BPR MCS | Bulletin of Pioneering Researches of Medical and Clinical Science

Bulletin of Pioneering Researches of Medical and Clinical Science

Available online: https://bprmcs.com
2022 | Volume 2 | Issue 1 | Page: 145-153

Genetic Screening for Detecting Individuals at Elevated Risk of Heart Failure

Thomas R. Nguyen[1]*, Daniel K. Moore[1], Jessica A. White[1]

[1]Department of Medical and Clinical Sciences, University of Melbourne, Melbourne, Australia.

Abstract

Heart failure (HF) is a common and serious condition, yet identifying individuals at greatest risk—especially before symptoms appear—remains a challenge. In this study, we introduce and validate a new genetic testing approach designed to predict HF susceptibility, leveraging data from three separate Australian and US cohorts. The first phase utilized the Baker Biobank case–control cohort, revealing 41 genetic variants associated with HF risk through genome-wide association and interaction analyses. A second phase expanded the panel with 29 additional single-nucleotide polymorphisms, and combining both phases produced a comprehensive test demonstrating strong predictive performance, with an Area Under the Curve (AUC) of 0.93 and balanced accuracy of 0.89. Participants identified as high genetic risk in the Baker Biobank cohort showed an odds ratio of 533.2. External validation in the Busselton Health Study and Atherosclerosis Risk in Communities cohorts confirmed the test's reliability, with AUCs of 0.83 and 0.72, balanced accuracies of 0.76 and 0.67, and odds ratios of 12.3 and 4.6, respectively. These results highlight the significant contribution of genetic factors to HF and indicate that this test could provide a powerful tool for early, personalized HF risk prediction.

Keywords: Precision medicine, Personalized risk score, Genetic risk, Complex genetic diseases, Heart failure

Corresponding author: Thomas R. Nguyen
E-mail: thomas.nguyen@outlook.com

How to Cite This Article: Nguyen TR, Moore DK, White JA. Genetic Screening for Detecting Individuals at Elevated Risk of Heart Failure. Bull Pioneer Res Med Clin Sci. 2022;2(1):145-53. https://doi.org/10.51847/XefKtOGquq

Introduction

Heart failure (HF) continues to pose a serious public health challenge globally, affecting more than 26 million individuals and showing rising prevalence among older populations [1]. In the United States, forecasts suggest that by 2030 over 8 million people will live with HF, with direct healthcare expenditures (adjusted to 2010 USD) expected to escalate from USD 21 billion to USD 53 billion [2]. HF is a multifactorial syndrome in which the heart is unable to pump blood effectively or fill adequately, commonly presenting as either systolic or diastolic dysfunction [3, 4]. Systolic HF is characterized by weakened myocardial contraction and reduced ejection fraction, often resulting from myocardial infarction or other cardiac injuries [5], whereas diastolic HF involves stiffened ventricular walls that impede filling despite normal ejection fraction, frequently linked to hypertension and age-related myocardial remodeling [6].

The development of HF is driven by intricate interactions among hemodynamic stress, neurohormonal dysregulation, and cellular remodeling in the myocardium [7]. Structural adaptations such as cardiac hypertrophy and chamber dilation, coupled with processes like apoptosis and fibrosis, contribute to disease progression. Risk factors encompass both clinical and lifestyle components. Hypertension imposes extra cardiac load, coronary artery disease compromises myocardial performance, and diabetes independently increases susceptibility. Lifestyle

behaviors—including smoking, obesity, and unhealthy diets—further exacerbate risk [8, 9].

Early detection and prevention are critical, yet conventional population-level HF risk tools, such as PCP-HF equations, offer only modest accuracy and are heavily weighted toward age rather than capturing the full spectrum of contributing factors [10]. Identifying high-risk individuals before clinical manifestations appear remains particularly challenging [11]. Diagnosing HF is complicated by symptom variability and overlap with other conditions [3, 12]. Patients may present with fatigue, shortness of breath, edema, or weight gain, symptoms that are often nonspecific and can delay recognition [13].

Standard diagnostic approaches, though essential, have limitations. Echocardiography evaluates cardiac structure and function but may fail to detect early or subtle HF changes [14]. Biomarkers such as BNP and NT-proBNP assist in diagnosis but are influenced by factors like renal function, age, and body mass, which can complicate interpretation [15]. Misdiagnosis or delayed recognition can worsen outcomes, increase hospitalization rates, and raise mortality risk [16]. There is a clear need for more accurate, timely diagnostic strategies to improve patient management and prognosis [17].

Genetic testing presents a promising avenue to identify individuals at elevated HF risk long before clinical signs appear [18]. Despite this potential, no reliable genetic risk assessment exists for HF [19]. To address this gap, we aimed to develop a genetic test using three rigorously characterized independent cohorts. In this study, we present the design, validation, and predictive performance of a genetic assay capable of detecting individuals at increased HF risk.

