Penetrance and disease expression of (likely) pathogenic variants associated with inherited cardiomyopathies in the general population

Mimount Bourfiss, MD*; Marion van Vugt, MSc*; Abdulrahman I. Alasiri, MSc²; Bram Ruijsink, MD, PhD³; Jessica van Setten, MSc, PhD⁴; Amand F. Schmidt, MSc, PhD⁵; Dennis Dooijes, PhD⁶; Esther Puyol-Antón, PhD⁷; Birgitta K. Velthuis, MD, PhD⁸; J. Peter van Tintelen, MD, PhD⁹; Anneline S.J.M. te Riele, MD, PhD¹⁰; Annette F. Baas, MD, PhD¹¹; Folkert W. Asselbergs, MD, PhD¹²,³.⁶

*Shared first authorship

Affiliations
  a. Department of Cardiology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
  b. School of Biomedical Engineering and Imaging Sciences, King’s College London, London, United Kingdom
  c. Institute of Cardiovascular Science, London, Faculty of Population Health Sciences, University College London, United Kingdom
  d. Department of Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
  e. Department of Radiology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
  f. Netherlands Heart Institute, Utrecht, the Netherlands
  g. Health Data Research UK and Institute of Health Informatics, University College London, London, United Kingdom

Short title: ACM, DCM and HCM variants in the general population

Corresponding author: Mimount Bourfiss; Department of Medicine, Division of Cardiology; University Medical Center Utrecht; room E.02.561; Heidelberglaan 100; 3584 CX Utrecht, the Netherlands. Email address: m.bourfiss-2@umcutrecht.nl Telephone number: +31-88-77570240. Fax number: +31-88-7555660.

Journal Subjects Terms: Cardiomyopathy, Genetics, Heart failure, Hypertrophy, Magnetic Resonance imaging

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Background. (Likely) pathogenic variants associated with arrhythmogenic cardiomyopathy (ACM), dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM) are recommended to be reported as secondary findings in genome sequencing studies. This provides opportunities for early diagnosis, but also fuels uncertainty in variant carriers (G+), since disease penetrance is incomplete. We assessed the prevalence and disease expression of G+ in the general population.

Methods. We identified (likely) pathogenic variants associated with ACM, DCM and/or HCM in 200,643 UK Biobank individuals, who underwent whole exome sequencing. We calculated the prevalence of G+ and analysed the frequency of cardiomyopathy/heart failure diagnosis. In undiagnosed individuals, we analysed early signs of disease expression.

Results. We found a prevalence of 1:578, 1:251 and 1:149 for (likely) pathogenic variants associated with ACM, DCM and HCM respectively. Compared to controls, cardiovascular mortality was higher in DCM G+ (OR 1.67 [95% CI 1.04;2.59], p=0.030), but similar in ACM and HCM G+ (p≥0.100). More specifically, cardiomyopathy or heart failure diagnosis were more frequent in DCM G+ (OR 3.66 [95% CI 2.24;5.81], p=4.9×10⁻⁷) and HCM G+ (OR 3.03 [95% CI 1.98;4.56], p=5.8×10⁻⁷), but comparable in ACM G+ (p=0.172). In contrast, ACM G+ had more ventricular arrhythmias (p=3.3×10⁻⁴). In undiagnosed individuals, left ventricular ejection fraction was reduced in DCM G+ (p=0.009).

Conclusions. In the general population, (likely) pathogenic variants associated with ACM, DCM or HCM are not uncommon. Although G+ have increased mortality and morbidity, disease expression in these carriers from the general population remains low. Decisions on application of cascade screening and frequency of cardiological examination should be based on multiple factors, such as the variant and disease expression.
**Key Words:** whole exome sequencing, genetics, arrhythmogenic cardiomyopathy, dilated cardiomyopathy, hypertrophic cardiomyopathy

**Background**

The major inherited cardiomyopathies arrhythmogenic cardiomyopathy (ACM), dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM) are characterized by ventricular dysfunction and ventricular arrhythmias that can lead to progressive heart failure and sudden cardiac death. ACM – especially its right dominant subform – is mainly caused by pathogenic variants in desmosomal genes, whereas DCM and HCM are mainly caused by sarcomeric gene variants. These inherited cardiomyopathies are typically inherited in an autosomal dominant fashion with incomplete penetrance and variable expressivity. As such, phenotypic expression may vary greatly, even among individuals from the same family or those carrying the same pathogenic variant.

With the implementation of next-generation sequencing (NGS), genetic testing has become an important part of routine clinical care in the diagnosis of inherited cardiomyopathies. Technical advances and commercial availability of NGS, have led to less costly and more accessible genetic testing. The American College of Medical Genetics and Genomics (ACMG) has developed recommendations for the reporting of incidental or secondary findings unrelated to the test indication. In this framework, variants in genes associated with ACM, DCM and HCM are recommended to be reported as secondary findings from clinical exome and other genome sequencing tests. Although this offers the potential to prevent morbidity and mortality of heart failure and sudden cardiac death by
early treatment, it also fuels uncertainty in (likely) pathogenic variant carriers (G+) and their family members, since factors that influence disease penetrance in the general population are largely unknown. More knowledge about disease penetrance of these variants in an unselected population cohort is needed to determine screening protocols in these individuals.

