Supplemental Online Content

Huggins GS, Kinnamon DD, Haas GJ, et al; DCM Precision Medicine Study of the DCM Consortium. Prevalence and cumulative risk of familial idiopathic dilated cardiomyopathy. *JAMA*. doi:10.1001/jama.2021.24674

This supplemental material has been provided by the authors to give readers additional information about their work.

© 2022 American Medical Association. All rights reserved.

eAppendix. Statistical Methods

Throughout, uppercase letters denote random variables and lowercase versions of the same letters denote realized values. Probands are indexed by i, and $m_i = 1, \ldots, 28$ indexes the site of enrollment of proband i. Twenty-six of these were advanced heart failure programs (one was inactivated after 3 probands were enrolled leaving 25 who completed the study), one was a satellite site of a program, and one was a virtual site at the coordinating center. This study included an embedded open-label randomized controlled trial of a behavioral intervention that was administered to probands in order to increase screening and surveillance uptake in family members.[1] Analyses described below included families in both the control and intervention arms because the analysis would be unaffected by variation in screening uptake among families as long as other assumptions were correct.

Prevalence of familial DCM among probands

Let n_i denote the number of living first-degree relatives of proband i that could have been screened at some point during the study period, $d_i \leq n_i$ denote the number of these who would have disease by a particular definition (dilated cardiomyopathy [DCM] or DCM, left ventricular systolic dysfunction [LVSD], or left ventricular enlargement [LVE] without known cause) if all n_i first-degree relatives were screened, $s_i \leq n_i$ denote the number actually screened by the study, and z_i denote a set of proband-level characteristics. A proband had familial DCM per a particular definition if $d_i > 0$; the goal was to model $\Pr(D_i > 0 | z_i, m_i)$, the prevalence of familial DCM among probands with characteristics z_i at site m_i. However, because it was not possible to observe or confirm disease in unenrolled first-degree relatives, the observed quantity was $d_i^* \leq \min(s_i, d_i)$, the number of living first-degree relatives of proband i who were enrolled in the study and found to have disease by a particular definition, rather than d_i. The original analysis plan[1] was to follow prior studies[2] in modeling $\Pr(D_i^* > 0 | z_i, m_i)$ directly. At each site, this probability represented a lower bound on $\Pr(D_i > 0 | z_i, m_i)$ because, in order for $D_i^* > 0$ to occur, $D_i > 0$ must have occurred as well, which implies that:

$$\Pr(D_i^* > 0 | z_i, m_i) = \Pr(D_i^* > 0, D_i > 0 | z_i, m_i) = \Pr(D_i > 0 | z_i, m_i) - \Pr(D_i^* = 0, D_i > 0 | z_i, m_i)$$

If all first-degree relatives that could have been screened were screened for all probands, then $S_i \equiv N_i$, the event $\{D_i^* = 0, D_i > 0\}$ would have probability zero, and modeling $\Pr(D_i^* > 0 | z_i, m_i)$ would directly model $\Pr(D_i > 0 | z_i, m_i)$. However, if $S_i < N_i$ for some probands, those with $D_i > 0$ may have had $D_i^* = 0$ simply because not all first-degree relatives were screened. In this case, $\Pr(D_i^* = 0, D_i > 0 | z_i, m_i) \geq 0$ and $\Pr(D_i^* > 0 | z_i, m_i) \leq \Pr(D_i > 0 | z_i, m_i)$. Moreover, if the joint distribution of S_i and N_i depended on z_i, $\Pr(D_i^* = 0, D_i > 0 | z_i, m_i)$ would also vary with z_i, and variation in $\Pr(D_i^* > 0 | z_i, m_i)$ at a given site might not reflect variation in $\Pr(D_i > 0 | z_i, m_i)$. Put differently, differences in the lower bound between proband subpopulations at a given site might reflect differences in first-degree relative participation or family structure in addition to differences in familial DCM prevalence.

An additional consideration was that $\Pr(D_i > 0 | z_i, m_i)$ depends on the distribution of N_i because, for larger n_i, there are more chances that at least one first-degree relative who could be screened had disease. More formally:

$$\Pr(D_i > 0 | z_i, m_i) = \sum_{n=1}^{\infty} \Pr(D_i > 0 | z_i, m_i, n) \Pr(N_i = n | z_i, m_i)$$

where $\Pr(D_i > 0 | z_i, m_i, n = 0) = 0$ trivially. Thus, even if $\Pr(D_i > 0 | z_i, m_i, n)$ and $\Pr(D_i > 0 | z_i', m_i, n)$ were identical for all n, it could happen that $\Pr(D_i > 0 | z_i, m_i) \neq \Pr(D_i > 0 | z_i', m_i)$ if $\Pr(N_i = n | z_i, m_i) \neq \Pr(N_i = n | z_i', m_i)$ for some n. As a result, an approach using the observed data $\{d_i^*, s_i, n_i\}$ to model $\Pr(D_i > 0 | z_i, m_i, n_i)$ was chosen instead.