Results and Discussion

First-phase HF genetic test development

The first phase focused on generating a genetic model to discriminate HF cases from controls within the Baker Biobank (BB) cohort. Controls were participants over 70 years old with no history of HF. The cohort was split into discovery and testing subsets. Within the training dataset, 23 single-nucleotide polymorphisms (SNPs) showing the strongest associations were selected, alongside 41 SNPs demonstrating significant interactions.

A multi-layer deep neural network (DNN) was employed to enhance predictive accuracy by iteratively adjusting weights from prior training cycles. The selected SNP genotypes were input into the DNN to construct a predictive model capable of distinguishing HF cases from controls.

From this model, a Quantitative Risk Score (QRS) was derived for each participant. The risk threshold was set at the 90th percentile of control scores (0.402), meaning only 10% of controls exceeded this value. Individuals with QRS above this threshold were classified as high-risk, whereas those below were considered low-risk.

The model achieved an odds ratio (OR) of 2.53 and a hazard ratio (HR) of 5.7, substantially outperforming traditional polygenic risk scores (PRSs) for cardiovascular disease, which generally show ORs below 2. These results demonstrate the potential of this approach to identify individuals at elevated HF risk, supporting its application in preventive cardiology.

Table 1. Evaluation of the first-phase heart failure (HF) genetic test in the reserved "test set" from the Baker Biobank (BB). Statistical analysis of these results yielded a Fisher's exact test p-value of 4.97×10^{-25}, an Area Under the Curve (AUC) of 0.72, a balanced accuracy (BAC) of 0.67, and an odds ratio (OR) of 2.53. The corresponding hazard ratio (HR) was 5.7.

Genetic Risk	No HF	HF Cases
High	530	801
Low	523	313

As presented in **Table 1**, the first-phase test successfully identified HF cases, correctly classifying approximately 72% of them as high-risk. However, its ability to accurately recognize control subjects—those without HF—was limited. Two factors may explain this observation. First, some individuals classified as high-risk may develop HF later in life. Second, HF likely arises from multiple molecular pathways, each influenced by different sets of genetic variants. To address this genetic heterogeneity, we designed a second-phase genetic test.

Development of the second-phase HF genetic test

For the second-phase analysis, all 41 SNPs used in the first test were excluded from the genome build, and association studies were conducted specifically on the subset of individuals flagged as high-risk by Test 1. Applying the same methodological approach as in the first phase, we identified 29 additional SNPs to serve as the foundation for Test 2. The risk threshold for this test was established at the 90th percentile of the control group, corresponding to a QRS value of 0.541. Using this cutoff, the performance of Test 2 was then assessed in the reserved "test set" of BB participants who had been classified as high-risk by the initial test **(Table 2)**.

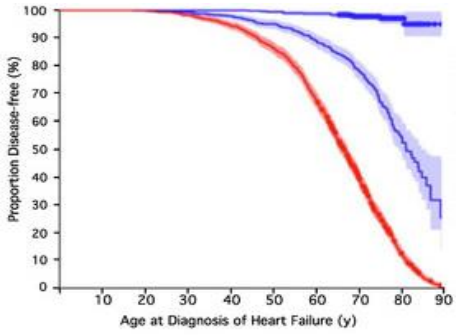
Table 2. Evaluation of the second-phase heart failure (HF) genetic test in the reserved validation set. Analysis of these results produced a Fisher's exact test p-value of 8.1×10^{-170}, an Area Under the Curve (AUC) of 0.89, a balanced accuracy (BAC) of 0.81, and an odds ratio (OR) of 50.29. The hazard ratio (HR), calculated using the full test set including controls younger than 70, was 37.4.

Genetic Risk	No HF	HF Cases
High	118	749
Low	412	52

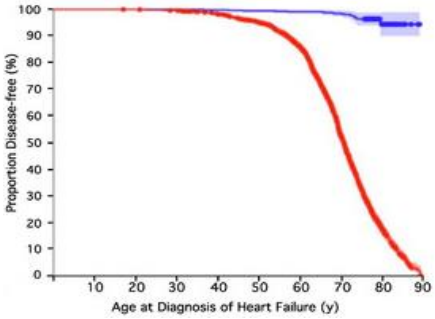
Integrating the two genetic HF risk assessments

The first- and second-phase tests were designed to capture complementary aspects of HF's genetic complexity. To generate a comprehensive measure of overall genetic susceptibility, we combined the results of both tests, producing four distinct participant groups: high-risk on both tests (1_1), high-risk only on the first test (1_0), high-risk only on the second test (0_1), and low-risk on both tests (0_0).