In this study, we aimed to assess the G+ prevalence in the general population using a set of recently curated genes for ACM, DCM and HCM in two international databases. In order to do so, we leveraged data from the UK Biobank (UKB), a population-based cohort with whole exome sequencing (WES) data available of 200,643 individuals. Furthermore, we looked into the UKB-reported phenotypical characteristics of these G+ and assessed the occurrence of early signs of disease in undiagnosed G+.

**Methods**

*Genetic variants in the study population*

We identified carriers of a (likely) pathogenic variant associated with ACM, DCM or HCM in individuals from the UKB who underwent WES (n=200,643 at time of analysis). For each inherited cardiomyopathy we selected curated genes classified to have definite, strong or moderate evidence of pathogenicity. For ACM we included DES, DSC2, DSG2, DSP, JUP, PKP2, PLN and TMEM43; for DCM we included ACTC1, ACTN2, BAG3, DES, DSP, FLNC, JPH2, LMNA, MYH7, NEXN, PLN, RBM20, SCNSA, TNNC1, TNNI3, TNNT2, TPM1, TTN and VCL; and for HCM we included ACTC1, CSRP3, JPH2, MYBPC3, MYH7, MYL2, MYL3, TNNC1, TNNI3, TNNT2 and TPM1 (Figure 1 and Table S1). Some variants are associated with two cardiomyopathies (Figure 1 and Table S1); individuals carrying these variants were included in the G+ groups of both cardiomyopathies. Next, likely pathogenic and pathogenic variants
in these genes were identified using the ClinVar NCBI-NIH database⁸ and the Dutch Society for Clinical Genetic Laboratory Diagnostics (Vereniging Klinische Genetische Laboratoriumdiagnostiek, VKGL) database⁹. An elaborate overview of the Clinvar and VKGL search criteria is given in Figure 2. In short, ClinVar was queried using the disease name(s) and filtered for (likely) pathogenic variants in the curated genes. For pathogenic variants mentioned in the VKGL database, which does not specify disease associations, association with one of the cardiomyopathies was confirmed in ClinVar. The minor allele frequency (MAF) cut-off was defined at 0.001. Pathogenic variants were classified as missense or loss of function (LoF), with LoF being defined as frameshift, stop gain, start lost and canonical splice site variants.

We matched G+ individuals in a 1:4 ratio to UKB individuals without a (likely) pathogenic variant associated with one of the cardiomyopathies (G-). Matching of this G-control group was based on age, sex, ethnicity and presence of cardiac magnetic resonance imaging (CMR) measurements. Controls are referred to as G- throughout this study.

Data extraction UKB

Disease definitions

An elaborate overview of the disease definitions used in this study is available in Table S2. In short, individuals were defined to be phenotype positive (P+) if they had an ICD-10 or self-reporting code for cardiomyopathy, DCM, HCM or heart failure, without a diagnosis of chronic ischemic heart disease. No ICD-10 or self-reporting code was available for ACM in the UKB.

CMR and ECG data analysis
We investigated disease expression on cardiac magnetic resonance imaging (CMR) and electrocardiography (ECG) in a subset of G+P- individuals. The full CMR protocol of the UKB has been described in detail\textsuperscript{11}. In short, all CMR examinations were performed on a 1.5 Tesla scanner (Magnetom Aera, Syngo Platform VD13A, Siemens Healthcare, Erlangen, Germany). We used a previously developed and validated deep-learning methodology (AI-CMR\textsuperscript{QC}) to extract left (LV) and right ventricular (RV) CMR measurements\textsuperscript{12}. In short, cine images of short-axis and 2- and 4-chamber long-axis views were used to automatically calculate LV and RV functional measures (ejection fraction [EF], stroke volume [SV]) and structural measures (end diastolic volume [EDV], end-systolic volume [ESV], LV end diastolic mass [EDM], LV mass to EDV ratio [LVMVR] and LV maximal and regional [16 segments model according the American Heart Association\textsuperscript{13}] wall thickness).

The electrocardiography (ECG) variables P duration, P axis, PQ interval, QRS duration, R axis, QTc interval and T axis were extracted from the UKB for G+P- individuals.

\textit{Statistical analysis}

Statistical analysis was performed using R version 4.0.2\textsuperscript{14}. Continuous values are presented as median [interquartile range] and for comparisons of two groups, Mann-Whitney-U test was used. Categorical data was displayed as absolute frequency (n) and percentages (%) and Fisher’s exact test was used to test for differences. The strength of the association between cardiac outcomes and G+ ACM, DCM and HCM was calculated by the odds ratio with 95\% confidence intervals. A p-value of less than 0.05 was considered significant. See the \textbf{Supplemental Information} for a more elaborate overview of the definitions and analysis performed.
**Results**

*Population characteristics*

We identified 2,207/200,643 unique G+ individuals classified as 1) ACM G+ (n=347, 54% female, median age of 57 [50-64] years); 2) DCM G+ (n=800, 56% female, median age of 58 [51-64] years) and; 3) HCM G+ (n=1,346, 54% female, median age of 56 [49-63] years). The matched control group consisted of 9,972 individuals (55% females, median age of 57 [49-63] years). **Table 1** and **Table S3** show the baseline characteristics of the included individuals. The majority of G+ were of white ethnicity (ACM 90%, DCM 96%, HCM 75%), followed by asian (ACM 3%, DCM 1%, HCM 19%) and black ethnicity (ACM 2%, DCM 2% and HCM 2%).

