds, were obtained from ICONRELOADED. The diagnostic test accuracy of clinical decision alone was taken from the IMPROVE CHF study. The diagnostic test accuracy of the NT-proBNP '*rule-out*' strategy was based on an updated meta-analysis, initially undertaken for NICE CG187. Diagnostic test accuracy of NT-proBNP strategy with differential rule-in and *rule-out* cut-offs used the estimates of sensitivity and specificity from the '*rule-out*' meta-analysis to quantify the proportion of patients with and without AHF falling below the 300pg/ml cut-off. Above the 300pg/ml threshold, the BASEL V study was used. Diagnostic test accuracy for people in the NT-proBNP *rule-in/rule-out* 'grey zone' was assumed to be the same as clinical assessment alone. This was validated using a logistic regression model for predicting AHF in the individual patient data from the ICON-RELOADED cohort. Inpatient mortality was obtained from the NICOR dataset.	**Results (NICE structure, DA)** **Costs** CD: £5,543 NT-proBNP RI/RO: £5,583 NT-proBNP RO only: £5,694 **QALYs** CD: 124.6 NT-proBNP RI/RO: 126.4 NT-proBNP RO only: 127.0 **ICER (cost per QALY)** NT-proBNP RI/RO vs CD: £2,203 per QALY NT-proBNP RO only vs NT-proBNP RI/RO: £34,611 per QALY **Alternative model structure (with node for probability of admission)** **Costs** CD: £4,720 NT-proBNP RI/RO: £4,747 NT-proBNP RO only: £5,086 **QALYs** CD: 123.6 NT-proBNP RI/RO: 126.2 NT-proBNP RO only: 126.6 **ICER (cost per QALY)** NT-proBNP RI/RO vs CD: £1,068 per QALY NT-proBNP RO only vs NT-proBNP RI/RO: £74,354 per QALY **Sensitivity and scenario analyses** Scenario analyses:	**Applicability** The study is directly applicable as it considers the UK NHS and PSS perspective. **Limitations** Some potential limitations were identified: • Echocardiogram is assumed to be 100% accurate. • False positive and true negative patients do not enter the long-term component of the model. This may underestimate costs and health consequences of false positive results. • There is no direct evidence available on how patients are diagnosed if they fall within the grey zone between rule-in and *rule-out* thresholds. Therefore this aspect of the

Study details	Study population and design	Data sources	Results	Quality assessment
	• Structure based on NICE CG187 model, with no decision node accounting for the probability of admission following a diagnostic test accuracy result. assigned an average length of stay to each type of patient • Chance node accounting for the probability of admission. Longer term component of the model uses a semi-Markov structure, based on a parametric survival curve. It comprises three states: 'alive no admission', 'alive admission' and dead. The occupancy of the dead state is dictated by the proportion above a parametric survival curve in any given cycle and, among those that are alive, the proportion of people who are admitted and not admitted in any given cycle is dictated by cycle-specific probabilities. In the long-term model a proportion of TPs start on each follow-on service (Aldosterone Agonists, ACEi/ARB, Beta-blocker, Sacubitril Valsartan).	A Kaplan-Meier curve showing up to 17-year survival for patients with AHF first diagnosed in hospital in the UK in the period 2000-2017 from a published study was used for long-term survival in true positives, and the curve for FNs was calculated by applying weighted hazard ratios to the TP curve, to account for these patients not receiving treatment (ACEi/ARB, Beta-blockers, cardiologist follow-up). Probabilities for readmission for TPs in the long-term model were taken from the NICE CG187 economic model. For FNs, admissions were adjusted by a relative risk as in the NICE CH187 model. **Source of resource use and cost data:** Resource items: NT-proBNP tests, echocardiograms, length of stay, cardiology follow-up (outpatient visits, NT-proBNP tests, blood tests), heart failure nurse follow-up (community visit, GP visit), drugs Data sources: • TPs: total average spell costs for a new diagnosis of heart failure from an analysis of HES data • TNs: assigned the average non-elective inpatient spell cost from the NHS reference costs • FN or FP led to a 2 day length of stay penalty. FNs: NHS reference costs 2017/2018 excess bed day for heart failure unit costs were used, inflated to 2020 values.	• Alternative sources of diagnostic accuracy data for clinical decision • Alternative diagnostic accuracy data for *rule-out* cutoff • Alternative proportions of true positives with NT-proBNP higher than the rule-in threshold • Alternative proportions of false positives with NT-proBNP higher than the rule-in threshold • Grey zone assumptions • Admission probability assumptions • Non-admitted mortality assumptions • Cardiology follow-up for FPs • Alternative AHF prevalence data • FN status correction assumptions • QALY penalties for long term FPs (maximum -0.2) • Alternative survival curve assumptions • Alternative bed day cost sources • Assumptions on NT-proBNP in CD arm • Long term NT-proBNP tests annually • True positives using Sacubitril/Valsartan For the NICE structure, there were two scenarios where RO only replaced RI/RO was as the top ranked strategy. When the Se/SP in the grey zone was reduced in the NICE structure model to	analysis is based upon assumptions. • No consideration of LVSD and non-LVSD. • In the base case, studies using non-Roche NT-proBNP tests were not included in the meta-analysis for sensitivity and specificity of the *rule-out* threshold. • 17.9% of hospitalised TP and FP patients in the clinical decision alone arm were assumed to receive an NT-proBNP test at some point during their stay, but in the NT-proBNP arms no further tests were included beyond the initial emergency department test. • The model excluded the costs and benefits of Sacubitril/Valsartan from the model in the base case (explored in sensitivity analysis). • A meta-analysis was conducted for the diagnostic accuracy

Study details	Study population and design	Data sources	Results	Quality assessment
	A proportion of patients are re-admitted each cycle.	Non-AHF bed days from NHS reference costs (all HRGs) applied to FPs. • Proportion of patients receiving cardiology follow-up and heart failure nurse follow-up was taken from the national heart failure audit. Resource use based on NICE CG187. Unit costs from NHS Reference Costs and PSSRU. • Unit costs for drugs from BNF. Proportions of drugs prescribed based on English prescription cost analysis 2019 (NHS Business Services Authority). • The cost of the NT-proBNP test, was based on the list price from Roche Diagnostics Ltd (£24.53) **Source of quality of life data:** Utility values for cycles where patients were and were not admitted to hospital were taken directly from the NICE model.	70%/70%, RO became cost effective in several scenarios. For the alternative model structure with the node for admission probability, the model's conclusions are very insensitive to changing input parameters, with NHB always being highest in the RI/RO strategy and RO almost always having a higher NHB than CD. <u>One way sensitivity analyses</u> For the NICE structure model show that the results are most sensitive to the bed day increment for FPs. For a low bed day increment (1 day) RO becomes cost effective. Using the alternative structure, there are no parameters which influence the conclusions when changed. <u>Probabilistic sensitivity analysis</u> (NICE structure) NT-proBNP rule in/rule out was cost effective compared to the rule out strategy in 92% of simulations, and cost effective compared to clinical decision in all simulations.	of the rule-in/*rule-out* strategy but the resulting values were not used in base case. • The population of the study from which the long-term mortality data for TPs were derived (survival from diagnosis) does not match the exact population required (survival from discharge), but is thought to be a good approximation. • Computational errors identified in some of the model calculations

Abbreviations
AHF: acute heart failure; BNF: British National Formulary; CD: clinical decision alone strategy; DA: deterministic analysis; FN: false negative; FP: false positive; HES: Hospital Episode Statistics; HRG: Healthcare Resource Group; ICER: incremental cost-effectiveness ratio; NT-proBNP: N-terminal pro-B-type natriuretic peptide; NICE: National Institute of Health and Care Excellence; NHB: Net health benefit; PSSRU: Personal Social Services Research Unit; QoL: quality of life; QALY: quality-adjusted life year; RCT: randomised controlled trial; RI/RO: rule-in/rule/out combined strategy; RO: *Rule-out* only; TP; true positive; TN: true negative

6.3 HTW cost utility analysis

To address the remaining uncertainty and ensure that all relevant comparators are considered, HTW researchers constructed a cost-utility analysis based on both the model built for NICE CG187 (NICE 2014) and the manufacturer-submitted model summarised in section 6.2.

The model comprises a decision tree and a Markov model to evaluate the total costs and QALYs of the following included strategies:

1. Clinical decision alone
2. NT-proBNP using a single *rule-out* threshold
3. NT-proBNP using *rule-in/rule-out* thresholds
4. BNP using a *rule-out* threshold only

The analysis took the perspective of the UK NHS and personal social services (PSS). A lifetime time horizon was considered, and future costs and benefits were discounted at rate of 3.5%. Full details of the methods and results are available in Appendix 5.

The results of the base case analysis are presented in Table 11. All three of the BNP and NT-proBNP testing strategies were found to be more costly and more effective than clinical decision alone. The resulting ICERs were below a threshold of £20,000 per QALY indicating that the BNP and NT-proBNP strategies were cost-effective in comparison to clinical decision alone. When comparing all strategies against each other, the NT-proBNP *rule-in/rule-out* strategy was found to be the optimal strategy.