These four groups were further categorized into three broader risk tiers: individuals in the 1_1 group were classified as high-risk, those in 0_0 as low-risk, and participants in 1_0 and 0_1 were designated as intermediate-risk. Kaplan–Meier survival curves and the corresponding 2 × 2 contingency table **(Figure 1 and Table 3)** illustrate the performance of the combined testing approach. The separation of survival trajectories across these groups was highly significant, confirming that the integrated risk assignments reflect meaningful biological differences in HF susceptibility. Most HF cases were concentrated within the high- and intermediate-risk categories (1_1, 1_0, and 0_1), whereas the low-risk group (0_0) contained relatively few cases. The right panel of **Figure 1** additionally shows the effect of merging the intermediate- and high-risk groups, highlighting the stratification power of the combined test.



a)



b)

Figure 1. Kaplan–Meier survival curves depicting the genetic risk stratification for heart failure in the reserved Baker Biobank (BB) "test set," which was not included in model training. Risk categories were determined exclusively from genetic information, without using any clinical data. Left panel: The high-risk group is shown in red, whereas intermediate- and low-risk groups are depicted in blue, with shaded regions representing 95 percent confidence intervals (CI). Right panel: The combined HF test merges intermediate- and high-risk individuals, resulting in an AUC of 0.93, balanced accuracy (BAC) of 0.89, and an odds ratio (OR) of 123.1 ($p = 5.6 \times 10^{-222}$). Across the entire BB cohort, the hazard ratio (HR) for the integrated test was 157.6, highlighting its strong predictive power.

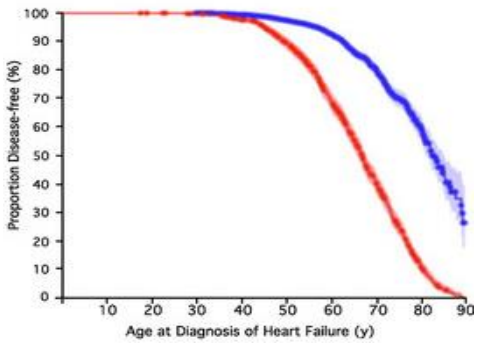
Table 3. Performance metrics for the combined heart failure (HF) genetic test in the reserved "test set" of subjects. Statistical comparisons between risk groups yielded the following results: high- versus low-risk individuals, $p < 10^{-300}$ with an odds ratio (OR) of 533.2; low- versus intermediate-risk individuals, $p = 1.4 \times 10^{-57}$ with an OR of 27.5.

Genetic Risk Group	No HF	HF Cases
High	85	902
Low	603	12

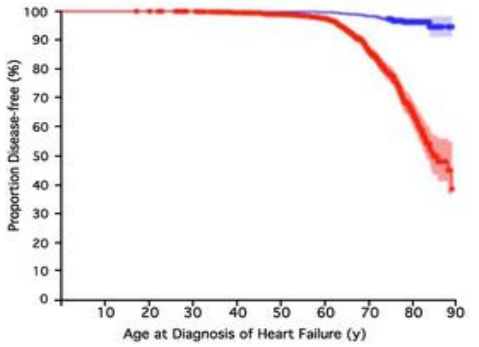
The combined HF genetic test demonstrated strong predictive power, as reflected in the Fisher's exact test p-values and odds ratios shown in **Table 3**. Individuals categorized as high-risk had an odds of developing HF more than forty times greater than those in the low-risk group. Remarkably, over 90% of high-risk participants were diagnosed with HF before reaching 80 years of age, whereas fewer than 5% of low-risk participants ever developed the condition. The intermediate-risk group showed an elevated incidence compared with the low-risk group, with approximately half of these individuals experiencing HF by age 80.

External validation of genetic HF risk tests

To test the generalizability of the genetic risk models, we applied them to two independent population cohorts: the Atherosclerosis Risk in Communities (ARIC) HF cohort and the Busselton Health Study (BHS). Because these were population-based studies, the proportion of HF cases was considerably lower than in the Baker Biobank case–control cohort. Test 1 showed significant predictive performance in both cohorts, with odds ratios exceeding 4 in the BHS cohort **(Table 4)**. Test 2 was similarly validated. For the combined genetic test—merging intermediate- and high-risk groups—the relative risk of HF was 2.6 in the ARIC cohort and 10.5 in the BHS cohort **(Figure 2 and Table 5)**, indicating that individuals with elevated genetic risk were substantially more likely to develop HF compared to those at low genetic risk, particularly in the BHS population.



a)



b)

Figure 2. External validation of the combined genetic HF risk test in the ARIC cohort (left panel) and the BHS cohort (right panel). For the ARIC and BHS cohorts, the p-values were 2.3×10^{-57} and 1.2×10^{-70}, respectively. In this analysis, the intermediate- and high-risk groups were combined. The red curves indicate individuals classified as high-risk, while the blue curves represent those at low or intermediate risk. Shaded regions correspond to 95% confidence intervals (CI).