The distribution of D_i^* in a proband randomly selected from among those with characteristics $\{z_i, n_i, s_i\}$ at site m_i was obtained by treating D_i as a latent variable and summing over its possible values with observed data $\{d_i^*, s_i, n_i\}$. With $n_i = 0$, $d_i^* \equiv d_i \equiv 0$, and $d_i^* \equiv 0$ with $s_i = 0$ regardless of n_i. Thus, for $n_i = 0$ or $s_i = 0$, $d_i^* = 0$ with probability 1. If $n_i > 0$ and $s_i > 0$, there were d_i^* individuals who were screened and had disease and $s_i - d_i^*$ who did not, which implies that at least d_i^* individuals and at most $n_i - (s_i - d_i^*)$ would have had disease if all n_i were screened. It follows that:

© 2022 American Medical Association. All rights reserved.

$$\Pr(D_i^* = d_i^* | z_i, m_i, n_i, s_i) = \begin{cases} I[d_i^* = 0] & \text{if } n_i = 0 \text{ or } s_i = 0 \\ \sum_{d=d_i^*}^{n_i-(s_i-d_i^*)} \Pr(D_i^* = d_i^* | D_i = d, z_i, m_i, n_i, s_i) \Pr(D_i = d | z_i, m_i, n_i) & \text{if } n_i > 0 \text{ and } s_i > 0 \end{cases} \quad (1)$$

In the above expression, the conditional independence of D_i from S_i given $\{z_i, m_i, n_i\}$ follows from assumptions made below about the probability model for first-degree relative screening.

To model $\Pr(D_i = d | z_i, m_i, n_i)$ in (1), D_i was conceptualized as the sum of n_i Bernoulli disease indicators in first-degree relatives that were correlated due to shared genetic and other risk factors. Although the disease probability in a particular first-degree relative may have depended on his or her age or gender, the former was not readily available for unenrolled first-degree relatives. A family-specific first-degree relative disease probability, $p_i(z_i, m_i)$, that was the same for all first-degree relatives of proband i with characteristics z_i at site m_i was therefore assumed; disease statuses for first-degree relatives in a family were assumed conditionally independent given $p_i(z_i, m_i)$. The quantity $p_i(z_i, m_i)$ can be conceptualized as an average disease probability for first-degree relatives in a particular family. Within-family correlation arose because $p_i(z_i, m_i)$ was allowed to vary randomly across families with the same z_i at site m_i, which implies a constant correlation between the disease statuses in pairs of first-degree relatives, also known as an exchangeable correlation structure. Allowing random variation in $p_i(z_i, m_i)$ across families also captured differences in this average first-degree relative disease probability between families arising from factors such as the age and gender composition of living first-degree relatives. Each family-specific first-degree relative disease probability was assumed to be an independent draw from a $\text{Beta}\big(p(z_i, m_i)\vartheta^{-1}, \big(1 - p(z_i, m_i)\big)\vartheta^{-1}\big)$ distribution, where $0 < p(z_i, m_i) < 1$ is the marginal disease probability in the population of first-degree relatives of probands with characteristics z_i at site m_i. As a result, D_i for a randomly selected proband from the population with characteristics $\{z_i, n_i\}$ at site m_i was assumed to follow a beta-binomial distribution given by:[3]

$$\Pr(D_i = d | z_i, m_i, n_i) = \binom{n_i}{d} \frac{\prod_{v=0}^{d-1}(p(z_i, m_i) + \vartheta v) \prod_{v=0}^{n_i-d-1}(1 - p(z_i, m_i) + \vartheta v)}{\prod_{v=0}^{n_i-1}(1 + \vartheta v)} \quad (2)$$

with mean $n_i\, p(z_i, m_i)$ and variance $n_i\, p(z_i, m_i)\big(1 - p(z_i, m_i)\big)\Big(1 + (n-1)\frac{\vartheta}{1+\vartheta}\Big)$, where $\frac{\vartheta}{1+\vartheta}$ is the common correlation between disease outcomes in pairs of first-degree relatives. Note that this probability mass function is defined only for $\vartheta \geq \max\big(-p(z_i, m_i)(n_i - 1)^{-1}, -\big(1 - p(z_i, m_i)\big)(n_i - 1)^{-1}\big)$. A major benefit of this beta-binomial parameterization, as opposed to a generalized linear mixed model (GLMM) with family-level random effects in the linear predictor, was that it replaced an intractable high-dimensional integral with the closed form in (2) for evaluating the marginal distribution of D_i in a subpopulation of probands at a particular site, $\Pr(D_i = d | z_i, m_i, n_i)$.