The probabilistic sensitivity analysis showed that the BNP *rule-out* strategy had a 58% probability of being cost effective while the NT-proBNP *rule-in* and *rule-out* strategy had a 42% probability of being cost-effective, clinical decision alone had a 0% probability of being cost-effective and NT-proBNP *rule-out* had a 0% probability of being cost-effective, at a threshold of £20,000 per QALY.

In deterministic sensitivity analyses, the results were largely robust with the NT-proBNP *rule-in* and *rule-out* strategy found to be optimal in most modelled scenarios. The one exception was a scenario in which it was assumed that false negative results are not corrected during the hospital stay.

Table 11. Base case results

Diagnostic strategy	Cost		QALYs		ICER (cost per QALY)
	Total	Incremental	Total	Incremental	
Clinical decision	£5,664	-	4.740	-	-
BNP *rule-out*	£5,713	£49	4.757	0.017	£2,882
NT-proBNP *rule-out*	£5,785	£121	4.762	0.022	£5,500
NT-proBNP *rule-in* and *rule-out*	£5,684	£20	4.758	0.018	£1,111

BNP: brain natriuretic peptide; N: N-terminal pro-B-type natriuretic peptide; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life-year

7. Organisational Issues

Expert reviewers stated a number of barriers to using BNP and NT-proBNP in the ED:

Poor selection of patients for natriuretic peptide testing may generate unnecessary investigations and demand on cardiology services. Inappropriate requesting of the test may lead to a lack of confidence in the usefulness of test.

- Funding a testing service.
- Understanding of limitations of the test.
- Training and education of clinicians.
- Turnaround times for results.
- Provision of a testing service on-site for each ED.
- Access to, and review of, results in a timely fashion before patients leave the ED.
- Agreement of a robust protocol to guide testing a management of patients.
- Availability of specialists (e.g. cardiologists or heart failure nurses) for patient review.
- Availability of urgent echocardiography.
- Failure to act on a result below the cut-off (for example, to cancel a planned echocardiogram).

7.1 Welsh context

The NHS Wales (2019) Cardiovascular Atlas of Variation was produced at the request of Welsh Government to identify and investigate unwarranted variation in key aspects of cardiac care in Wales. According to the Cardiovascular Atlas of Variation, between the years of 2016 and 2018, only one hospital in Wales was routinely using natriuretic peptides in AHF admissions, and it recommends that access to timely BNP and echocardiography is improved in NHS Wales (NHS Wales 2019). Expert reviewers suggest that although timely access to echocardiography still remains a problem in Wales, more hospital EDs now have access to natriuretic peptide testing for AHF, but it is a rapidly changing situation and the COVID-19 pandemic has also had an impact on the provision of these tests.

8. Patient issues

An Ontario Health Technology Assessment of the use of BNP and NT-proBNP as diagnostic tests in adults with suspected heart failure was published in 2021. As part of the appraisal, they interviewed six people and received survey feedback from an additional 15 people from Canada: six of the 21 participants had undergone diagnostic testing for heart failure, all of whom were diagnosed with heart failure. The remaining participants were family members and caregivers of patients who had received diagnostic testing for heart failure. They reported that people they interviewed gave BNP and NT-proBNP testing strong support. The main reason for support from participants was the potential time saved by receiving a speedier diagnosis. The overall process, from diagnosis to treatment, was described as being a substantial emotional burden for patients and caregivers, which also impacts them financially and disrupts their quality of life (Ontario Health 2021).

9. Conclusions

Based on evidence from two meta-analyses published since NICE AHF guidelines in 2014, and additional observational studies, natriuretic peptide biomarker levels usually have higher sensitivity than specificity, with sensitivity decreasing and specificity increasing as the natriuretic peptide threshold increases. This suggests that natriuretic peptides may be more useful for ruling-out than ruling-in AHF in the ED setting. However, the majority of the evidence we identified did not identify diagnostic test accuracy using a *rule-in* and *rule-out* test in combination. The evidence identified found that using different BNP and NT-proBNP thresholds leads to different trade-offs in sensitivity and specificity. If BNP and NT-proBNP were to be recommended, consideration would need to be given to the most appropriate threshold(s) to use.

Whilst the majority of the evidence identified evaluated the diagnostic accuracy of NT-proBNP and BNP, there is some evidence that NT-proBNP and BNP testing might improve clinical outcomes compared to clinical judgement. One meta-analysis found that BNP or NT-proBNP testing can reduce the length of hospital stay by at least one day (Lam et al. 2010). The authors suggest that this may be because knowledge of natriuretic peptide levels facilitates better acute management and subsequently more rapid discharge. However, it is unclear what *rule-in*/*rule-out* thresholds were used in this meta-analysis. RCTs found that NT-proBNP testing reduced the 60-day re-admission rate (Moe et al. 2007), and that BNP testing reduced the time spent in hospital over the following year (Breidthardt et al. 2007).

Secondary evidence identified reports that NT-proBNP and BNP levels rise with increasing age. An ESC Position Statement notes that age-dependent *rule-in* cut-offs are preferred for NT-proBNP, as NT-proBNP concentration correlates more strongly with age and renal dysfunction than BNP. But independently of age, an NT-proBNP concentration of less than 300 ng/L and BNP concentration of less than 100 ng/L provides a high NPV for AHF. However, one of the studies we identified that used NT-proBNP *rule-in* and *rule-out* thresholds in combination, found limited differences in accuracy for people with advanced age (Darche et al. 2017).

Using a *rule-in* and *rule-out* strategy for natriuretic peptide testing means that there is a 'grey area' between the thresholds. Patients with levels in the 'grey area' need extra physician attention and ancillary testing. One study we identified found that the diagnostic performance of NT-proBNP within the 'grey zone' was comparable with that in the entire population (Darche et al. 2017).

The quality of the studies included in this EAR is variable. Many of the included studies have significant weaknesses. A lot of the studies were heterogeneous in terms of patient characteristics and specificity, and no UK-specific data were available, making it's generalisability to the Welsh NHS uncertain. Most of the studies investigating NT-proBNP and BNP were observational (prospective cohort) studies: results from the observational studies are likely to be of lower certainty than those from RCTs. There is no single diagnostic test for heart failure. The majority of studies used a reference standard of clinical assessment and interpretation of radiographic findings. With no definitive reference standard, the evaluation of natriuretic peptides needs to be considered in the context of a condition with a variable presentation.

No studies included in the EAR compared a BNP or NT-proBNP strategy using both *rule-in* and *rule-out* with a strategy using *rule-out* thresholds only. We adapted a manufacturer-submitted cost-utility analysis and the model undertaken for NICE CG187 (NICE 2014). In the deterministic analysis, the NT-proBNP *rule-in*/*rule-out* strategy was found to be the optimal strategy when using the dominance rank approach. In deterministic sensitivity analyses, the results were largely robust with the NT-proBNP *rule-in* and *rule-out* strategy found to be optimal in most modelled scenarios. The one exception was a scenario in which it was assumed that false negative results

are not corrected during the hospital stay. The probabilistic sensitivity analysis showed that the BNP *rule-out* strategy had a 58% probability of being cost effective while the NT-proBNP *rule-in* and *rule-out* strategy had a 42% probability of being cost-effective, clinical decision alone had a 0% probability of being cost-effective and NT-proBNP *rule-out* had a 0% probability of being cost-effective, at a threshold of £20,000 per QALY.

10. Contributors

This topic was proposed by Ross Maconachie, Health Economics Manager, Roche Diagnostics Limited.

The HTW staff and contract researchers involved in writing this report were:

- J Washington – literature searches & information management
- J Williams – clinical author
- S Hughes – health economics author
- M Prettyjohns – lead author
- K McDermott – project management
- A Evans – patient and public involvement author

The HTW Assessment Group advised on methodology throughout the scoping and development of the report.

A range of clinical experts from the UK provided material and commented on a draft of this report. Their views were documented and have been actioned accordingly. All contributions from reviewers were considered by HTW's Assessment Group. However, reviewers had no role in authorship or editorial control, and the views expressed are those of Health Technology Wales.

Experts who contributed to this appraisal:

- Mr Geoffrey Armstrong - Principal Clinical Biochemist, Department of Blood Sciences, Wrexham Maelor Hospital, BCUHB
- Dr Mohamed Subhan Anwar - Cardiology research fellow and registrar, Cardiology department, Edinburgh Royal Infirmary
- Mr Ryan Walkley – Health Economist, Roche

11. References

AHTA. (2007). B-type natriuretic peptide assays in the diagnosis of heart failure.: part A - in the hospital emergency setting (November 2006): part B - in the non-hospital setting (May 2007). MSAC Application 1087. Adelaide Health Technology Assessment. Available at: http://www.msac.gov.au/internet/msac/publishing.nsf/Content/B68AAF8993749793CA25801000123B83/$File/1087-Assessment-Report.pdf [Accessed 5 Aug 2021].