Table 4. Validation of Test 1 for predicting heart failure (HF) risk in the population-based ARIC and BHS cohorts. Due to the design of these studies, HF cases represent a smaller fraction of participants compared with the BB case–control cohort. Results are stratified by low- and high-genetic-risk groups and HF status. One-sided p-values from Fisher's exact test and odds ratios (ORs) are provided to facilitate comparison with the BB cohort results shown in **Table 1**. The 95% confidence intervals (CIs) indicate the estimated range of effect sizes and the uncertainty of the OR estimates. The analyses revealed highly significant associations, including a relative risk (RR) of 4.18 in the BHS cohort.

Cohort	Genetic Risk Group	HF Cases	No HF	OR	95% CI	p
ARIC	Low	342	507	1.7	1.45–1.99	3.1×10^{-8}
	High	562	498			
BHS	Low	57	1870	4.6	2.97–7.12	1.2×10^{-30}
	High	262	1853			

Table 5. Validation of the combined heart failure (HF) genetic tests in the ARIC and BHS cohorts. Participants were classified into low- and high-genetic-risk groups and stratified by HF status. One-sided p-values (p) and odds ratios (ORs) are provided to illustrate the enhanced predictive accuracy of the combined tests relative to Test 1 alone. The 95% confidence intervals (CIs) reflect the precision of the OR estimates. The relative risk (RR) values were 1.97 for ARIC and 2.28 for BHS.

Cohort	Genetic Risk Group	HF Cases	No HF	95% CI	p	OR
ARIC	Low	241	630	3.06–6.91	2.3×10^{-57}	4.6
	High	663	375			

BHS	Low	36	2272	7.07–21.39	1.2×10^{-70}	12.3
	High	283	1451			

Comparison with clinical risk prediction

In the BB cohort, 1,604 participants without a history of HF at baseline were analyzed to compare genetic versus clinical risk prediction. Thirty commonly assessed clinical factors—including demographics (age, sex, socioeconomic status), lifestyle (alcohol use, smoking), anthropometrics (BMI, waist circumference), blood pressure, serum biomarkers (HDL, LDL, triglycerides, glucose, total cholesterol), comorbidities (stroke, diabetes, MI, hypertension), and use of medications for cholesterol, blood pressure, or diabetes—were incorporated into a Cox proportional hazards model to estimate the risk of developing HF over an average follow-up of 10.1 years. The resulting model produced a C-statistic of 0.75. By comparison, the genetic test alone achieved a higher predictive value with a C-statistic of 0.804. Since HF incidence is higher in older men, including age and sex further increased the C-statistic to 0.835, while adding the remaining clinical variables yielded only a modest increase, reaching 0.86. Overall, combining the genetic test with all clinical predictors improved discrimination by 0.11 relative to using clinical variables alone.

We developed and validated a novel genetic test for HF risk across three independent cohorts, representing an advance in precision cardiovascular medicine by enabling earlier identification of high-risk individuals. The two-tiered design of the test offers insight into the complex genetic underpinnings of HF.

The first-phase test, comprising 41 variants, outperformed conventional polygenic risk scores (PRSs) typically applied to cardiovascular conditions, which generally generate hazard ratios below 1.5 [20]. This enhanced performance highlights the ability of our approach to capture meaningful genetic contributions to HF risk, consistent with evidence that both common and rare variants, together with environmental interactions, influence disease susceptibility [21, 22].

The second-phase test targeted individuals flagged as high-risk in the first phase and incorporated 29 additional SNPs, refining risk prediction and addressing HF's genetic heterogeneity. This two-step approach markedly improved performance, achieving an AUC of 0.93 and balanced accuracy of 0.89 in the BB cohort. This strategy builds on prior findings demonstrating that patient stratification strengthens the detection of genetic effects [23, 24], emphasizing the value of iterative refinement in constructing robust genetic risk models. The integration of machine learning methods for risk stratification further enhances predictive capability, setting a high standard for genetic risk assessment in cardiovascular disease.

Validation in geographically and ethnically diverse cohorts (ARIC and BHS) confirmed the test's reproducibility. It also showed predictive utility in African-American participants within ARIC, indicating potential cross-ethnic applicability. Nevertheless, performance differences across cohorts underscore the need for population-specific refinement, reflecting known variations in allele frequencies and linkage disequilibrium patterns across ethnic groups [25]. Future research should focus on identifying population-specific variants to optimize predictive accuracy in diverse populations.