*Genotypic characteristics of (likely) pathogenic variant carriers*

**Prevalence of (likely) pathogenic variants in the general population**

We found a prevalence of 1:578 [1:521; 1:644] ACM G+ and identified 75 variants out of the 593 (13%) (likely) pathogenic variants described in Clinvar and VKGL: 13 missense and 62 LoF (**Table S4**). As shown in **Figure 3**, most ACM G+ harbored a pathogenic variant in *PKP2* (n=185, 53%), followed by *DSP* (n=49, 14%), *DSC2* (n=42, 12%), *DSG2* (n=31, 9%), *JUP* (n=24, 7%), *DES* (n=15, 4%), and *PLN* (n=1, 0.3%).

We found a prevalence of 1:251 [1:234; 1:269] DCM G+ and identified 216 out of 3,460 (6%) (likely) pathogenic variants described in Clinvar and VKGL: 80 missense and 136 LoF (**Table S4**). Variants in *TTN* (n=272, 34%) and *MYH7* (n=158, 20%) were most prevalent among DCM G+, followed by *SCN5A* (n=59, 7%), *FLNC* (n=56, 7%), *DSP* (n=49, 6%), *DES* (n=49, 6%), *LMNA* (n=42, 5%), *TNNT3* (n=35, 4%) and *TNNT2* (n=32, 4%). Eight more genes
with a frequency of less than 3% were identified: **BAG3, PLN, TNNC1, ACTN2, RBM20, NEXN, TPM1,** and **ACTC1** (Figure 3).

We found a prevalence of 1:149 [1:141; 1:157] HCM G+ and identified 131 out of 1,512 (9%) (likely) pathogenic variants from Clinvar and VKGL: 98 missense and 23 LoF (Table S4). Most individuals carried a (likely) pathogenic variant in **MYBPC3** (n=723, 54%), followed by **TNNT2** (n=274, 20%), **MYH7** (n=232, 17%), and **TNNI3** (n=50, 4%) (Figure 3). A frequency of less than 3% was found in **CSRP3, MYL2, TNNC1, JPH2, TPM1, ACTC1,** and **MYL3** (Figure 3). Interestingly, a variant in **TNNT2** (c.862 C>T p.Arg288 Cys) affected 242 individuals (18%). Four of these carriers were diagnosed with heart failure, of whom one also with HCM. All of them also suffered from chronic ischemic heart disease. Furthermore, a variant in **MYBPC3** (c.3628-41_3628-17del) was mainly seen in individuals with an Asian ethnicity (n=237, 18% of the total HCM G+). Four were diagnosed with heart failure, of whom two also had coronary artery disease and one was diagnosed with DCM, however none were diagnosed with HCM. When excluding these two variants, we found a HCM G+ prevalence of 1:250 [1:234; 1:269]. **MYBPC3** remained the most prevalent gene (52%), whereas the **TNNT2** frequency decreased to 4%.

The prevalence of G+ per gene for all cardiomyopathies is depicted in Table S5.

**Overlapping variants and individuals**

Some (likely) pathogenic variants were identified in multiple cardiomyopathy subtypes. First, 26 (likely) pathogenic variants were described in both ACM and DCM, affecting 53 individuals. Most of these variants (n=20/26 variants, 77%) were found in **DSP** (n=37 individuals, 70%), of whom one individual (3%) had heart failure and one (3%) was diagnosed with cardiomyopathy. Five variants out of 26 (19%) were found in **DES** (n=15...
individuals, 28%) of whom two individuals (13%) had heart failure, and one was diagnosed with both DCM and HCM. One variant out of 26 (4%) in PLN (c.26G>A; p.Arg9His, NM_002667.5) was found in one individual (2%) who was not diagnosed with either a cardiomyopathy or heart failure.

Second, 52 (likely) pathogenic variants were described in DCM and HCM, affecting 232 individuals. Most of these variants (n=33/52 variants, 63%) were found in MYH7 (n=158 individuals, 68%), followed by 10 variants (19%) in TNNT2 (n=29 individuals, 13%), 6 variants (12%) in TNNI3 (n=35 individuals, 15%), and 1 (2%) variant in TNNC1, ACTC1 as well as TPM1. In this group of 232 individuals, 9 (4%) individuals had a cardiomyopathy or heart failure diagnosis, of whom 5 were diagnosed with HCM and none with DCM.