The model in (2) required specification of $p(z_i, m_i)$. Let $j_i = 1$ (non-Hispanic White), 2 (non-Hispanic Black), 3 (Hispanic White), 4 (Hispanic Black) denote the ethnicity-race group of proband i, $k_i = 1,2,3,4$ denote quartile of proband age of enrollment, and $l_i = 1$ (male), 2 (female) denote the proband sex. For $z_i = \{j_i, k_i, l_i\}$, the chosen model was:

$$\text{logit}\big(p(z_i, m_i)\big) = \alpha + \beta_{j_i} + \gamma_{k_i} + \delta_{l_i} + u_{m_i} + u_{j_i m_i} \quad (3)$$

where $\beta_1 \equiv \gamma_4 \equiv \delta_1 \equiv 0, u_{m_i} \sim N(0, \sigma_m^2)$ are random effects designed to reflect site-specific random variation in the marginal disease odds for a first-degree relative, and $u_{j_i m_i} \sim N(0, \sigma_{jm}^2)$ are random effects designed to reflect variation across sites in the differences in these odds between proband ethnicity-race groups at the same site. Under this model, $\exp\big(\gamma_{k_i}\big)$ is the ratio of the disease odds for a first-degree relative of a proband in enrollment age quartile k_i to those for a first-degree relative of a proband in enrollment age quartile 4 at the same site who is identical on all other factors. The odds ratio $\exp\big(\delta_{l_i}\big)$ is interpreted similarly. The odds ratio comparing a first-degree relative of a proband in ethnicity-race group j_i to a first-degree relative of a non-Hispanic White proband at the same site who is identical on all other factors is a random variable given by $\exp\big(\beta_{j_i} + u_{j_i m_i} - u_{1 m_i}\big)$. As a result, $\exp\big(\beta_{j_i}\big)$ is interpretable as the within-site odds ratio for a typical advanced heart failure program in the US (i.e., one at the mean or mode of the random effects distribution describing the population of such programs)[4-6] or the median within-site odds ratio across such programs if $\sigma_{jm}^2 > 0$.[7] In the event that $\sigma_{jm}^2 = 0$, all sites have the same within-site odds ratio.

© 2022 American Medical Association. All rights reserved.

Estimating the parameters in (2) and (3) required modeling $\Pr(D_i^* = d_i^* | D_i = d, z_i, m_i, n_i, s_i)$ in (1) to obtain the likelihood component for each proband. S_i was also conceptualized as the sum of n_i correlated Bernoulli random variables that arose when each of the n_i first-degree relatives in family i completed or did not complete screening independently with a probability $\pi_i(z_i, m_i)$ that varied randomly across probands with the same z_i at site m_i. This probability was assumed to depend on proband/family characteristics in a way that varied across sites but not on first-degree relative characteristics, such as age and disease status. Under these conditions, $\Pr\big(D_i^* = d_i^*, S_i = s_i | D_i = d, z_i, m_i, n_i, \pi_i(z_i, m_i)\big)$ was simply the probability of screening d_i^* of the d first-degree relatives with disease and $s_i - d_i^*$ of the $n_i - d$ without disease when each is screened independently with probability $\pi_i(z_i, m_i)$. As a result:

$$\begin{aligned}
&\Pr(D_i^* = d_i^* | D_i = d, z_i, m_i, n_i, \pi_i(z_i, m_i), s_i) \\
&= \frac{\binom{d}{d_i^*}\binom{n_i - d}{s_i - d_i^*}\pi_i(z_i, m_i)^{s_i}\big(1 - \pi_i(z_i, m_i)\big)^{n_i - s_i}}{\sum_{d_i^*=0}^{\min(d, s_i)}\binom{d}{d_i^*}\binom{n_i - d}{s_i - d_i^*}\pi_i(z_i, m_i)^{s_i}\big(1 - \pi_i(z_i, m_i)\big)^{n_i - s_i}} \\
&= \frac{\binom{d}{d_i^*}\binom{n_i - d}{s_i - d_i^*}}{\binom{n_i}{s_i}}
\end{aligned} \tag{4}$$

where the second equality followed from applying the Vandermonde identity in the denominator.[8] Thus, $\Pr(D_i^* = d_i^* | D_i = d, z_i, m_i, n_i, \pi_i(z_i, m_i), s_i)$ has the same hypergeometric distribution depending on only $\{d, n_i, s_i\}$ for every family, and it immediately follows that the marginal $\Pr(D_i^* = d_i^* | D_i = d, z_i, m_i, n_i, s_i)$ in (1) was also given by (4). Conditioning on s_i removes the impact of any family-level factor that could influence $\pi_i(z_i, m_i)$, including the randomized trial intervention, as well as any site-level factor. It also immediately follows from the denominator in the first equality in (4) that $\Pr\big(S_i = s_i | D_i = d, z_i, m_i, n_i, \pi_i(z_i, m_i)\big) = \binom{n_i}{s_i}\pi_i(z_i, m_i)^{s_i}\big(1 - \pi_i(z_i, m_i)\big)^{n_i - s_i}$, which does not depend on d. This implies that, after marginalizing over $\pi_i(z_i, m_i)$, $\Pr(S_i = s_i | D_i = d, z_i, m_i, n_i) = \Pr(S_i = s_i | z_i, m_i, n_i)$, thereby demonstrating conditional independence of D_i from S_i given $\{z_i, m_i, n_i\}$ assumed in (1).