Limitations of this study include the predominance of European ancestry (73%) in the BB cohort, with 27% representing other ethnicities, which may restrict generalizability. Despite adjustment for population structure via principal component analysis (PCA), residual confounding cannot be excluded. Age and sex differences between cases and controls could introduce bias, given that HF is more prevalent in older men; in the BB cohort, males represented 47% of controls and 53% of cases. However, the genetic test successfully identified women who subsequently developed HF. Genotype imputation was required in ARIC and BHS for missing data, which could slightly reduce accuracy.

HF diagnosis relied on ICD codes assigned independently within each cohort. Although some variability may arise from this approach, prior studies have demonstrated that ICD coding for HF is generally reliable [26]. The large cohort sizes further minimize potential inconsistencies, reinforcing the robustness of our findings.

Our investigation primarily targeted the genetic underpinnings of HF, though environmental and lifestyle elements—such as dietary habits, physical activity, smoking, and socioeconomic factors—were only partially considered, leaving the possibility of residual confounding. Additional influences, including comorbidities and medication usage (e.g., treatments for hypertension, diabetes, or other cardiovascular conditions), may independently affect HF risk. Remarkably, our findings demonstrated that the genetic markers identified improved risk prediction beyond conventional clinical measures, suggesting they capture additional hereditary components. Environmental exposures may explain the limited number of HF cases observed among participants classified as genetically low-risk. Furthermore, survivorship bias—stemming from the inclusion of only controls older than 70 without HF—could lead to underestimation of genetic effects. Expanding validation efforts and integrating these factors into the model are essential for enhancing reliability.

HF represents a genetically and clinically heterogeneous condition, encompassing multiple subtypes such as myocardial ischemia, valvular disease, arrhythmias, and cardiomyopathies [27, 28]. Each subtype is characterized by distinct pathophysiology, which can influence genetic test performance. While our two-step testing framework addressed some of this heterogeneity, predictive accuracy may still vary across HF forms. For example, Joseph *et al*. [28] recently highlighted genetic differences between HFrEF and HFpEF, where most loci were associated with coronary artery disease and hypertension. Future investigations should assess test performance across HF subtypes and consider incorporating echocardiographic parameters to refine predictions further.

The translation of genetic HF testing into clinical settings involves ethical and logistical challenges. Conveying risk information to patients requires careful counseling to ensure comprehension and informed decision-making. Nevertheless, identifying individuals with elevated genetic risk before clinical manifestation offers opportunities for proactive management, including lifestyle interventions such as weight control, blood pressure management, and adherence to a cardioprotective diet, which may delay or prevent HF onset.

This study also highlights directions for future research. The SNPs identified may point to novel regulatory mechanisms, particularly in non-coding genomic regions [29], influencing gene expression through enhancer activity or chromatin remodeling. Investigating these mechanisms could shed light on HF pathogenesis and reveal potential therapeutic targets. Incorporating data from additional ethnic populations and integrating environmental and lifestyle variables could enhance both predictive accuracy and clinical relevance. Longitudinal studies will also be crucial to evaluate test performance over time and its utility in monitoring disease progression or response to interventions [30].

Our results align with recent GWAS and multi-trait analyses of HF, which identified several loci associated with disease risk [31, 32]. Eight SNPs in our test overlapped with previously reported loci, supporting the validity of our approach. Non-overlapping SNPs likely reflect differences in study design, particularly our focus on SNP–SNP interactions, as well as variations in study populations and specific HF-related traits. These findings underscore the need for collaborative efforts to standardize genetic data and optimize HF risk prediction models.

Materials and Methods

Cohort descriptions

Genetic and clinical data were collected from three distinct cohorts with varied demographic and clinical characteristics:

Baker Biobank (BB) [33]: Individuals aged 18–69 were enrolled between 2000 and 2011, with detailed data collected on demographics, lifestyle, anthropometrics, medical history, medication use, and blood biomarkers. Biological samples included Guthrie cards, plasma, serum, buffy coat, and whole blood preserved in Tempus tubes for RNA extraction; some samples also contained extracted DNA and RNA. Data were linked to echocardiography, hospital admissions, pathology, and mortality records. HF status was determined from medical records at enrollment or ICD-coded linked datasets [34]. DNA from all HF cases and controls over 70 without HF was genotyped using the Illumina GSA array at the Australian Genome Research Facility, covering more than 700,000 SNPs.

Atherosclerosis Risk in Communities (ARIC) [35]: This community-based study followed 15,792 adults aged 45–64 from multiple ethnic groups to monitor MI, CHD, and mortality. Its primary goal was to identify factors contributing to subclinical atherosclerosis and CHD. Clinical and genetic data were accessed through NIH dbGAP.