Furthermore, three individuals carried two pathogenic variants. Individual 1 carried variants in MYBPC3 (c.3628-41_3628-17del, NM_000256.3) and TNNT2 (c.460C>T; p.Arg154Trp, NM_001276345.2) and was diagnosed with heart failure, with underlying chronic ischemic heart disease. Individual 2 carried variants in FLNC (c.7450G>A; p.Gly2484Ser, NM_001458.5) and PKP2 (c.1867G>T; p.Glu623Ter, NM_001005242.3) and was therefore included in both the ACM as well as the DCM G+ group. Individual 3 carried variants in MYBPC3 (c.1504C>T; p.Arg502Trp, NM_000256.3) and MYH7 (c.5655G>A; p.Ala1885=, NM_000257.4). Individuals 2 and 3 were not diagnosed with either cardiomyopathy or heart failure.

**Phenotypic characteristics of (likely) pathogenic variant carriers**

**Cardiovascular risk factors**

Hypertension, BMI, and level of activity in metabolic equivalent of task minutes (MET) per week were comparable between G- and G+ for all cardiomyopathies (p≥0.055; Table 1 and
Table S6). Diabetes was more prevalent in G+ HCM (9.2% (G-) vs 11.4% (G+), p=0.008), while smoking was more prevalent in DCM G+ (41.4% vs 46.4%, p=0.007) (Table S7).

Cardiovascular disease

As seen in Figure 4 and Table S7, compared to G-, cardiomyopathy/heart failure without previous ischemic heart disease (P+) was more often diagnosed in DCM G+ (OR 3.66 [95% CI 2.24;5.81], p=4.9•10^{-7}) and HCM G+ (OR 3.03 [95% CI 1.98;4.56], p=5.8•10^{-7}). Among DCM G+, 25 individuals (3.1%, genes: 8 MYH7, 8 TTN, 2 BAG3, 2 DSP, 2 FLNC, 1 DES, 1 SCN5A and 1 TNNT2) were P+. Within HCM G+, 32 individuals (4.0%, genes: 20 MYBPC3, 10 MYH7, 1 TNNI3 and 1 TNNT2) were P+. For ACM G+, 4 individuals (1.2%, genes: 2 DSP, 1 DES and 1 PKP2) were P+, being comparable to G- controls (87 subjects, 0.8%).

Ventricular arrhythmias occurred more often in G+ compared to G-, reaching statistical significance for ACM (OR 6.20 [95% CI 2.30;14.38], p=3.3•10^{-4}) and DCM (OR 4.97 [95% CI 2.39;9.75], p=1.9•10^{-6}). Atrial arrhythmias were more prevalent among DCM G+ (OR 2.27 [95% CI 1.52;3.31], p=8.2•10^{-5}). Finally, all-cause mortality (OR 1.39 [95% CI 1.02;1.85], p=0.032) and death due to a cardiovascular cause were more prevalent in DCM G+ (OR 1.67 [95% CI 1.04;2.59], p=0.030) but did not reach statistical significance for ACM G+ and HCM G+ (p>0.100). Figure 5 depicts the overlap in cardiomyopathy, heart failure, ventricular arrhythmia and ischemic heart disease diagnosis. See Figure S1 for the forest plots when excluding the more prevalent TNNT2 and MYBPC3 variants in HCM G+ individuals.

Deep phenotyping of undiagnosed pathogenic variant carriers
Next, we set out to study early signs of disease in G+ without a cardiomyopathy/heart failure diagnosis (P-) using registered ICD-10 codes, self-reported cardiac symptoms, and abnormal ECG and CMR values.

Diagnosis and symptoms
Ventricular arrhythmias were more prevalent in ACM G+P- (OR 5.85 [95% CI 1.98;14.40], p=0.001) and DCM G+P- (OR 3.43 [95% CI 1.35;7.68], p=0.005) but not in HCM G+P- (OR 1.01 [95% CI 0.26;2.86], p=1.000) compared to G-P- controls. Also, atrial arrhythmias (OR 2.12 [95% CI 1.36;3.19], p=7.9•10^{-4}) were more frequent in DCM G+P- compared to G-P- controls. Finally, angina pectoris occurred more often in HCM G+P- (OR 1.38 [95% CI 1.01;1.85], p=0.038), but not in ACM G+P- and DCM G+P- (p≥0.117; Table S6).

Electrocardiography
In total, 231 out of 2,207 G+P- and 1,058 out of 9,885 G-P- had various ECG variables available. None of these ECG variables differed significantly between all undiagnosed G+ and control individuals (Table S7).

Cardiac magnetic resonance imaging
CMR data was available in 225 G+P- of the 2,207 unique G+P- individuals: n=33 for ACM G+P-, n=87 for DCM G+P- and n=130 for HCM G+P-) and 986 of the 9,885 G-P- controls.

As shown in Figure 6 and Table S7, all RV (p≥0.546) and LV (p≥0.052) functional and structural parameters in ACM G+P- were comparable to G-P- controls. Three ACM G+P-individuals had an RVEDV corrected for body surface area (RVEDVi) between 100-110 mL/m² for males or 90-100 mL/m² for females, meeting the minor CMR task force criteria (TFC) if
wall motion abnormalities were present, and two ACM G+P- individuals had an RVEDVi larger than 110 mL/m² for males or 100 mL/m² for females, meeting the major CMR TFC.

In addition, ACM G+P- had reduced inferior and posterolateral wall thickness compared to controls (p≤0.035).