Berdagué P, Caffin P-Y, Barazer I, et al. (2006). Use of N-terminal prohormone brain natriuretic peptide assay for etiologic diagnosis of acute dyspnea in elderly patients. American Heart Journal. 151(3): 690-8. doi: https://doi.org/10.1016/j.ahj.2005.04.004

Bombelli M, Maloberti A, Rossi S, et al. (2015). Clinical value of NT-proBNP assay in the emergency department for the diagnosis of heart failure (HF) in very elderly people. Archives of Gerontology and Geriatrics. 61(2): 296-300. doi: https://doi.org/10.1016/j.archger.2015.05.001

Breidthardt T, Laule K, Strohmeyer AH, et al. (2007). Medical and economic long-term effects of B-type natriuretic peptide testing in patients with acute dyspnea. Clinical Chemistry. 53(8): 1415-22. doi: https://doi.org/10.1373/clinchem.2006.081448

Chenevier-Gobeaux C, Delerme S, Allo J-C, et al. (2008). B-type natriuretic peptides for the diagnosis of congestive heart failure in dyspneic oldest-old patients. Clinical Biochemistry. 41(13): 1049–54. doi: https://doi.org/10.1016/j.clinbiochem.2008.05.012

Curtis L, Burns A. (2019). Unit costs of health and social care 2019. Personal Social Services Research Unit, University of Kent, Canterbury. Available at: https://kar.kent.ac.uk/79286/ [Accessed 9 Aug 2021].

Darche FF, Baumgärtner C, Biener M, et al. (2017). Comparative accuracy of NT-proBNP and MR-proANP for the diagnosis of acute heart failure in dyspnoeic patients. ESC Heart Failure. 4(3): 232-40. doi: https://doi.org/10.1002/ehf2.12150

Hill SA, Booth RA, Santaguida PL, et al. (2014). Use of BNP and NT-proBNP for the diagnosis of heart failure in the emergency department: a systematic review of the evidence. Heart Failure Reviews. 19(4): 421-38. doi: https://dx.doi.org/10.1007/s10741-014-9447-6

Ibrahim I, Kuan WS, Frampton C, et al. (2017). Superior performance of N-terminal pro brain natriuretic peptide for diagnosis of acute decompensated heart failure in an Asian compared with a Western setting. European Journal of Heart Failure. 19(2): 209-17. doi: http://dx.doi.org/10.1002/ejhf.612

Januzzi JL, Camargo CA, Anwaruddin S, et al. (2005). The N-terminal Pro-BNP Investigation of Dyspnea in the Emergency department (PRIDE) study. The American Journal of Cardiology. 95(8): 948-54. doi: https://doi.org/10.1016/j.amjcard.2004.12.032

Januzzi JL, Jr, Chen-Tournoux AA, Christenson RH, et al. (2018). N-terminal pro-B-type natriuretic peptide in the emergency department: the ICON-RELOADED study. Journal of the American College of Cardiology. 71(11): 1191-200. doi: https://dx.doi.org/10.1016/j.jacc.2018.01.021

Januzzi JL, van Kimmenade R, Lainchbury J, et al. (2006). NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the international collaborative of NTproBNP study. European Heart Journal. 27(3): 330-7. doi: https://doi.org/10.1093/eurheartj/ehi631

Kozhuharov N, Sabti Z, Wussler D, et al. (2019). Prospective validation of N-terminal pro B-type natriuretic peptide cut-off concentrations for the diagnosis of acute heart failure. European Journal of Heart Failure. 21(6): 813-5. doi: https://doi.org/10.1002/ejhf.1471

Lam LL, Cameron PA, Schneider HG, et al. (2010). Meta-analysis: effect of B-type natriuretic peptide testing on clinical outcomes in patients with acute dyspnea in the emergency setting. Annals of Internal Medicine. 153(11): 728-35. doi: https://doi.org/10.7326/0003-4819-153-11-201012070-00006

Liteplo AS, Marill KA, Villen T, et al. (2009). Emergency thoracic ultrasound in the differentiation of the etiology of shortness of breath (ETUDES): sonographic B-lines and N-terminal pro-brain-type natriuretic peptide in diagnosing congestive heart failure. Academic Emergency Medicine. 16(3): 201-10. doi: https://doi.org/10.1111/j.1553-2712.2008.00347.x

Magnussen C, Blankenberg S. (2018). Biomarkers for heart failure: small molecules with high clinical relevance. Journal of Internal Medicine. 283(6): 530-43. doi: https://doi.org/10.1111/joim.12756

Martinez-Rumayor AA, Vazquez J, Rehman SU, et al. (2010). Relative value of amino-terminal pro-B-type natriuretic peptide testing and radiographic standards for the diagnostic evaluation of heart failure in acutely dyspneic subjects. Biomarkers. 15(2): 175-82. doi: https://doi.org/10.3109/13547500903411087

Moe GW, Howlett J, Januzzi JL, et al. (2007). N-terminal pro-B-type natriuretic peptide testing improves the management of patients with suspected acute heart failure: primary results of the Canadian prospective randomized multicenter IMPROVE-CHF study. Circulation. 115(24): 3103-10. doi: https://doi.org/10.1161/circulationaha.106.666255

Mueller C, Laule-Kilian K, Schindler C, et al. (2006). Cost-effectiveness of B-type natriuretic peptide testing in patients with acute dyspnea. Archives of Internal Medicine. 166(10): 1081-7. doi: https://doi.org/10.1001/archinte.166.10.1081

Mueller C, McDonald K, de Boer RA, et al. (2019). Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations. European Journal of Heart Failure. 21(6): 715-31. doi: https://doi.org/10.1002/ejhf.1494

National Institute for Cardiovascular Outcomes Research. (2020). Heart failure (heart failure audit): National Heart Failure Audit 2018/2019. Summary Report. Available at: https://www.nicor.org.uk/national-cardiac-audit-programme/heart-failure-heart-failure-audit/ [Accessed 23 Feb 2021].

NHS England. (2021). National Cost Collection for the NHS. 2019/2020. Available at: https://www.england.nhs.uk/national-cost-collection/ [Accessed 9 Aug 2021].

NHS Wales. (2019). NHS Wales cardiovascular atlas of variation. Public Health Wales & NHS Wales Health Collaborative. Available at: https://collaborative.nhs.wales/networks/wales-cardiac-network/nhs-wales-cardiovascular-atlas-of-variation/ [Accessed 28 Jan 2021].

NICE. (2014). Acute heart failure: diagnosis and management. Clinical guideline CG187. National Institute for Health and Care Excellence. Available at: https://www.nice.org.uk/guidance/cg187 [Accessed 23 Feb 2021].

NICE. (2018). Chronic heart failure in adults: diagnosis and management. NICE guideline NG106. National Institute for Health and Care Excellence. Available at: https://www.nice.org.uk/guidance/ng106 [Accessed 23 Feb 2021].

Ontario Health. (2021). Use of B-Type Natriuretic Peptide (BNP) and N-Terminal proBNP (NT-proBNP) as diagnostic tests in adults with suspected heart failure: a health technology assessment. Ontario Health Technology Assessment Series 21(2). Available at: https://www.hqontario.ca/Portals/0/Documents/evidence/reports/hta-use-of-b-type-

natriuretic-peptide-and-n-terminal-probnp-as-diagnostic-tests-in-adults-with-suspected-heart-failure.pdf [Accessed 28 July 2021].

Ponikowski P, Voors AA, Anker SD, et al. (2016). 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 European Heart Journal. 37(27): 2129-200. doi: https://doi.org/10.1093/eurheartj/ehw128

Robaei D, Koe L, Bais R, et al. (2011). Effect of NT-proBNP testing on diagnostic certainty in patients admitted to the emergency department with possible heart failure. Annals of Clinical Biochemistry. 48(3): 212-7. doi: https://doi.org/10.1258/acb.2010.010137

Roberts E, Ludman AJ, Dworzynski K, et al. (2015). The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. BMJ. 350: h910. doi: http://dx.doi.org/10.1136/bmj.h910

Rutten JHW, Steyerberg EW, Boomsma F, et al. (2008). N-terminal pro-brain natriuretic peptide testing in the emergency department: beneficial effects on hospitalization, costs, and outcome. American Heart Journal. 156(1): 71-7. doi: https://doi.org/10.1016/j.ahj.2008.02.021

Shaikh K, Ahmad M. (2011). Diagnostic significance of NT-proBNP estimation in patients with acute dyspnea. Journal of the College of Physicians and Surgeons Pakistan. 21(10): 584-8. doi: https://doi.org/10.2011/JCPSP.584588

Siebert U, Milev S, Zou D, et al. (2021). Economic evaluation of an N-terminal pro B-type natriuretic peptide-supported diagnostic strategy among dyspneic patients suspected of acute heart failure in the emergency department. American Journal of Cardiology. 147: 61-9. doi: https://doi.org/10.1016/j.amjcard.2021.01.036