Busselton Health Study (BHS) [36]: As one of the longest-running epidemiologic programs globally, BHS tracks residents of Busselton, Western Australia, with some participants enrolled since 1966. The study focuses on cardiovascular and respiratory risk factors through population-wide surveys, longitudinal follow-up, and collection of serum and DNA. Genetic and clinical data from a 1990s survey were made available by the study coordinators upon approval.

Genetic analyses

We applied three separate approaches to construct genetic signatures based on disease-related data structures and associated weights, utilizing both conventional association analysis and SNP-SNP interaction searches implemented in Plink 1.9 (http://pngu.mgh.harvard.edu/purcell/plink/, last accessed 14 February 2025) [37]. Initially, a case–control study was established, selecting controls from the BB cohort who were over 70 years old and free of HF, to reduce the likelihood of including younger individuals who might develop HF later, consistent with strategies used in other late-onset disease genetics studies (e.g., [38]). The dataset was randomly split into a discovery subset (500 participants) and a testing subset (2167 participants). The discovery subset was used to implement a multi-stage deep neural network (DNN) pipeline, efficiently utilizing extensive weight coverage while incorporating weights from previous cycles. To minimize demographic confounding, only individuals of European descent were included in the training set, with ancestry verified via PCA against known cohorts; these individuals represented 73% of the BB cohort. The analysis required substantial computational resources, performed on the

Pawsey Supercomputer Facility (www.pawsey.org.au, last accessed 16 February 2025). This pipeline identified 41 SNPs with significant associations and interactions, enabling the model to effectively discriminate cases from controls.

Significant SNP-SNP interactions were integrated into an artificial neural network, serving as weighted factors in the construction of a quantitative risk score (QRS), thereby enhancing predictive accuracy. Statistical models included data from all participants to comprehensively define risk, and each subject's QRS was calculated. Individuals were classified as high- or low-risk using QRS thresholds set at the 90th percentile of the control group.

A second-phase test focused exclusively on high-risk individuals identified in the first test, including all true positives and false positives. By excluding the SNPs from Test 1, 29 new SNPs were identified to define risk for this second-phase assessment. Both tests were developed using only whole-genome genotyped data. After assigning risk via Tests 1 and 2, validation was conducted in a reserved BB test set, followed by further confirmation in independent ARIC and BHS cohorts.

Genotype imputation

All SNPs used in the BB cohort were directly genotyped; however, variations in reference panels and genotyping methods across cohorts caused substantial missingness for several of the 41 HF-associated SNPs. Specifically, in the ARIC and BHS cohorts, eight SNPs from Test 1 and five SNPs from Test 2 exhibited low genotyping quality, resulting in missing data for roughly 10% of individuals. To remedy this, genotype imputation was performed using the TopMed Imputation Server with the Michigan Imputation Server pipeline employing Minimac4 [39], based on GRCh38/hg38 reference data. Older genome builds for the BHS cohort were lifted over to GRCh38/hg38. Accuracy of imputation was assessed by comparing the imputed SNPs to their original genotyped values, achieving over 99% concordance, a substantial improvement over the previous 68% accuracy obtained with the 1000 Genomes Panel in hg19.

Statistical analyses

Statistical significance for the analyses was assessed using Fisher's exact test via the online tool (https://www.langsrud.com/fisher.htm) (last accessed 16 February 2025), with a two-sided p-value below 10E-5 considered significant. Additional diagnostic measures, including the odds ratio (OR) and relative risk (RR), were computed using standard procedures (see ref. [40] for a review), with significance defined as OR > 2 or RR > 1.5. The OR quantifies the relationship between an exposure (e.g., genetic predisposition) and an outcome (e.g., heart failure), where OR > 1 indicates higher odds of the outcome with the exposure, and OR < 1 implies a protective effect. The ARIC and BHS cohorts, being community-based, were evaluated using RR, whereas the BB cohort, as a case–control study, employed OR for analysis.

Risk prediction performance was visualized using Kaplan–Meier (KM) survival curves, with the x-axis representing temporal variables such as age at diagnosis, and the y-axis showing the proportion of disease-free individuals. KM curves for different genetic risk groups were generated via Eureka Statistics (https://eurekastatistics.com) (last accessed 16 February 2025), which also provided 95% confidence intervals. Greater separation between curves indicates more pronounced differences between groups.

Balanced Accuracy (BAC), a metric particularly useful for imbalanced datasets in classification tasks, was considered meaningful if exceeding 0.6. BAC reflects model performance by averaging sensitivity (true positive rate) and specificity (true negative rate), offering a more holistic assessment of classification across both classes.