Overall, DCM G+P- and G-P- controls had comparable RV functional and structural measures (p≥0.048). However, DCM G+P- had lower LVEF [57.3% [52.6, 62.8] vs. 59.5% [55.3, 63.5] vs, p=0.009) and less negative LV peak longitudinal strain (-22.3% [-24.6, -19.86] vs. -23.3% [-26.0, -21.4], p=0.009). Although LVEDVi was not significantly increased in DCM G+P-, the LVEDV/RVEDV ratio (0.9 [0.9, 1.0] vs 1.0 [0.9, 1.1], p=8.2•10⁻⁴) and LVESVi (30.0 ml/m² [25.1, 35.7] vs 31.7 ml/m² [26.2, 39.8], p=0.032) were increased. Six individuals had an LVEF below 45%, but none of the individuals met the Henry criteria for DCM (LVEF below 45% and LVEDVi two times the normal SD).

For HCM G+P-, most RV and LV functional and dimension parameters were comparable to G-P- controls (p≥0.051). Only RVEF was higher than controls (58.4% [54.2, 62.7] vs 59.6% [54.8, 64.0], p=0.025). Importantly, wall thickness was not significantly different between HCM G+P- without a cardiomyopathy/heart failure diagnosis and G-P- (p≥0.160). None of the G+P- individuals met HCM criteria of ≥15 mm wall thickness, but two individuals met the criteria for limited hypertrophy (13-15 mm) in the presence of a positive genetic test.

**Exclusion of the more prevalent TNNT2 and MYBPC3 variants**

When excluding the more prevalent TNNT2 and MYBPC3 variants in HCM G+P- individuals, the occurrence of ventricular arrhythmias (OR 1.72 [95% CI 0.44;4.89], p=0.306) and atrial arrhythmias (OR 1.43 [95% CI 0.84;2.32], p=0.156) was comparable to G-P- controls.
However, the maximum wall thickness (8.47mm [7.59, 9.94] vs. 8.09mm [7.24, 9.01], p=0.008) and basal anterior wall thickness (7.93mm [6.97, 9.11] vs. 7.65mm [6.81, 8.49], p =0.029) were significantly increased in HCM G+P- compared to controls (Table S7). Two individuals had a maximum wall thickness between 13-15mm.

Discussion
In this study we leveraged the largest European population database including WES and phenotyping data to evaluate the prevalence and penetrance of previously reported (likely) pathogenic variants associated with ACM, DCM and HCM. Our study has several interesting findings. First, we found a prevalence of 1:578, 1:251, and 1:149 for (likely) pathogenic variants previously associated with ACM, DCM and HCM respectively. Second, 1.2% of ACM G+, 3.1% of DCM G+ and 2.6% of HCM G+ were diagnosed with a cardiomyopathy or heart failure without previous chronic ischemic heart disease. Finally, 3.2% of the undiagnosed ACM G+, 1.8% of the undiagnosed DCM G+, 0.5% of the undiagnosed HCM G+ reported ventricular arrhythmias or had CMR abnormalities. These results confirm the low disease penetrance in G+ among the general population.

Prevalence of pathogenic variant carriers in the general population
Since the major contributor to inherited cardiomyopathies are rare genetic variants, a large dataset is needed to accurately identify the population prevalence of these variants.

Although prevalence of pathogenic variants in populations has been the focus of several previous studies\(^4\)\(^\text{18-20}\), they were mostly limited by the number of included individuals. At time of analysis, we had access to an unprecedented high number of 200,643 individuals.
The reported prevalence of ACM G+ in the general population varied widely between different previous studies, ranging between 1:143 to 1:1,706\textsuperscript{18,21}. This variability is likely to be explained by heterogeneity in study populations and definitions of variant pathogenicity. For example, many previous studies did not include all eight curated genes with strong or moderate disease-gene association but also marked other genes (e.g. \textit{TGFB3}) with only limited evidence as associated with ACM\textsuperscript{19,20}. In addition, we included both missense and LoF variants whereas prior studies restricted themselves to LoF variants only.

For DCM, little is known about the prevalence of DCM causing variants in the general population. Studies focusing on truncating \textit{TTN} variants in the general population have found a prevalence ranging between 1:33 and 1:526 depending on the definition of pathogenicity\textsuperscript{22,23}. This could in part be due to the changed locations of truncating variants over the years. When solely focussing on (likely) pathogenic \textit{TTN} variants, we found a prevalence of only 1:735. This differs from the previous studies, probably because not all \textit{TTN} variants are reported as (likely) pathogenic in ClinVar and VKGL. Including all curated DCM-associated genes, we report a prevalence of 1:251.