Substituting (4) and (2) into (1) and simplifying the binomial coefficients yields the conditional likelihood component for a proband randomly selected from the population with characteristics $\{z_i, n_i, s_i\}$ at site m_i:

$$\begin{aligned}
&\Pr(D_i^* = d_i^* | z_i, m_i, n_i, s_i) \\
&= \begin{cases}
I[d_i^* = 0] & \text{if } n_i = 0 \text{ or } s_i = 0 \\
\displaystyle\sum_{d=d_i^*}^{n_i - (s_i - d_i^*)}\binom{s_i}{d_i^*}\binom{n_i - s_i}{d - d_i^*}\frac{\prod_{v=0}^{d-1}(p(z_i, m_i) + \vartheta v)\prod_{v=0}^{n_i - d - 1}(1 - p(z_i, m_i) + \vartheta v)}{\prod_{v=0}^{n_i - 1}(1 + \vartheta v)} & \text{if } n_i > 0 \text{ and } s_i > 0
\end{cases}
\end{aligned} \tag{5}$$

for $0 < p(z_i, m_i) < 1$ and $\vartheta \geq \max\big(-p(z_i, m_i)(n_i - 1)^{-1}, -\big(1 - p(z_i, m_i)\big)(n_i - 1)^{-1}\big)$. Probands' D_i^* were assumed to be conditionally independent within each site given the random effects, so the conditional likelihood was simply the product of the components in (5) over probands. To obtain estimates of the model parameters, SAS/STAT PROC NLMIXED was used to maximize the marginal likelihood obtained by integrating the conditional likelihood over the distributions of the u_{m_i} and $u_{j_i m_i}$ using adaptive Gaussian quadrature. Probands without living or screened first-degree relatives do not influence the likelihood because they have $d_i^* = 0$ with probability 1 for all parameter values; such probands were therefore excluded from likelihood calculation. To prevent optimization steps outside of the feasible region, constraints $\vartheta \geq 0$, $\sigma_m \geq 0$, and $\sigma_{jm} \geq 0$ were imposed. Although the probability mass function in (5) is defined for some $\vartheta < 0$, $\vartheta \geq 0$ was taken as the plausible range because correlation in the Bernoulli disease status between first-degree relatives was expected to be positive in general. Negative correlation was implausible and would likely indicate a model specification issue, so this constraint was not expected to be active at the solution. The loglikelihood component was also set to -1e-20 for any $p(z_i, m_i)$ within 1e-12 of 0 or 1 to avoid optimization steps into these regions. Model-based estimates of the covariance matrix were obtained as the inverse of the Hessian matrix at the solution.

Parameter estimates, standard errors, and odds ratios with Wald 95% confidence intervals (CIs) constructed using standard normal quantiles are presented in eTables 1 and 3. This model fit was used to obtain marginally standardized

© 2022 American Medical Association. All rights reserved.

estimates of familial DCM prevalence among probands at a typical US advanced heart failure program, as described below.

Age-specific cumulative risk for first-degree relatives

For each screened first-degree relative, it was possible to determine whether DCM or partial phenotypes were present by the age of enrollment. This observation scheme yielded current status data, a special type of interval-censored survival data[9, 10] that can be used to estimate age-specific cumulative risks. Let $r = 1, \ldots, s_i$ index the screened first-degree relatives of proband i, $j_i = 1$ (non-Hispanic White), 2 (non-Hispanic Black), 3 (Hispanic White), 4 (Hispanic Black) denote the proband ethnicity-race group, $k_i = 1, 2, 3, 4$ denote quartile of proband age at first diagnosis, $l_i = 1$ (male), 2 (female) denote the proband sex, and $l_{ir} = 1$ (male), 2 (female) denote first-degree relative sex. Age at disease onset, T_{ir}, in a first-degree relative with characteristics $\{j_i, k_i, l_i, l_{ir}\}$ at site m_i was assumed to have a marginal distribution with a Weibull baseline survivor function $S_0(t) = \exp[-\exp(a)\, t^b]$ influenced by covariates and random effects through a proportional hazards model,[9, 10] yielding the following marginal survivor function:

$$S(t|j_i, k_i, l_i, l_{ir}, m_i) = \exp\left[-\exp(a)\, t^b \exp\left(\theta_{j_i} + \phi_{k_i} + \tau_{l_i} + \omega_{l_{ir}} + v_{m_i} + v_{j_i m_i}\right)\right] \quad (6)$$