To compare the genetic tests with HF prediction based on conventional clinical variables, survival analyses were conducted on 1604 participants free of HF at baseline using multivariable Cox regression. In the BB cohort, 30 previously established HF-associated variables were included, and model discrimination was evaluated via C-statistics, paralleling prior analyses [41]. The model included all variables without selection, prioritizing maximum discriminatory ability over parsimony for cross-cohort applicability. Missing clinical data were handled through multiple imputation with chained equations (10 datasets) using the "mice" package in R v4.2.2, with pooled results calculated via Rubin's rules.

To address the risk of Type I errors due to multiple simultaneous hypotheses, Bonferroni correction was applied to control the family-wise error rate (FWER), thereby reducing the probability of false positives across all statistical tests.

Conclusion

Overall, this study advances understanding of HF genetics by developing a refined genetic test capable of predicting HF risk. This tool holds promise for enabling early intervention and tailoring personalized treatment strategies. Although further validation and integration with non-genetic factors are needed, the findings offer potential to improve outcomes in HF, a condition associated with substantial morbidity and mortality worldwide. Future work should focus on refining the test and ensuring its ethical clinical application, and we encourage collaborative efforts to extend its use to diverse populations.

Acknowledgments: We thank Alexis Espinosa and Sam Yates from the Pawsey Supercomputing Centre for their support and contributions to this work.

Conflict of interest: G.M. holds shares in Advanced Genetic Diagnostics, which developed the genetic risk test described here. X.G., B.B., L.D. and D.M.K. have no conflicts of interest to declare. The funders had no role in the design of this study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

Financial support: This research was funded by MTPConnect through their BioMedTech Horizons Program, grant number BMTH2_029, awarded to Advanced Genetic Diagnostics.

Ethics statement: The Australian Baker Heart and Diabetes Institute Biobank, the Busselton Health Study, and the Atherosclerosis Risk in Communities obtained their own approval from their respective institutional review boards or ethics committees. This study itself was conducted in accordance with the Declaration of Helsinki and approved by the following ethics committees: the Human Research Ethics Committee of the Department of Health WA (2015/39 (renamed RGS0000002849) on 19 August 2015 and the Human Research Ethics Committee of the University of Western Australia (RA_4_1_6937 on 19 February 2015 superseded by 2020/ET000284 on 20 November 2020.

Informed consent was obtained from all of the subjects involved in this study by the respective teams of the Australian Baker Heart and Diabetes Institute Biobank, the Busselton Health Study, and the Atherosclerosis Risk in Communities study.

References

1. Savarese G, Lund LH. Global public health burden of heart failure. Card Fail Rev. 2017;3(1):7–11.
2. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States: A policy statement from the American Heart Association. Circ Heart Fail. 2013;6(3):606–19.
3. Malik A, Brito D, Vaqar S, Chhabra L. Congestive Heart Failure. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430873/. Accessed 20 Dec 2025.
4. Schwinger RHG. Pathophysiology of heart failure. Cardiovasc Diagn Ther. 2021;11(1):263–76.
5. Hajouli S, Ludhwani D. Heart failure and ejection fraction. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK553115/. Accessed 20 Dec 2025.
6. Vasan RS. Diastolic heart failure. BMJ. 2003;327(7426):1181–2.
7. Hartupee J, Mann DL. Neurohormonal activation in heart failure with reduced ejection fraction. Nat Rev Cardiol. 2017;14(1):30–8.
8. Rodgers JL, Jones J, Bolleddu SI, Vanthenapalli S, Rodgers LE, Shah K, et al. Cardiovascular risks associated with gender and aging. J Cardiovasc Dev Dis. 2019;6(2):19.
9. Saheera S, Krishnamurthy P. Cardiovascular changes associated with hypertensive heart disease and aging. Cell Transplant. 2020;29:0963689720920830.
10. Khan SS, Ning H, Shah SJ, Yancy CW, Carnethon M, Berry JD, et al. 10-year risk equations for incident heart failure in the general population. J Am Coll Cardiol. 2019;73(19):2388–97.
11. Inamdar AA, Inamdar AC. Heart failure: Diagnosis, management and utilization. J Clin Med. 2016;5(6):62.
12. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. Eur Heart J. 2012;33(14):1787–847.
13. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. AHA/ACC/HFSA guideline for the management of heart failure. Circulation. 2022;145(18):e895–e1032.
14. Omerovic S, Jain A. Echocardiogram. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK558940/. Accessed 20 Dec 2025.
15. Pan Y, Li D, Ma J, Shan L, Wei M. NT-proBNP test with improved accuracy for the diagnosis of chronic heart failure. Medicine. 2017;96(37):e9181.
16. Long B, Koyfman A, Gottlieb M. Diagnosis of acute heart failure in the emergency department: An evidence-based review. West J Emerg Med. 2019;20(6):875–84.
17. Palmieri G, D'Ambrosio MF, Correale M, Brunetti ND, Santacroce R, Iacoviello M, et al. The role of genetics in the management of heart failure patients. Int J Mol Sci. 2023;24(22):15221.
18. Shrivastava A, Haase T, Zeller T, Schulte C. Biomarkers for heart failure prognosis: Proteins, genetic scores and non-coding RNAs. Front Cardiovasc Med. 2020;7:601364.
19. Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, National Academies of Sciences,