For HCM, we found a prevalence ranging between 1:250 and 1:149 individuals carrying a (likely) pathogenic variant, which approaches previous estimates of 1:207\textsuperscript{24}. In a recent study, including the UKB population, a prevalence of 1:407 was reported\textsuperscript{25}. They included 8 sarcomere-encoding genes described to be associated with HCM (\textit{ACTC1, MYBPC3, MYH7, MYL2, MYL3, TNNI3, TNNT2 and TPM1}) and variants that were described as (likely) pathogenic in Clinvar or annotated as (likely) pathogenic according to ACMG criteria and filtered variants for an allele frequency of 0.00004. We included additional genes (\textit{CSRP3, JPH2 and TNNC1}) and (likely) pathogenic variants from the VKGL database and filtered for a minor allele frequency of 0.001.
Disease expression of (likely) pathogenic variants in the general population

Most information on disease penetrance in ACM, DCM or HCM G+ is based on observations in G+ relatives. Previous studies have shown that 37\% of ACM G+ relatives and up to 50\% of HCM G+ relatives with sarcomeric variants show disease expression during follow-up. Our findings suggest that disease penetrance in the general population is much lower. We found that 1.2\% of ACM G+, 3.1\% of DCM G+ and 2.6\% of HCM G+ in the UKB were diagnosed with a cardiomyopathy or heart failure, in the absence of chronic ischemic heart disease. Our additional analysis of ventricular function and ECG in undiagnosed G+ subjects also suggests a low disease penetrance. We found significantly worse LVEF and strain parameters in DCM G+P- compared to controls, however none met the diagnostic Henry criteria (LVEF below 45\% and LVEDVi two times the normal SD)\textsuperscript{16}. Although CMR data was only available in a subgroup of undiagnosed G+ patients, these findings make it unlikely that the low penetrance found in our study arises from missed diagnoses or covert disease in the G+ cohort. In G+P- ACM and DCM we found a low but significantly higher prevalence of ventricular arrhythmias compared to controls (1.7\% vs. 0.3\% (OR 5.85 [95\% CI 1.98;14.40]) and 1.0\% vs. 0.3\% (OR 3.43 [95\% CI 1.35;7.68]) respectively). In ACM, electrical abnormalities are known to precede structural abnormalities\textsuperscript{27}. Therefore, these findings may suggest early disease penetrance in a small subset of undiagnosed G+ individuals. The discrepancy between the high disease prevalence found in G+ family members and the low prevalence in the G+ general population points towards the interaction of possible other (unidentified) genetic and environmental factors leading to this variation in disease penetrance between relatives and general population.
Interestingly, the South Asian MYBPC3 and the TNNT2 variant, showed a relatively high prevalence in our cohort. In total, 19% of HCM G+ was Asian and mainly carried the c.3628-41_3628-17del variant in the MYBPC3 gene. Although this variant is indicated as likely pathogenic in Clinvar, a previous study suggests that this variant may be reclassified as benign. In our study, none of these mutation carriers were diagnosed with HCM. Four were diagnosed with heart failure of whom one was diagnosed with DCM. This suggests that this variant is associated with heart failure in the setting of multiple forms of cardiomyopathy, and not simply HCM. Second, the c.862C>T p.Arg288Cys variant in TNNT2 was previously found in HCM individuals, and is suggested as a non-primary or only cause of HCM. Interestingly, when excluding these variants from our G+P- population, a significantly higher wall thickness is measured compared to control. These two examples emphasize that when (likely) pathogenic variants are identified as a secondary finding, other factors such as the specific variant and the family history are crucial to inform screening frequency.

Interestingly, self-reported health-related quality of life and psychological well-being of 89 asymptomatic HCM knowing to be G+ were previously evaluated in a Dutch cohort and found to be at least similar to the general population, which suggests that reporting incidental findings will not harm psychological well-being of G+. However, frequent cardiological examination of G+ and family members turning out to be carriers after cascade screening will put a burden on health care and societal costs. Since the overall disease penetration in the general population is quite low, it should be considered to apply a modest frequency of cardiovascular screening in individuals identified after secondary finding and their relatives. It may even be considered to only perform cascade screening.
after disease expression is identified, but in all these decisions, family history, gene and variant type should be considered.

**Limitations**

Disease expression in inherited cardiomyopathies typically occurs before the age of 40 years. Since the mean age of our cohort was around 60 years, we would expect disease expression to have occurred in most of the G+. However, further follow-up of G+P- individuals is necessary to evaluate if they will develop a cardiomyopathy later in life. On the other hand, since disease expression is present at a younger age, these individuals will be less likely to participate in such large-scale biobank study.

Despite recent efforts to harmonize knowledge on genes associated with inherited cardiomyopathies, and guidelines for variant classification, the adjudication of the clinical significance of single variants can still differ between diagnostic laboratories which has led to interpretation differences. Furthermore, not all (likely) pathogenic variants are reported in these databases, especially family-specific variants and pathogenic variants in non-Caucasian populations are underreported.

**Conclusion**

In a cohort of 200,643 individuals with WES and phenotype data we identified a prevalence of pathogenic variants associated with ACM, DCM and HCM of 1:578, 1:251 and 1:149 respectively. Among the identified G+ individuals, cardiomyopathy, heart failure and ventricular arrhythmias were more common compared to G-. However, overall disease penetrance was low (1.2-3.1%). Therefore, decisions on application of cascade screening
and frequency of cardiological examination should be based on multiple factors, such as variant and gene type, family history and disease expression.