Taylor CJ, Ordóñez-Mena JM, Roalfe AK, et al. (2019). Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. BMJ. 364: l223. doi: https://doi.org/10.1136/bmj.l223

Appendix 1. Table of commercially available BNP, NT-proBNP and MR-proANP assays

Manufacturer and test name	Assay type
NT-proBNP	
Abbott Alinity	Hospital laboratory
BioMérieux VIDAS	Hospital laboratory
Mitsubishi Chemical PATHFAST	POC
Nanogen LifeSign DXpress Reader	POC
Ortho-Clinical Diagnostic NT-proBNP II (analysed using VITROS ECi/ECiQ/3600 Immunodiagnostic Systems and the VITROS 5600/XT 7600 Integrated Systems)	Hospital laboratory
PerkinElmer NT-proBNP AlphaLISA (analysed using EnVision-Alpha Reader)	Hospital laboratory
Radio-meter AQT90 FLEX	POC
Response Biomedical RAMP	POC
Roche CARDIAC proBNP+ (analysed using Cobas h 232)	POC
Roche NT-proBNP I and NT-proBNP II (analysed using Elecsys E170, Cobas e601, e602)	Hospital laboratory
Siemens (Dade Behring): Atellica Solution and ADVIA Centaur Systems, Dimension EXLTM systems, Dimension Vista Systems, Stratus CS systems, Immulite	Hospital laboratory
BNP	
Abbott Alinity I BNP Assay and ARCHITECT systems	Hospital laboratory
Abbott i-STAT	POC
Biosite Triage	POC
ET Healthcare Pylon	POC
Fujirebio Lumipulse G G1200 and G600II	Hospital laboratory
Siemens Advia Centaur CP System	Hospital laboratory
Tosoh AIA	Hospital laboratory
Quidel Triage BNP Test	POC
BNP: B-type natriuretic peptide; NT-proBNP: N-terminal pro B-type natriuretic peptide; POC: point-of-care	

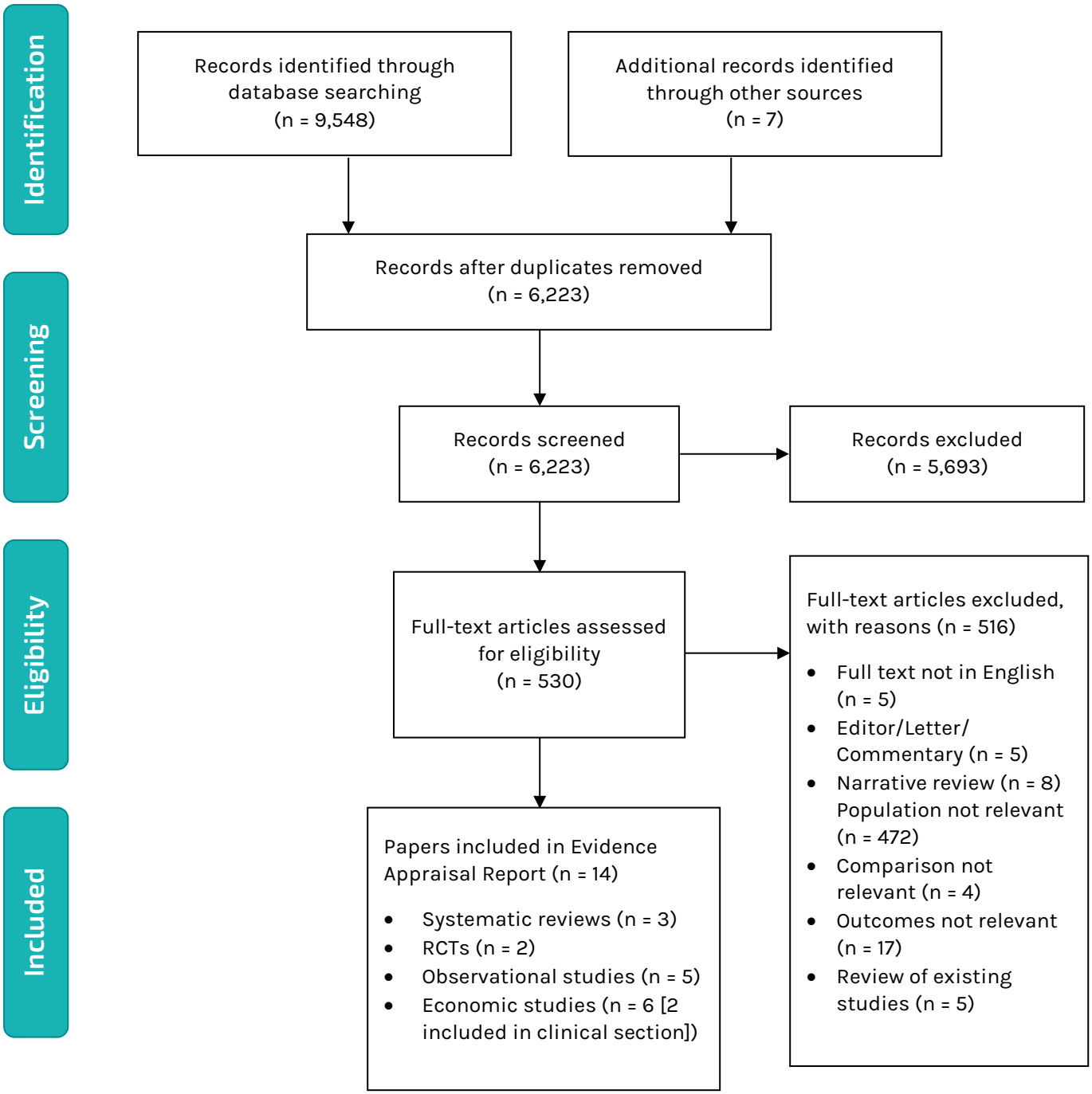
Appendix 2. PICO framework

Research question	What is the clinical and cost effectiveness of N-terminal pro B-type natriuretic peptide (NTproBNP) and B-type natriuretic peptide (BNP) as diagnostic tests to *rule-in* and *rule-out* acute heart failure (AHF) in adults in the emergency department (ED) setting, compared with NT-proBNP and BNP to solely *rule-out* AHF

	Inclusion criteria	**Exclusion criteria**
Population	Adults with suspected acute heart failure in the emergency department Children with suspected acute heart failure	Non acute care (Primary care and community) settings Suspected cardiovascular conditions other than heart failure
Intervention	N-terminal pro-brain natriuretic peptide (NT-proBNP) (to rule in and rule out heart failure) in addition to standard of care B-type natriuretic peptide (BNP) (to rule in and rule out heart failure in addition to standard of care	
Comparison/ Comparators	NT-proBNP (rule out threshold **<300pg/ml**) in addition to standard of care BNP (rule out) in addition to standard of care Standard of care: Clinical examination and standard investigations (e.g. electrocardiography, chest X-ray)	
Reference standard	Using ECG, chest X-ray and blood tests plus clinical judgement	
Outcome measures	Diagnostic accuracy outcomes • Sensitivity • Specificity • Positive predictive value (PPV) • Negative predictive value (NPV) Clinical effectiveness and patient outcomes • Changes in patient management • Mortality • Morbidity • Quality of life	

Study design	We will include the following clinical evidence in order of priority: • Systematic reviews. • Randomised trials (e.g. test and treat studies) • Non-randomised trials • Cross sectional studies, retrospective or prospective case reviews and cohort studies We will only include evidence for "lower priority" evidence where outcomes are not reported by a "higher priority" source. We will also search for economic evaluations or original research that can form the basis of an economic assessment.
Search limits	We will only include evidence published in English language A date limit of 2014 will be applied as our aim is to consider evidence that was not previously reviewed by NICE and SHTG.
Other factors	There is NICE guidance on NT-proBNP for 'ruling out' heart failure, so this appraisal should focus on 'ruling in' heart failure

Appendix 3. PRISMA flow diagram outlining selection of papers for clinical and cost effectiveness (from 2014 – present)



Appendix 4. Description and findings of primary diagnostic accuracy studies for NT-proBNP

Ibrahim et al. (2017) compared the diagnostic test performance of the Elecsys NT-proBNP immunoassay (Roche) in EDs in 606 people in Singapore and 500 people in New Zealand, to see whether ethnicity influences test performance. The study also aimed to identify potential differences in the relationship of age to NT-proBNP and the relationship of age to the diagnosis of acute decompensated heart failure (ADHF) between countries.

NT-proBNP discriminated ADHF from other causes of acute dyspnoea better in Singapore than New Zealand, with an area under the receiver operating characteristic (ROC) curve of 92.6% and 86.6%, respectively (p = 0.012). Application of the single cut-off *rule-out* value of 300 ng/L yielded comparable sensitivity and NPVs in both countries, but better specificity (73% versus 42%) and accuracy (79% versus 62%) in Singapore, indicating that the results would correctly allocate patients to ADHF versus non-ADHF categories 17 times more per 100 patients assessed in Singapore compared with New Zealand (Table 1 and Table 3).