where $\theta_1 \equiv \phi_4 \equiv \tau_1 \equiv \omega_1 \equiv 0$, $v_{m_i} \sim N(0, \sigma_m^2)$ are random effects designed to reflect site-specific random variation in the marginal disease hazard for a first-degree relative, and $v_{j_i m_i} \sim N(0, \sigma_{jm}^2)$ are random effects designed to reflect variation across sites in the differences in these hazards between proband ethnicity-race groups at the same site. Under this model, $\exp(\phi_{k_i})$ is the ratio of the disease hazard for a first-degree relative of a proband in enrollment age quartile k_i to that for a first-degree relative of a proband in enrollment age quartile 4 at the same site who is identical on all other factors. The hazard ratios $\exp(\tau_{l_i})$ and $\exp(\omega_{l_{ir}})$ are interpreted similarly. The hazard ratio comparing a first-degree relative of a proband in ethnicity-race group j_i to a first-degree relative of a non-Hispanic White proband at the same site who is identical on all other factors is a random variable given by $\exp(\theta_{j_i} + v_{j_i m_i} - v_{1 m_i})$. As a result, $\exp(\theta_{j_i})$ is interpretable as the within-site hazard ratio for a typical advanced heart failure program in the US (i.e., one at the mean or mode of the random effects distribution describing the population of such programs)[4-6] or the median within-site hazard ratio across such programs if $\sigma_{jm}^2 > 0$.[7] In the event that $\sigma_{jm}^2 = 0$, all sites have the same within-site hazard ratio.

The parameters in (6) were estimated as follows using current status data. The age at disease onset was unobserved because DCM and partial phenotypes are typically asymptomatic for months or years before presentation. However, enrolling a first-degree relative at a particular age (C_{ir}) and examining him or her allowed determination of whether T_{ir} was before or after C_{ir} on the basis of whether the individual had disease at C_{ir}. Defining the observable random variable $Y_{ir} = I(T_{ir} \leq c_{ir})$ and assuming conditional independence of C_{ir} and T_{ir} given $\{j_i, k_i, l_i, l_{ir}, m_i\}$ yielded:

$$\Pr(Y_{ir}|C_{ir} = c_{ir}, j_i, k_i, l_i, l_{ir}, m_i) = [1 - S(c_{ir}|j_i, k_i, l_i, l_{ir}, m_i)]^{Y_{ir}}\, S(c_{ir}|j_i, k_i, l_i, l_{ir}, m_i)^{1-Y_{ir}}$$

Thus, Y_{ir} was a Bernoulli random variable with conditional success probability $\Pr(Y_{ir} = 1|c_{ir}, j_i, k_i, l_i, l_{ir}, m_i) = 1 - S(c_{ir}|j_i, k_i, l_i, l_{ir}, m_i)$. Furthermore, this probability can be related to the parameters in (6) by applying the complementary log-log link:

$$\ln(-\ln(1 - \Pr(Y_{ir} = 1|c_{ir}, j_i, k_i, l_i, l_{ir}, m_i))) = a + b \ln c_{ir} + \theta_{j_i} + \phi_{k_i} + \tau_{l_i} + \omega_{l_{ir}} + v_{m_i} + v_{j_i m_i} \quad (7)$$

With a single first-degree relative per proband, the model could have been fit using a standard GLMM with a binary outcome, site-level random effects, a complementary log-log link, and a linear predictor given by (7). However, with multiple first-degree relatives per proband, the effect of non-independence on estimated standard errors needed to be taken into account. The parameters and variance components for the model in (7) were therefore estimated using a generalized estimating equation (GEE)-type GLMM[4, 11] with a binary outcome, site-level random effects, a complementary log-log link, a linear predictor given by (7), and a compound symmetry working correlation matrix within families. This model was fit with residual subject-specific pseudolikelihood as implemented in SAS/STAT PROC GLIMMIX, and inference on fixed effects used the Morel-Bokossa-Neerchal corrected empirical covariance estimator with sites as independent units.[4, 11, 12] Estimation within this framework assumed that, given covariates and site, first-degree relative participation did not depend on Y_{ir}.

© 2022 American Medical Association. All rights reserved.

As in the familial DCM prevalence analysis, DCM or partial phenotypes in a first-degree relative were required to have no known cause for $Y_{ir} = 1$. Thus, $Y_{ir} = 0$ for individuals without LVSD or LVE as well as for those with DCM or partial phenotypes with a probable environmental cause. This approach is tantamount to assuming that first-degree relatives with a DCM phenotype arising from an environmental cause would not have developed disease absent this exposure. As these first-degree relatives were genetically at-risk, they may still have developed disease absent the exposure, and so this approach likely underestimates the age-specific cumulative risk.

Parameter estimates, standard errors, and hazard ratios with Wald 95% CIs constructed using standard normal quantiles are presented in eTables 5 and 6. Select hazard ratios and their 95% CIs are also shown in Table 4 of the main text. This model fit was also used to obtain marginally standardized estimates of age-specific cumulative risk among first-degree relatives at a typical US advanced heart failure program, as described below.