Acknowledgements

None

Funding sources

The work was financially supported by the Netherlands Cardiovascular Research Initiative, an initiative supported by the Dutch Heart Foundation (CardioVascular Onderzoek Nederland (CVON) projects: DOUBLE-DOSE 2020B005, PREDICT2 2018-30, eDETECT 2015-12), PREDICT Young Talent Program (AtR), 2015T058 (AtR), 2015T041 (AB) and 2019T045 (MvV and JvS). In addition, MB is supported by the Alexandre Suerman Stipend of the UMC Utrecht (2017), AtR by the UMC Utrecht Fellowship Clinical Research Talent and FA by the UCL Hospitals NIHR Biomedical Research Center.

Disclosures

None

References


14. R Core team. R: A Language and Environment for Statistical Computing [Internet]. 2020;Available from: https://www.r-project.org/


19. Haggerty CM, James CA, Calkins H, Tichnell C, Leader JB, Hartzel DN, Nevius CD,


## Tables

### Table 1: Baseline characteristics of mutation carriers and controls

<table>
<thead>
<tr>
<th></th>
<th>ACM n=347</th>
<th>DCM n=800</th>
<th>HCM n=1,346</th>
<th>Controls n=9,972</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (%)</td>
<td>187 (54)</td>
<td>450 (56)</td>
<td>720 (54)</td>
<td>5,436 (55)</td>
</tr>
<tr>
<td>Age, years</td>
<td>57 [50-64]</td>
<td>58 [51-64]</td>
<td>56 [49-63]</td>
<td>57 [49-63]</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>311 (90)</td>
<td>760 (96)</td>
<td>1,001 (75)</td>
<td>8,288 (84)</td>
</tr>
<tr>
<td>Asian</td>
<td>10 (3)</td>
<td>8 (1)</td>
<td>251 (19)</td>
<td>1,076 (11)</td>
</tr>
<tr>
<td>Black</td>
<td>7 (2)</td>
<td>12 (2)</td>
<td>22 (2)</td>
<td>164 (2)</td>
</tr>
<tr>
<td>Other</td>
<td>17 (5)</td>
<td>15 (2)</td>
<td>55 (4)</td>
<td>348 (4)</td>
</tr>
<tr>
<td><strong>Cardiovascular risk factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>35 (10)</td>
<td>62 (8)</td>
<td>154 (11)*</td>
<td>914 (9)</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>116 (33)</td>
<td>287 (36)</td>
<td>475 (35)</td>
<td>3,420 (34)</td>
</tr>
<tr>
<td>Hypercholesterolemia (%)</td>
<td>86 (25)</td>
<td>211 (26)</td>
<td>369 (27)*</td>
<td>2,416 (24)</td>
</tr>
<tr>
<td>Ever Smoked (%)</td>
<td>161 (46)</td>
<td>371 (46)*</td>
<td>543 (40)</td>
<td>4,132 (41)</td>
</tr>
<tr>
<td>MET minutes per week, ml/kg/min</td>
<td>2,001 [923-3,551]</td>
<td>1,695 [784-3,536]</td>
<td>1,762 [848-3,426]</td>
<td>1,773 [810-3,453]</td>
</tr>
<tr>
<td>Family heart disease (%)</td>
<td>179 (52)*</td>
<td>380 (48)</td>
<td>623 (46)</td>
<td>4,458 (45)</td>
</tr>
<tr>
<td><strong>Cardiac disease/outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiomyopathy (%)</td>
<td>3 (0.9)</td>
<td>22 (3)**</td>
<td>27 (2)**</td>
<td>37 (0.4)</td>
</tr>
<tr>
<td>DCM (%)</td>
<td>2 (0.6)</td>
<td>9 (1)**</td>
<td>1 (0.1)</td>
<td>14 (0.1)</td>
</tr>
<tr>
<td>HCM (%)</td>
<td>1 (0.3)</td>
<td>7 (1)**</td>
<td>20 (2)**</td>
<td>8 (0.1)</td>
</tr>
<tr>
<td>Heart failure (%)</td>
<td>9 (3)</td>
<td>36 (5)**</td>
<td>33 (3)</td>
<td>182 (2)</td>
</tr>
<tr>
<td>Ventricular arrhythmias (%)</td>
<td>7 (2)**</td>
<td>13 (2)**</td>
<td>8 (1)</td>
<td>33 (0.3)</td>
</tr>
<tr>
<td>Condition</td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
<td>Group 4</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Atrial arrhythmias (%)</td>
<td>7 (2)</td>
<td>34 (4)**</td>
<td>32 (2)</td>
<td>191 (2)</td>
</tr>
<tr>
<td>Chronic ischemic heart disease (%)</td>
<td>35 (10)</td>
<td>73 (9)</td>
<td>93 (7)</td>
<td>725 (7)</td>
</tr>
<tr>
<td>Acute myocardial infarction (%)</td>
<td>15 (4)</td>
<td>27 (3)</td>
<td>36 (3)</td>
<td>298 (3)</td>
</tr>
<tr>
<td>Cardiac arrest (%)</td>
<td>0 (0)</td>
<td>6 (1)</td>
<td>5 (0.4)</td>
<td>34 (0.3)</td>
</tr>
<tr>
<td>Cardiovascular death (%)</td>
<td>11 (3)</td>
<td>24 (3)*</td>
<td>18 (1)</td>
<td>181 (2)</td>
</tr>
<tr>
<td>All cause mortality (%)</td>
<td>19 (6)</td>
<td>56 (7)*</td>
<td>62 (5)</td>
<td>513 (5)</td>
</tr>
</tbody>
</table>