Even though the use of the designated age-adjusted *rule-in* cut-off points improved results in both countries (with specificity improving to 86% and 71%, and accuracy to 86% and 77% in Singapore and New Zealand, respectively), the better performance in Singapore remained apparent. (Table 2 and Table 4) (Ibrahim et al. 2017).

Table 1. Diagnostic accuracy of NT-proBNP for adults of all ages (*rule-out*) presenting to the ED in New Zealand (Ibrahim et al. 2017)

Index test	Reference test		
All ages (threshold 300 ng/L)	**AHF present**	**AHF absent**	
AHF present	175	185	**PPV**: 49%
AHF absent	5	135	**NPV**: 96%
	Sensitivity: 97%	**Specificity**: 42%	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

Table 2. Diagnostic accuracy of age-specific *rule-in* thresholds of NT-proBNP for adults presenting to the ED in New Zealand (Ibrahim et al. 2017)

Index test	Reference test		
Age-specific *rule-in* thresholds	**AHF present**	**AHF absent**	
AHF present	158	93	**PPV**: 63%
AHF absent	22	227	**NPV**: 92%
	Sensitivity: 88%	**Specificity**: 71%	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

Table 3. Diagnostic accuracy of NT-proBNP for adults of all ages *(rule-out)* presenting to the ED in Singapore (Ibrahim et al. 2017)

Index test	Reference test		
All ages (threshold 300 ng/L)	**AHF present**	**AHF absent**	
AHF present	141	125	**PPV**: 54%
AHF absent	4	336	**NPV**: 99%
	Sensitivity: 97%	**Specificity**: 73%	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

Table 4. Diagnostic accuracy of age-specific *rule-in* thresholds of NT-proBNP for adults presenting to the ED in Singapore (Ibrahim et al. 2017)

Index test	Reference test		
Age-specific *rule-in* thresholds	**AHF present**	**AHF absent**	
AHF present	130	64	**PPV**: 67%
AHF absent	16	396	**NPV**: 96%
	Sensitivity: 89%	**Specificity**: 86%	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

Januzzi et al. (2018) aimed to validate the age-specific NT-proBNP *rule-in* cut-offs identified in the first ICON study (Januzzi et al. 2006). Of 1,461 patients from 19 EDs in the USA and Canada, 19% were adjudicated as having AHF.

The NT-proBNP assay used in the study was Elecsys proBNP II (Roche). Sensitivity for age-stratified cut-offs of 450 , 900, and 1,800 ng/L was 85.7%, 79.3%, and 75.9%, respectively; specificity was 93.9%, 84.0%, and 75.0%, respectively. PPVs were 53.6%, 58.4%, and 62.0%, respectively. Overall, positive LRs across age-dependent cut-offs was 5.99 (95% CI: 5.05 to 6.93) (Table 5 and 6). The study concluded that elevated age-stratified *rule-in* cut-off results may be used as an aid in the diagnosis of AHF, whereas an NT-proBNP concentration < 300 ng/L provides a substantial ability to exclude the presence of AHF (Januzzi et al. 2018).

Table 5. Diagnostic accuracy of NT-proBNP to *rule-in* AHF in patients in the ED (Januzzi et al. 2018)

Index test	Reference test		
Age-specific *rule-in* thresholds	**AHF present**	**AHF absent**	
AHF present	220	157	**PPV**: 58.4% (95% CI [%]: 54.5 to 62.1)
AHF absent	57	1,027	**NPV**: 94.7% (95% CI [%]: 93.5 to 95.8)
	Sensitivity: 79.4% (95% CI [%]: 74.7 to 84.2)	**Specificity**: 86.7% (95% CI [%]: 84.8 to 88.7)	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

Table 6. Diagnostic accuracy of NT-proBNP to *rule-out* AHF in patients in the ED (Januzzi et al. 2018)

Index test	Reference test		
All ages (threshold 300 ng/L)	**AHF present**	**AHF absent**	
AHF present	260	335	**PPV** (%): 43.7 (95% CI [%]: 41.4 to 46.1)
AHF absent	17	849	**NPV** (%): 98.0 (95% CI [%]: 96.9 to 98.8)
	Sensitivity (%): 93.9 (95% CI [%]: 91.0 to 96.7)	**Specificity** (%): 71.7 (95% CI [%]: 69.1 to 74.3)	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

In a research letter by Kozhuharov et al. (2019), NT-proBNP concentrations were taken from 2,053 patients with acute dyspnea in two Swiss EDs (the assay used was not reported); 51% had an adjudicated diagnosis of AHF. For the *rule-in* of AHF, age-dependent cut-off concentrations of NT-proBNP (450 ng/L if < 50 years old, 900 ng/L if 50 to 75 years old, and 1,800 ng/L if > 75 years old) achieved a specificity of 91% (95% CI: 87% to 95%), 84% (95% CI: 81% to 87%), and 81% (95% CI: 76% to 85%), respectively, and a PPV of 60% (95% CI: 45% to 73%), 79% (95% CI: 74% to 82%), and 90% (95% CI: 88% to 92%), respectively, allowing to *rule-in* AHF in 19%, 45%, and 62% of patients, respectively (Table 13). The *rule-out* cut-off for AHF of 300 ng/L, achieved a sensitivity of 98% (95% CI: 97% to 99%), an NPV of 97% (95% CI: 95% to 98%), and allowed to *rule-out* AHF in 29% of patients (Table 7) (Kozhuharov et al. 2019).

Table 7. Diagnostic accuracy of age-specific *rule-in* thresholds of NT-proBNP (Kozhuharov et al. 2019)

Index test	Reference test		
Age-specific *rule-in* thresholds	**AHF present**	**AHF absent**	
AHF present	872	158	**PPV**: 85 (82 to 87)
AHF absent	171	852	**NPV:** 83 (81 to 85)
	Sensitivity: 84% (81 to 86)	**Specificity:** 84% (82 to 87)	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

Table 8. Diagnostic accuracy of NT-proBNP for adults of all ages (*rule-out*) (Kozhuharov et al. 2019)

Index test	Reference test		
All ages (threshold 300 ng/L)	**AHF present**	**AHF absent**	
AHF present	1,023	438	**PPV**: 70% (95% CI: 68 to 72)
AHF absent	20	572	**NPV:** 97% (95% CI: 95 to 98)
	Sensitivity: 98% (95% CI: 97% to 99%)	**Specificity:** 57% (95% CI: 54% to 60%)	

AHF: acute heart failure; CI: confidence interval; ED: emergency department; ng/L: nanograms per litre; NPV: negative predictive value; PPV: positive predictive value

Appendix 5. HTW cost utility analysis

1. Background and objective

An economic analysis was developed to identify the most cost effective strategy of the following:

1. Clinical decision alone
2. NT-proBNP using a single *rule-out* threshold
3. NT-proBNP using *rule-in/rule-out* thresholds
4. BNP using a *rule-out* threshold only

A 'BNP using *rule-in/rule-out* thresholds' strategy was not considered due to the lack of diagnostic accuracy evidence for a *rule-in* threshold.

The cost-utility analysis was based on both the model built for NICE (2014) and the manufacturer-submitted model summarised in section 6.2 of this EAR.

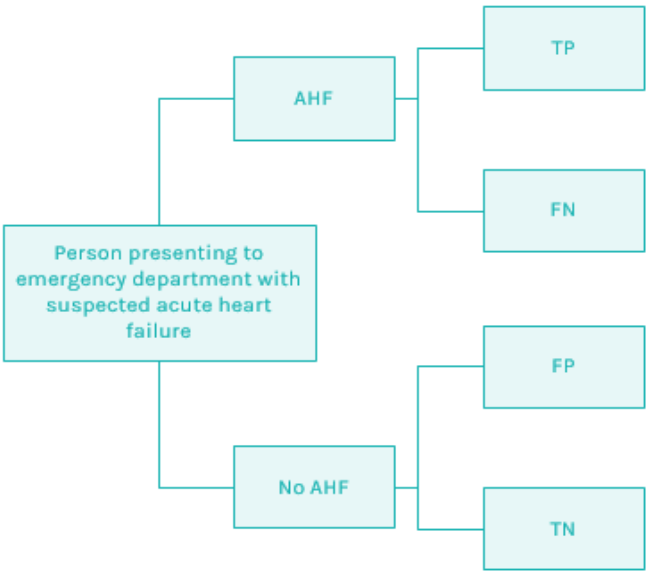
2. Methods

2.1 Model structure

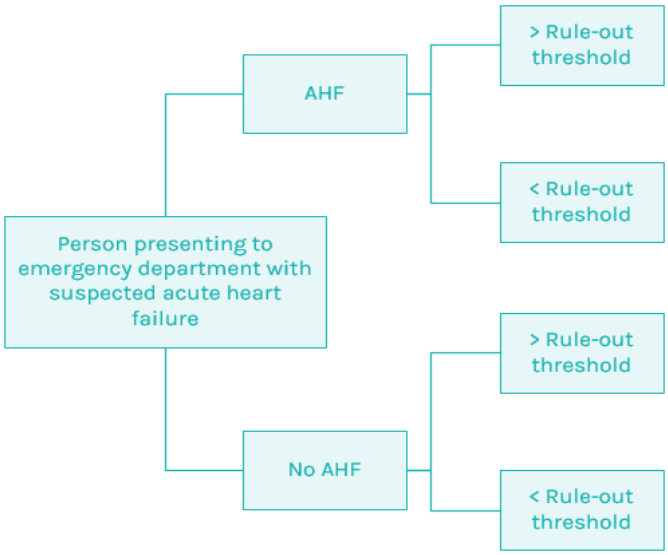
The model, constructed using Microsoft Excel, comprises a decision tree and a Markov model to evaluate the total costs and QALYs of the included strategies. The analysis took the perspective of the UK NHS and personal social services (PSS). A lifetime time horizon was considered, and future costs and benefits were discounted at rate of 3.5%.