Engineering, and Medicine. Improving Diagnosis in Health Care. Balogh EP, Miller BT, Ball JR, editors. Washington, DC: National Academies Press; 2015. Available from: https://www.ncbi.nlm.nih.gov/books/NBK338593/. Accessed 20 Dec 2025.
20. Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield M, Devlin JJ, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy. Lancet. 2015;385(9984):2264–71.
21. Praz F, Beyersdorf F, Haugaa K, Prendergast B. Valvular heart disease: From mechanisms to management. Lancet. 2024;403(10385):P1576–P1589.
22. Masarone D, Limongelli G, Rubino M, Valente F, Vastarella R, Ammendola E, et al. Management of arrhythmias in heart failure. J Cardiovasc Dev Dis. 2017;4(1):3.
23. Morahan G, Huang D, Tait BD, Colman PG, Harrison LC. Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. Science. 1996;272(5267):1811–3.
24. Morahan G, Huang D, Ymer S, Cancilla MR, Stephen K, Dabadghao P, et al. Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. Nat Genet. 2001;27(3):218–21.
25. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations. Cell. 2019;179(3):589–603.
26. Presley CA, Min JY, Chipman J, Greevy RA, Grijalva CG, Griffin MR, et al. Validation of an algorithm to identify heart failure hospitalisations in patients with diabetes within the Veterans Health Administration. BMJ Open. 2018;8(3):e020455.
27. Mascia G, Olivotto I, Brugada J, Arbelo E, Di Donna P, Di Bona R, et al. Sport practice in hypertrophic cardiomyopathy: Running to stand still. Int J Cardiol. 2021;345:77–82.
28. Joseph J, Liu C, Hui Q, Aragam K, Wang Z, Charest B, et al. Genetic architecture of heart failure with preserved versus reduced ejection fraction. Nat Commun. 2022;13(1):7753.
29. Ram R, Mehta M, Nguyen QT, Larma I, Boehm BO, Pociot F, et al. Systematic evaluation of genes and genetic variants associated with type 1 diabetes susceptibility. J Immunol. 2016;196(7):3043–53.
30. Morales A, Hershberger R. Clinical application of genetic testing in heart failure. Curr Heart Fail Rep. 2017;14(6):543–53.
31. Levin MG, Tsao NL, Singhal P, Liu C, Vy HMT, Paranjpe I, et al. Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. Nat Commun. 2022;13(1):6914.
32. Patel KK, Venkatesan C, Abdelhalim H, Zeeshan S, Arima Y, Linna-Kuosmanen S, et al. Genomic approaches to identify and investigate genes associated with atrial fibrillation and heart failure susceptibility. Hum Genom. 2023;17(1):47.
33. Haregu TN, Nanayakkara S, Kingwell B, Jennings G, Dart A, Carrington M, et al. The Baker Biobank: Understanding cardiovascular outcomes. Heart Lung Circ. 2020;29(7):1071–7.
34. Dawson LP, Carrington MJ, Haregu T, Nanayakkara S, Jennings G, Dart A, et al. Ten-year risk equations for incident heart failure in established atherosclerotic cardiovascular disease populations. J Am Heart Assoc. 2024;13(12):e034254.
35. Wright J, Folson AR, Coresh J, Sharrett AR, Couper D, Wagenknecht LE, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. J Am Coll Cardiol. 2021;77(23):2939–59.
36. Adris N, Chua ACG, Knuiman MW, Divitini ML, Trinder D, Olynyk JK. A prospective cohort examination of haematological parameters in relation to cancer death and incidence: The Busselton Health Study. BMC Cancer. 2018;18:863.
37. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007;81(3):559–75.
38. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013;368(2):117–27.
39. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48(10):1284–7.
40. Andrade C. Understanding relative risk, odds ratio, and related terms: As simple as it can get. J Clin Psychiatry. 2015;76(8):e857–61.
41. Dawson LP, Carrington MJ, Haregu T, Nanayakkara S, Jennings G, Dart A, et al. Differences in predictors of incident heart failure according to atherosclerotic cardiovascular disease status. ESC Heart Fail. 2023;10(6):3398–409.