**Cardiac symptoms**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac problem</td>
<td>3 (1)</td>
<td>3 (0.4)</td>
<td>5 (0.4)</td>
<td>41 (0.4)</td>
</tr>
<tr>
<td>Angina pectoris</td>
<td>16 (5)</td>
<td>30 (4)</td>
<td>56 (4)*</td>
<td>312 (3)</td>
</tr>
</tbody>
</table>

Number (percentages) are given or median [IQR]. * p=0.001-0.05 and ** p <0.001 difference compared to the control group. A more elaborate overview of the baseline characteristics are given in Supplementary Table 3.

Abbreviations: ACM= arrhythmogenic cardiomyopathy; DCM= dilated cardiomyopathy; HCM= hypertrophic cardiomyopathy; BMI= body mass index; MET= metabolic equivalent of task.
Figures

**Figure 1: Included curated genes per cardiomyopathy**

The Venn diagram of curated genes included in this study shows the overlap in genes per cardiomyopathy. Unless otherwise indicated, pathogenicity of genes are classified as definitive. If a superscript S or M is given, genes are classified as having a strong or moderate pathogenicity respectively. In the overlapping circles, yellow, black and red colors refer to ACM, DCM, and HCM respectively. *Supplementary Table 1 gives an overview of the included genes and pathogenicity classification per gene and abbreviation per gene.*

*Abbreviations: ACM= arrhythmogenic cardiomyopathy; DCM= dilated cardiomyopathy; HCM= hypertrophic cardiomyopathy.*

**Figure 2: Flowchart inclusion of variants**

Flowchart depicting the inclusion of (likely) pathogenic variants associated with arrhythmogenic cardiomyopathy, dilated cardiomyopathy and hypertrophic cardiomyopathy from the Clinvar\textsuperscript{11} and VKGL database\textsuperscript{12}.

*Abbreviations: ACM= arrhythmogenic cardiomyopathy; DCM= dilated cardiomyopathy; HCM= hypertrophic cardiomyopathy; VKGL= Vereniging Klinische Genetische Laboratoriumdiagnostiek.*

**Figure 3: Distribution of genes per cardiomyopathy**

Piecharts with the distribution of curated genes for A) arrhythmogenic cardiomyopathy (ACM); B) dilated cardiomyopathy (DCM); C) hypertrophic cardiomyopathy (HCM).

*Abbreviations of the different genes are given in Supplementary Table 3*  
*Abbreviations: G+= pathogenic variant carrier.*

**Figure 4: Forest plot cardiac outcomes stratified per inherited cardiomyopathy**

Odds ratios and 95% confidence interval are given for the associations between cardiac outcomes and ACM, DCM, or HCM pathogenic variant carriers.

*Abbreviations: ACM= arrhythmogenic cardiomyopathy; DCM= dilated cardiomyopathy; G+= pathogenic variant carrier; HCM= hypertrophic cardiomyopathy.*

**Figure 5: Overlap cardiac diagnoses per inherited cardiomyopathy**

The Venn diagram of the overlap between cardiomyopathy, heart failure, ventricular arrhythmia and chronic ischemic heart diagnoses in G+ individuals. De numbers in the diagram are the number of individuals.

*Abbreviations: ACM= arrhythmogenic cardiomyopathy; CM= cardiomyopathy; DCM= dilated cardiomyopathy; G+= pathogenic variant carrier; HCM= hypertrophic cardiomyopathy; HF= heart failure; VA= ventricular arrhythmias.*
**Figure 6: CMR parameters stratified per inherited cardiomyopathy**

Boxplots of CMR parameters to show the distribution of CMR parameters stratified in controls and individuals with a pathogenic variant associated with ACM, DCM, or HCM. *Abbreviations: ACM= arrhythmogenic cardiomyopathy; DCM= dilated cardiomyopathy; EDVi= body surface area corrected end-diastolic volume; EF= ejection fraction; G+= pathogenic variant carrier; HCM= hypertrophic cardiomyopathy; LV= left ventricular; RV= right ventricular.*
Likely pathogenic / pathogenic variants ClinVar
- ACM: 483 variants
- DCM: 3,268 variants
- HCM: 2,710 variants

Likely pathogenic / pathogenic variants VKGL in curated genes
- ACM: 440 variants
- DCM: 2,562 variants
- HCM: 1,000 variants

Variants present in UK biobank participants
- ACM: 90 variants
- DCM: 277 variants
- HCM: 148 variants

Removed VKGL variants with ambiguous annotation
- ACM: 75 variants
- DCM: 216 variants
- HCM: 131 variants

Variants and carriers in UK biobank
- ACM: 153 variants / 347 carriers
- DCM: 898 variants / 800 carriers
- HCM: 512 variants / 1,346 carriers