The decision tree structure is shown in Figure 1.

(a)



(b)



(c)

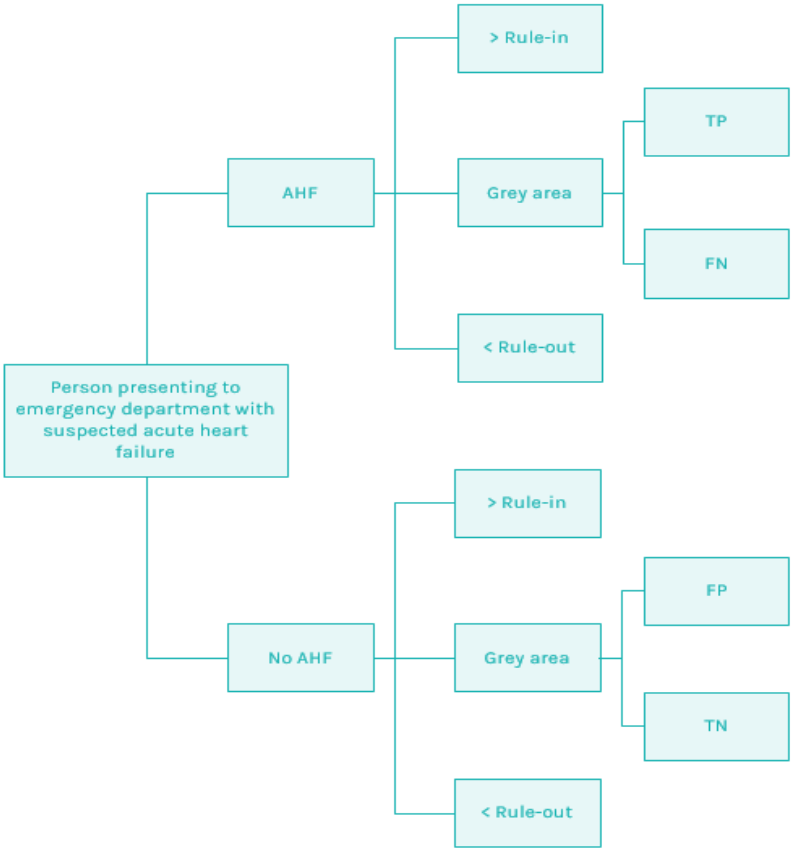


Figure 1. Decision tree structure (a) Decision tree structure for clinical decision alone (b) Decision tree structure for both NT-proBNP and BNP *rule-out* only strategies (c) Decision tree structure for strategy combining *rule-in* and *rule-out* thresholds.

AHF: Acute heart failure; TP: True positive; FN: False negative; FP: False positive; TN: True negative

The model structure mirrors the NICE CG187 model. The CG187 committee assigned an average length of stay to each type of patient, by diagnostic category. Patients have a probability of receiving an echocardiogram, which is assumed to be 100% accurate as in NICE CG187, and of short term mortality. Alive patients with AHF (TPs and FNs) proceed into the long term model whereas FP and TN patients do not. Patients who are alive at the end of the decision tree and have not had their FP status corrected by echocardiogram incur some one-off costs.

Along the lines of the NICE model, our long term model has 3 health states that a patient may occupy during a cycle; 'alive no admission', 'alive admission' and dead. The model has no 'transition probabilities' as such; the occupancy of the dead state is dictated by the proportion above a parametric survival curve in any given cycle and, among those that are alive, the proportion of people who are admitted and not admitted in any given cycle is dictated by cycle-specific probabilities.

2.2 Clinical data

2.2.1 Prevalence and accuracy data

The prevalence of AHF in people presenting to the emergency department with suspected AHF was taken from NICE CG187 (NICE 2014). The population entering the model was 44% female, with baseline age of 77.

Table 1. Model inputs: Acute heart failure prevalence

Scenario	Input	Mean	α, β, distribution	Source
Base case	AHF Prevalence	47%	722, 864, Beta	NICE (2014)
Scenario	AHF Prevalence	19%	-	Januzzi et al. (2018)

AHF: Acute heart failure

The diagnostic accuracy of clinical decision alone was obtained from NICE (2014). In a sensitivity analysis, the diagnostic accuracy used in the manufacturer-submitted model was applied. In the base case, it was assumed that the sensitivity and specificity of clinical judgement for people in the 'grey zone' was equal to the clinical judgement without natriuretic testing. In a sensitivity analysis, the diagnostic accuracy of clinical judgement in people with NT-proBNP results falling within the 'grey zone' from Darche et al. (2017) was used.

The diagnostic accuracy at the NT-proBNP and BNP *rule-out* thresholds was obtained from a systematic review included within this EAR (Roberts et al. 2015). In a sensitivity analysis, data from the manufacturer-submitted model were used for the diagnostic accuracy of NT-proBNP *rule-out* threshold. Data from the manufacturer-submitted model were used for the diagnostic accuracy of the NT-proBNP *rule-in* threshold. These data are calculated using the probability that true positives have NT-proBNP results above the *rule-in* threshold and the probability that false positives have NT-proBNP results above the *rule-in* thresholds. These probabilities were calculated in the manufacturer-submitted model using raw data from the BASEL RCT (Kozhuharov et al. 2019).

Table 2. Model inputs: Diagnostic accuracy of clinical judgement, NT-proBNP and BNP

Scenario	Input	Mean	α, β, distribution	Source
NT-proBNP *Rule-out* (<300ng/L)				
Base case	Sensitivity	99%	1678, 17, Beta	Roberts et al. (2015)
	Specificity	43%	711, 943, Beta	
Scenario	Sensitivity	98%	-	Manufacturer-submitted model
	Specificity	45%	-	
***Rule-in* probabilities**				
Base case	Probability that a true positive is above the age-specific NT-proBNP rule-in threshold	86.1%	880, 142, Beta	Manufacturer-submitted model
	Probability that a false positive is above the age-specific NT-proBNP rule-in threshold	36.4%	158, 276, Beta	
BNP *Rule-out* (<100ng/L)				
Base case	Sensitivity	95%	2897, 152, Beta	Roberts et al. (2015)
	Specificity	63%	2458, 1443, Beta	
Clinical decision				
Base case	Sensitivity	80%	595, 149, Beta	NICE (2014)
	Specificity	77%	648, 194, Beta	
Scenario	Sensitivity	78%		Manufacturer-submitted model
	Specificity	81%		
Clinical decision in the grey zone				
Base case	Sensitivity	80%	595, 149, Beta	Assumption
	Specificity	77%	648, 194, Beta	
Scenario	Sensitivity	71.4%	-	Darche et al. (2017)
	Specificity	82.6%	-	

AHF: Acute heart failure; BNP: B-type natriuretic peptide; NT-proBNP: N-terminal proB-type natriuretic peptide

2.2.2 Other probabilities

Patients have a probability of receiving an echocardiogram, which is assumed to be 100% accurate as in NICE CG187 (NICE 2014). This means there are no false positives in the long-term model.

NICE CG187 assumed that 80% of false negatives eventually get an echocardiogram, which converts them to true positives. Alive patients with AHF (TPs and FNs) proceed into the long term model whereas FP and TN patients do not.

Table 3. Model inputs: Echocardiogram and false negative correction

Input	Mean	α, β, distribution	Source
Proportion of false negative diagnoses corrected in hospital	80%	2.4, 0.6, Beta	NICE (2014)
Proportion receiving an echocardiogram	88%	42491, 5794, Beta	Manufacturer-submitted model, from scenario based on NICOR

The proportions of patients receiving different treatment options in the long-term model were obtained from the manufacturer model.