Marginally standardized estimates for a typical US advanced heart failure program

In addition to identifying factors affecting disease risk in first-degree relatives, the models in (3) and (7) can produce two types of estimates of familial DCM prevalence and age-specific cumulative risk: conditional estimates for a single advanced heart failure program in the US or marginal estimates across all advanced heart failure programs in the US. These two types of estimates will not be equal unless there is no heterogeneity in these programs,[4-6] and the appropriate choice depends on how the estimates will be applied.[13, 14] In the current context, we expect that these estimates will be used by clinicians to understand risks among different groups at their programs as well as for comparison with prior single-center studies, in which case conditional estimates for a single advanced heart failure program in the US are most relevant.[13, 14] For advanced heart failure programs in our study, empirical Bayes predictions of the random effects u_{m_i} and $u_{j_i m_i}$ (or v_{m_i} and $v_{j_i m_i}$) could have been used to generate program-specific estimates,[4, 13-15] but these would not apply to an external program not included in our sample.[4, 13, 14] To facilitate application at such a program, conditional estimates for a typical advanced heart failure program in the US with $u_{m_i} = u_{j_i m_i} = 0$ (or $v_{m_i} = v_{j_i m_i} = 0$),[4-6] which have the best expected within-program calibration for external programs,[13, 14] were presented. Such a typical US advanced heart failure program is defined by being at the mean or mode of the random effects distribution describing the population of such programs in the US.[4-6, 13, 14]

On the basis of the beta-binomial mixed model in (2) and (3), the familial DCM prevalence in the subpopulation of probands with characteristics $\{z_i, n_i\}$ at a typical US advanced heart failure program was obtained as:

$$\Pr\left(D_i > 0 | z_i, n_i, \mathbf{u}_{m_i} = 0\right) = 1 - \frac{\prod_{v=0}^{n_i-1}\left(1 - \left(1 + \exp\left(-\left(\alpha + \beta_{j_i} + \gamma_{k_i} + \delta_{l_i}\right)\right)\right)^{-1} + \vartheta v\right)}{\prod_{v=0}^{n_i-1}(1 + \vartheta v)} \tag{8}$$

To generate marginally standardized[16] ethnicity-race-specific familial DCM prevalence estimates at a typical US advanced heart failure program, the weighted average of prevalences in (8) was taken over $\{k_i, l_i, n_i\}$ in each ethnicity-race group assuming balance across the eight possible sex and age quartile combinations and $n_i \sim \text{Poisson}(4.53)$. The sum over n_i was truncated at 31, the smallest n_i for which the Poisson cumulative distribution function was within machine epsilon of 1. To marginalize over multiple ethnicity-race groups, the weighted average of the groups' familial DCM prevalences determined as described above was taken with weights given by the appropriate conditional distribution derived from 2019 US census population estimates.[17] For example, to calculate the marginally standardized familial DCM prevalence among Black probands, the average of the marginally standardized prevalences in Hispanic and non-Hispanic Black probands weighted by the estimated proportions of Hispanic and non-Hispanic Black individuals among those who identify as Black in the US population was taken. Further details regarding specific weights are provided in the footnotes to Table 2. The delta method[18] as implemented in SAS/STAT PROC NLMIXED was applied to these parameter estimates and their covariance matrix to obtain estimates and standard errors for marginally standardized familial DCM prevalence estimates, and Wald 95% CIs were obtained using standard normal quantiles. Marginally standardized familial DCM prevalence estimates for various ethnicity-race groups are presented in Table 2 of the main text, and differences between these estimates are presented in eTables 2 and 4.

The age-specific cumulative risk for a first-degree relative with characteristics $\{j_i, k_i, l_i, l_{ir}\}$ at a typical US advanced heart failure program was obtained as:

$$F\left(t | j_i, k_i, l_i, l_{ir}, \mathbf{v}_{m_i} = 0\right) = 1 - \exp\left[-\exp(a)\, t^b \exp\left(\theta_{j_i} + \phi_{k_i} + \tau_{l_i} + \omega_{l_{ir}}\right)\right] \tag{9}$$

© 2022 American Medical Association. All rights reserved.

Marginally standardized age-specific cumulative risk functions were obtained as the weighted average of $F(t|j_i, k_i, l_i, l_{ir}, \mathbf{v}_{m_i} = 0)$ over the distribution of $\{j_i, k_i, l_i, l_{ir}\}$ or a subset of these factors in a subpopulation of first-degree relatives for each $t = 1, \ldots, 80$. Further details regarding weights are provided in the legend to the Figure in the main text. The delta method[18] was applied to these parameter estimates and their covariance matrix to obtain estimates of marginally standardized age-specific cumulative risk functions and their standard errors at each t, which were used to construct Wald pointwise 95% CIs at each t using standard normal quantiles. These age-specific cumulative risk functions are shown in the Figure in the main text.

Different choices of weights could yield different absolute familial DCM prevalence and age-specific cumulative risk estimates than those reported. However, these choices were designed to reflect the likely experience at a typical advanced heart failure program serving a population representative of the US population in terms of the ethnicity-race groups considered, which should be a broadly relevant presentation for practicing clinicians. Moreover, differences in the odds or hazard of disease between subpopulations of first-degree relatives at such a program are also reflected as odds and hazard ratios that do not depend on our choices for marginal standardization. To assist researchers wishing to obtain their own marginally standardized estimates under different assumptions, parameter estimates and covariance matrices for each model have been provided in the following supplemental files:

1) fam_prev_ests_cov.csv (Supplement 2): For each model, this data set contains estimates for each parameter in the beta-binomial mixed model from (3) and (5) and their model-based covariance matrix. The MODEL column identifies the model in which a particular parameter was estimated, and the PARAMETER column identifies the parameter with data appearing in that row. The ESTIMATE column contains the parameter estimate, and the covariance matrix row for that parameter is displayed in the columns named for parameters. Each of these columns contains the covariance matrix column for the named parameter.
2) fdr_risk_ests_cov.csv (Supplement 3): For each outcome, this data set contains estimates for each fixed effect parameter in the binary-normal GEE-type GLMM in (7) and their Morel-Bokossa-Neerchal bias-corrected robust covariance matrix. The OUTCOME column identifies the outcome model in which a particular parameter was estimated, and combination of PARAMETER and LEVEL columns identifies the parameter with data appearing in that row. The ESTIMATE column contains the parameter estimate, and the covariance matrix row for that parameter is displayed in the columns with names COL*n*. The number *n* in the ROW column indicates that the covariance matrix column for that parameter appears in COL*n*; for example, the covariance between LOG_ENROLL_AGE and FEMALE = Yes can be found in COL3 of the LOG_ENROLL_AGE row and COL2 of the FEMALE = Yes row. Reference levels with estimates fixed to zero and the corresponding rows and columns of the covariance matrix that are identically zero have been omitted.

© 2022 American Medical Association. All rights reserved.

eTable 1. Model Fit for the Marginal Probability of DCM in the Population of First-Degree Relatives of Probands with Specified Characteristics at a Particular US Advanced Heart Failure Program

Proband Characteristic	Estimate	Standard Error	Odds Ratio (95% CI)	P value
Intercept	-2.7354	0.2351	-	-
Non-Hispanic White	0	-	1.00	-
Non-Hispanic Black	0.5146	0.1968	1.67 (1.14 – 2.46)	0.009
Hispanic White	0.0432	0.3536	1.04 (0.52 – 2.09)	0.90
Hispanic Black	-0.1020	1.0741	0.90 (0.11 – 7.41)	0.92
Enrollment age [15.78, 42.35]	0.3418	0.2428	1.41 (0.87 – 2.27)	0.16
Enrollment age [42.38, 52.81]	-0.1744	0.2662	0.84 (0.50 – 1.42)	0.51
Enrollment age [52.83, 61.83]	-0.1099	0.2598	0.90 (0.54 – 1.49)	0.67
Enrollment age [61.86, 85.12]	0	-	1.00	-
Female	0.3340	0.1776	1.40 (0.99 – 1.98)	0.06
Variance Parameter	**Estimate**	**Standard Error**		
Site (σ_m)	0.3274	0.1360		
Ethnicity-race within site (σ_{jm})	0[a]	-		
First-degree relative correlation ($\vartheta/(1+\vartheta)$)	0.0098	0.0316		

Estimated parameters for the beta-binomial mixed model from (3) and (5) were obtained using maximum likelihood with adaptive Gaussian quadrature. Model-based standard errors were obtained using the inverse Hessian, and two-sided p-values and Wald 95% confidence intervals were calculated using the standard normal distribution. Within-site odds ratios and their 95% confidence intervals were obtained by exponentiating corresponding estimates on the model scale. Probands without living or screened first-degree relatives provided no information on the parameters of interest in the likelihood; all model-based estimates were therefore derived from 822 probands (8 Hispanic Black, 64 Hispanic White, 290 Non-Hispanic Black, 460 Non-Hispanic White) with at least one screened first-degree relative.

[a] The final model excluded this random effect because convergence occurred on the boundary constraint of zero when it was included.

© 2022 American Medical Association. All rights reserved.

eTable 2. Differences in Marginally Standardized Familial DCM Prevalences (Standard Definition) between Proband Subpopulations Defined by Self-Reported Ethnicity and Race at a Typical US Advanced Heart Failure Program

Difference	% (95% CI)
Non-Hispanic Black – Non-Hispanic White	12.5 (2.9 – 22.1)
Hispanic Black – Hispanic White	-3.1 (-47.4 – 41.2)
Black – White	11.3 (1.9 – 20.8)
Hispanic White – Non-Hispanic White	0.9 (-14.3 – 16.2)
Hispanic Black – Non-Hispanic Black	-14.6 (-57.9 – 28.6)
Hispanic – Non-Hispanic	-1.4 (-15.9 – 13.1)

Differences between marginally standardized estimates for familial DCM prevalence (standard definition) in Table 2 with 95% confidence intervals (CIs) obtained using the delta method.