Table 4. Model inputs: Follow-up proportions

Input	Mean	α, β, distribution	Source
Heart failure nurse follow-up	58%	22020, 16038, Beta	Manufacturer-submitted model
Cardiology follow-up	47%	27676, 31209, Beta	Manufacturer-submitted model

Based on the manufacturer analysis of NICOR 2019, at the start of the long term model a proportion of true positives start on each follow-up service.

Table 5. Model inputs: Treatment proportions

Input	Mean	α, β, distribution	Source
MRA	53%	31209, 27676, Beta	Manufacturer-submitted model
Beta blockers	89%	52407, 6477, Beta	Manufacturer-submitted model
ACE/ARB	84%	49462, 9421, Beta	Manufacturer-submitted model

2.2.3 Readmissions

The manufacturer-submitted model used the same approach as NICE CG187 to calculate readmissions per cycle in the long-term model (NICE 2014). NICE CG187 used a UK study of AHF patients over a four-year follow-up period for readmissions in true positives. Beyond this follow-up period, it was assumed that the monthly proportion of admissions remained constant.

Table 6. Model inputs: Readmission data by cycle (NICE 2014)

Cycle	Mean	α, β for Beta distribution
1	17.8%	52.3, 241.7
2	8.1%	22.6, 256.7
3	7.1%	18.8, 245.8
4	5.8%	14.5, 235.4
5	3.2%	7.5, 227.7
6	5.7%	12.6, 207.9
7	2.5%	5.1, 200.7
8	6.0%	11.5, 179.6
9	3.8%	6.7, 169.7

Cycle	Mean	α, β for Beta distribution
10	3.6%	5.8, 155.9
11	3.4%	5.0, 142.0
12	3.3%	4.4, 127.9
13	3.1%	3.6, 114.0
14	3.0%	3.1, 99.8
15	2.9%	2.6, 85.6
16 (and thereafter)	2.8%	2.1, 71.4

For readmissions in false negatives, we used the same approach as the manufacturer-submitted model. The readmissions in table 6 were adjusted by the relative risks in table 7 from the NICE CG187 model.

Table 7. Model inputs: Treatment effect on hospital admission

Input	Mean	SE, distribution	Source
Log relative risk of hospital admission without AA	0.43	0.094, Normal	NICE (2014)
Log relative risk of hospital readmission without ACHi	0.22	0.035, Normal	NICE (2014)
Log relative risk of hospital readmission without beta blockers	0.26	0.039, Normal	NICE (2014)

2.2.4 Mortality

For in-hospital mortality, we used the same approach as the manufacturer-submitted model. The model used the NICOR dataset to estimate the proportion of people confirmed with AHF (true positives) accessing specialist care, and weighted mortality accordingly. It was assumed that no false negatives accessed specialist care. In a sensitivity analysis, it was assumed that 50% of false negatives accessed specialist care. In a second sensitivity analysis, we assumed no difference in in-hospital mortality. Alive patients with AHF (TPs and FNs) proceed into the long term model whereas FP and TN patients do not.

Table 8. Model inputs: In-hospital mortality

Input	Mean	α, β, distribution	Source
In-hospital mortality for people accessing specialist care (true positives)	8.6%	4153, 44133, Beta	Manufacturer-submitted model
In-hospital mortality for people not accessing specialist care (false negatives)	14.6%	1547, 9052, Beta	Manufacturer-submitted model
Proportion of true positives accessing specialist care	82%	48286, 10599, Beta	Manufacturer-submitted model
Proportion of false negatives accessing specialist care	0%	Fixed	Assumption (manufacturer)

For long-term mortality, we used the same approach as the manufacturer-submitted model. In the manufacturer model, a gamma curve (providing the best fit) was fitted to data from a Kaplan-Meier curve showing 17-year survival in people with AHF in the UK was obtained from (Taylor et al. 2019). For survival in false negatives, a composite hazard ratio was applied to the true positive

curve based on the treatments they are assumed to receive (beta blockers, ACE/ARB and cardiologist input).

Table 9. Model inputs: Treatment effect on mortality

Input	Mean	SE, distribution	Source
Log hazard ratio for mortality without beta blocker	0.14	0.047, Normal	Manufacturer-submitted model
Log hazard ratio for mortality without ACE inhibitor and/or ARB	0.36	0.042, Normal	Manufacturer-submitted model
Log hazard ratio for mortality without cardiology follow-up	0.43	0.040, Normal	Manufacturer-submitted model

2.3 Resource use and costs

The costs considered in the model reflect the perspective of the analysis, thus only costs that are relevant to the UK NHS & PSS were included. Where possible, all costs were estimated in 2021 prices.

The cost of the NT-proBNP test was based on the list price used in the manufacturer model (£24.53). The cost of the BNP test was estimated to be £21.69, based on the unit cost listed in the NICE guideline (NICE 2018). Blood test costs and the cost of an echocardiogram were based on values from NHS Reference costs (NHS England 2021).

The cost associated with hospital stay was based on estimates from NHS Reference costs (NHS England 2021). Following the approach applied in the manufacturer model, a separate cost was applied for AHF and non-AHF related admissions. Excess bed day costs for AHF and non-AHF conditions were also estimated and applied to bed day increments associated with a false negative or false positive diagnosis. These bed day increments were based on estimates used in the NICE CG187 model (NICE 2014).

The cost of treatments for AHF (beta blockers, ACE inhibitors and aldosterone agonists) were based on unit cost estimates applied in the manufacturer and NICE CG187 model (NICE 2014).

Resource use in cardiology follow-up, HF nurse follow-up and GP follow-up was based on estimates applied in the NICE CG187 model (NICE 2014). The cost associated with cardiology follow-up was based on the cost of a cardiology outpatient visit from NHS Reference costs (NHS England 2021). The cost of GP follow-up and nurse follow-up was based on estimates of GP and nurse time from the Personal Social Services Research Unit (PSSRU) (Curtis & Burns 2019).

Table 10. Model inputs: Unit costs

Input	Mean	α, β, distribution	Source
Test costs			
NT-proBNP	£24.53	6.829, 3.592	Manufacturer-submitted model
BNP	£21.69	Fixed	NICE (2018)
Echocardiogram	£72	Fixed	Manufacturer-submitted model
Blood test	£2	Fixed	Manufacturer-submitted model
Hospital stay costs			
Excess bed day (AHF)	£320.83	96, 3, Gamma	NHS England (2021)
Excess bed day (all conditions)	£349.90	96, 4, Gamma	NHS England (2021)

Input	Mean	α, β, distribution	Source
AHF spell	£3,690.17	16, 203.6, Gamma	NHS England (2021)
Non-AHF spell	£3,292.75	16, 205.8, Gamma	NHS England (2021)
Treatment costs			
ACE/ARB (per day)	£0.11	Fixed	Manufacturer-submitted model
Beta blocker (per day)	£0.06	Fixed	Manufacturer-submitted model
MRA (AA) (per day)	£0.12	Fixed	Manufacturer-submitted model
Follow-up costs			
GP visit	£42.32	Fixed	Manufacturer-submitted model
Community HFSN visit	£45.94	Fixed	Manufacturer-submitted model
Cardiology outpatient visit	£139	Fixed	Manufacturer-submitted model

AHF: Acute heart failure; HFSN: Heart failure specialist nurse

Table 11. Model inputs: Resource use

Input	Mean	α, β, distribution	Source
False positive and false negative increments			
Bed day increment for false positives	2	2, 0.51, Gamma	NICE (2014)
Bed day increment for false negatives	2	2, 0.51, Gamma	NICE (2014)
Cardiology follow-up			
Outpatient visits (first year)	2	Fixed	NICE (2014)
Outpatient visits (subsequent years)	1	Fixed	NICE (2014)
NP tests	2	Fixed	NICE (2014)
Blood tests	2	Fixed	NICE (2014)
HF nurse follow-up			
GP visits	3	Fixed	NICE (2014)
Community HFSN visits	4	Fixed	NICE (2014)
No HF nurse follow-up			
GP visits	7	Fixed	NICE (2014)

AHF: Acute heart failure; HFSN: Heart failure specialist nurse

2.4 Health-related quality of life

As recommended in the NICE reference case, the model estimates effectiveness in terms of quality adjusted life years (QALYs). These are estimated by combining life year estimates with quality of life (QoL) values associated with being in a particular health state.

Following the approach adopted in the NICE CG187 and manufacturer model, quality of life was estimated on the basis of whether patients were admitted or not (NICE 2014). The standard utility value for chronic heart failure was therefore applied as the baseline value. An additional disutility is then applied to patients that are admitted for acute heart failure. Using this approach, the QALY loss associated with receiving a false negative diagnosis is based entirely on the increased

probability of admission and death. Table 12 presents the QoL values applied in the economic analysis.

In the base case, it was assumed that there was no quality of life loss associated with receiving a false positive diagnosis. This reflects the high likelihood that an initial false positive diagnosis would be corrected during the subsequent inpatient admission. However, to account for the potential adverse effects of false positivity (primarily that there may be an underlying condition not being adequately treated), some alternative scenarios were explored in sensitivity analysis. We adopted the same approach as that used in the manufacturer model, in which small QALY decrements at arbitrary values were applied to false positive patients that remain undiagnosed after hospital admission.

Table 12. Quality of life values

Health state	QoL value	α, β, distribution	Source
Utility CHF	0.752	966, 318, Beta	NICE (2014)
Dis-utility for 3 months AHF	-0.064	16, 0.004, Gamma	NICE (2014)
QoL: quality of life			

3. Results

3.1 Base case results

The base case results of the analysis are shown in Table 13 and Table 14. Table 13 shows the results for BNP and NT-proBNP strategies in comparison to clinical decision alone. It can be seen that all three BNP and NT-proBNP strategies are more costly and more effective than clinical decision alone. The resulting ICERs are below £20,000 per QALY indicating that the BNP and NT-proBNP strategies are cost-effective in comparison to clinical decision alone.

Table 14 shows the results of the analysis using a 'dominance rank' approach, which allows for the optimal strategy to be determined. Under this approach, strategies are firstly arranged in order of total cost, from cheapest to most expensive. Incremental costs and QALYs are then calculated for each intervention by comparing it against the previous intervention that was found to be cost-effective (at a threshold of £20,000 per QALY). The NT-proBNP *rule-in/rule-out* strategy was found to be the optimal strategy when using the dominance rank approach.

Table 13. Base case results versus clinical decision alone

Diagnostic strategy	Cost		QALYs		ICER (cost per QALY)
	Total	Incremental	Total	Incremental	
Clinical decision	£5,664	-	4.740	-	-
BNP *rule-out*	£5,713	£49	4.757	0.017	£2,882
NT-proBNP *rule-out*	£5,785	£121	4.762	0.022	£5,500
NT-proBNP *rule-in* and *rule-out*	£5,684	£20	4.758	0.018	£1,111
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life-year					

Table 14. Base case results using 'dominance rank' approach

Diagnostic strategy	Cost		QALYs		ICER (cost per QALY)
	Total	Incremental	Total	Incremental	
Clinical decision	£5,664	-	4.740	-	-
NT-proBNP *rule-in* and *rule-out*	£5,684	£20	4.758	0.019	£1,044
BNP *rule-out*	£5,713	£29	4.757	-0.001	Dominated
NT-proBNP *rule-out*	£5,785	£101	4.762	0.003	£31,912

ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life-year

3.2 Deterministic sensitivity analysis results

A series of deterministic sensitivity analyses were conducted, whereby an input parameter is changed, the model is re-run and the new cost-effectiveness result is recorded. This is a useful way of estimating uncertainty and determining the key drivers of the model result. The results of the deterministic sensitivity analyses are presented in Table 15. The optimal strategy (using dominance rank) is presented for each scenario considered in the analysis.

It can be seen that the optimal strategy remained unchanged in most of the scenarios considered in the sensitivity analysis. The only exception was a scenario in which it was assumed that false negative results were not corrected in hospital. In this scenario, the better overall sensitivity of the BNP *rule-out* strategy leads to it being the optimal strategy. Another notable scenario was the scenario in which it was assumed that clinical decision is used in patients not ruled-out in BNP *rule-out* and NT-proBNP *rule-out* scenarios. In this scenario, diagnostic accuracy value associated with clinical judgement were applied to patients that were not *ruled-out*. This led to a change in the order of the strategies with NT-proBNP *rule-in*/*rule-out* found to be the most costly strategy but also the most effective. The resulting ICER was below £20,000 per QALY and this the NT-proBNP *rule-in*/*rule-out* remained cost effective and the optimal strategy overall.

Table 15. Deterministic sensitivity analysis results

Modelled scenario	Optimal strategy
Base case	NT-proBNP *rule-in* and *rule-out*
Prevalence of AHF from ICON study (19%)	NT-proBNP *rule-in* and *rule-out*
NT-proBNP *rule-out* accuracy from manufacturer model	NT-proBNP *rule-in* and *rule-out*
Clinical decision accuracy from manufacturer model	NT-proBNP *rule-in* and *rule-out*
Grey zone accuracy from Darche	NT-proBNP *rule-in* and *rule-out*
Assume clinical decision is used in patients not ruled-out in BNP *rule-out* and NT-proBNP *rule-out* scenarios	NT-proBNP *rule-in* and *rule-out*
Assume 50% of FN patients are seen by a specialist	NT-proBNP *rule-in* and *rule-out*
Assume no difference in in-hospital mortality	NT-proBNP *rule-in* and *rule-out*
In-hospital mortality estimate from NICE model	NT-proBNP *rule-in* and *rule-out*
Assume no FNs are corrected in hospital	BNP *rule-out*
Assume 50% of FNs are corrected in hospital	NT-proBNP *rule-in* and *rule-out*
Assume all FNs are corrected in hospital	NT-proBNP *rule-in* and *rule-out*
BNP test cost = NT-proBNP (as it was in NICE model)	NT-proBNP *rule-in* and *rule-out*

Modelled scenario	Optimal strategy
Equivalent excess bed day costs for AHF and non-AHF admissions	NT-proBNP *rule-in* and *rule-out*
Inpatient costs from NICE model (assuming 8-day stay)	NT-proBNP *rule-in* and *rule-out*
A bed day increase of 4 days for TPs and FPs	NT-proBNP *rule-in* and *rule-out*
QALY decrement for FP patients = 0.01	NT-proBNP *rule-in* and *rule-out*
QALY decrement for FP patients = 0.05	NT-proBNP *rule-in* and *rule-out*

AHF: acute heart failure; BNP: B-type natriuretic peptide; NT-proBNP: N-terminal proB-type natriuretic peptide QALY: quality-adjusted life-year; FN: false negative; FP: false positive

3.3 Probabilistic sensitivity analysis results

Probabilistic sensitivity analysis (PSA) was conducted to assess the combined parameter uncertainty in the model. In this analysis, the mean values that were utilised in the base case were replaced with values drawn from distributions around the mean values. The results of 10,000 runs of the PSA are presented using cost-effectiveness acceptability curves (CEACs) in figure 2, which shows the probability of each strategy being considered cost effective at various cost-effectiveness thresholds. At a threshold of £20,000 per QALY, the BNP *rule-out* strategy was found to have a 58% probability of being cost effective while the NT-proBNP *rule-in* and *rule-out* strategy had a 42% probability of being cost-effective, clinical decision alone had a 0% probability of being cost-effective and NT-proBNP *rule-out* had a 0% probability of being cost-effective.

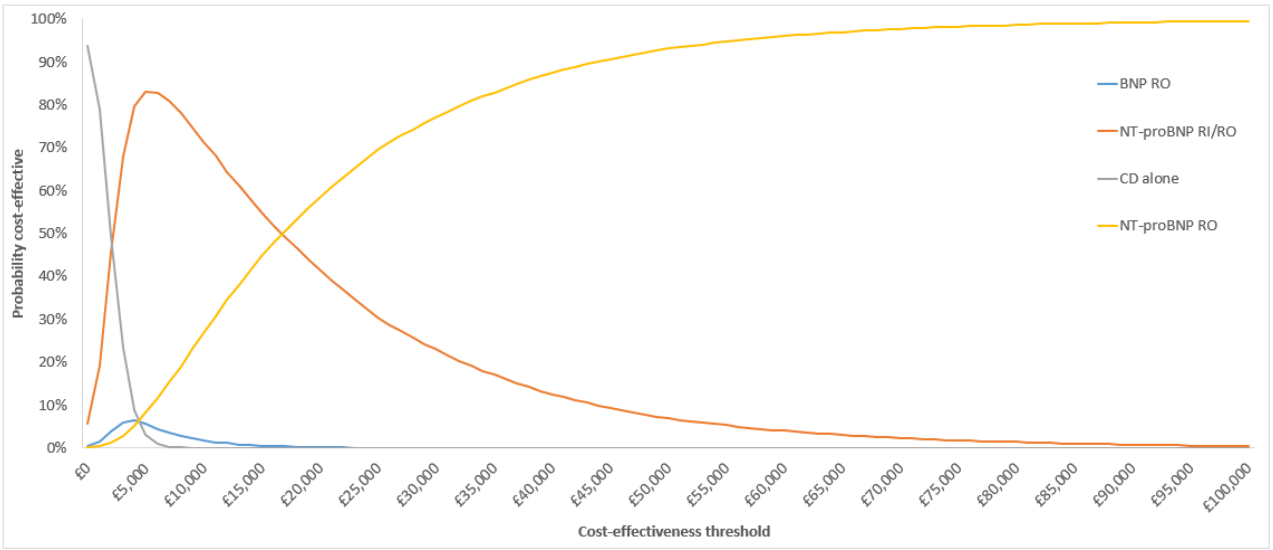


Figure 2. Cost-effectiveness acceptability curves (CEACs)