© 2022 American Medical Association. All rights reserved.

eTable 3. Model Fit for the Marginal Probability of DCM or Partial Phenotypes in the Population of First-Degree Relatives of Probands with Specified Characteristics at a Particular US Advanced Heart Failure Program

Proband Characteristic	Estimate	Standard Error	Odds Ratio (95% CI)	P value
Intercept	-1.6177	0.1661	-	-
Non-Hispanic White	0	-	1.00	-
Non-Hispanic Black	0.1692	0.1442	1.18 (0.89 – 1.57)	0.24
Hispanic White	-0.0916	0.2540	0.91 (0.55 – 1.50)	0.72
Hispanic Black	-0.4102	0.7955	0.66 (0.14 – 3.15)	0.61
Enrollment age [15.78, 42.35]	0.2800	0.1797	1.32 (0.93 – 1.88)	0.12
Enrollment age [42.38, 52.81]	0.2054	0.1825	1.23 (0.86 – 1.76)	0.26
Enrollment age [52.83, 61.83]	-0.0790	0.1867	0.92 (0.64 – 1.33)	0.67
Enrollment age [61.86, 85.12]	0	-	1.00	-
Female	0.2619	0.1261	1.30 (1.01 – 1.66)	0.04
Variance Parameters	**Estimate**	**Standard Error**		
Site (σ_m)	0.2417	0.0906		
Ethnicity-race within site (σ_{jm})	0[a]	-		
First-degree relative correlation ($\vartheta/(1+\vartheta)$)	0.0457	0.0330		

Estimated parameters for the beta-binomial mixed model from (3) and (5) were obtained using maximum likelihood with adaptive Gaussian quadrature. Model-based standard errors were obtained using the inverse Hessian, and two-sided p-values and Wald 95% confidence intervals were calculated using the standard normal distribution. Within-site odds ratios and their 95% confidence intervals were obtained by exponentiating corresponding estimates on the model scale. Probands without living or screened first-degree relatives provided no information on the parameters of interest in the likelihood; all model-based estimates were therefore derived from 822 probands (8 Hispanic Black, 64 Hispanic White, 290 Non-Hispanic Black, 460 Non-Hispanic White) with at least one screened first-degree relative.

[a] The final model excluded this random effect because convergence occurred on the boundary constraint of zero when it was included.

© 2022 American Medical Association. All rights reserved.

eTable 4. Differences in Marginally Standardized Familial DCM Prevalences (Expanded Definition) between Proband Subpopulations Defined by Self-Reported Ethnicity and Race at a Typical US Advanced Heart Failure Program

Difference	% (95% CI)
Non-Hispanic Black – Non-Hispanic White	4.9 (-3.2 – 13.0)
Hispanic Black – Hispanic White	-9.2 (-54.4 – 36.1)
Black – White	4.3 (-3.9 – 12.5)
Hispanic White – Non-Hispanic White	-2.7 (-17.2 – 11.8)
Hispanic Black – Non-Hispanic Black	-16.7 (-60.8 – 27.3)
Hispanic – Non-Hispanic	-4.0 (-17.8 – 9.9)

Differences between marginally standardized estimates for familial DCM prevalence (expanded definition) in Table 2 with 95% confidence intervals (CIs) obtained using the delta method.

© 2022 American Medical Association. All rights reserved.

eTable 5. Model Fit for the Age-Specific Cumulative Risk of DCM in a First-Degree Relative of a Patient with DCM at a Particular US Advanced Heart Failure Program

Weibull Baseline Survivor Function Parameters	**Estimate**	**SE**	**95% CI**		
a	-8.1069	0.9188	-9.9078 – -6.3061		
b	1.3554	0.2205	0.9232 – 1.7876		
Covariates	**Estimate**	**SE**	***P***	**Hazard Ratio (95% CI)**	
First-degree relative					
Female	-0.3339	0.2014	0.10	0.72 (0.48 – 1.06)	
Proband					
Non-Hispanic White	0	-	-	1.00	
Non-Hispanic Black	0.6365	0.2063	0.002	1.89 (1.26 – 2.83)	
Hispanic White	0.2468	0.3499	0.48	1.28 (0.64 – 2.54)	
Hispanic Black	0.4102	0.8397	0.63	1.51 (0.29 – 7.82)	
Diagnosis age [4.73, 34.33]	0.7816	0.3088	0.01	2.19 (1.19 – 4.00)	
Diagnosis age [34.34, 44.16]	0.4936	0.3009	0.10	1.64 (0.91 – 2.95)	
Diagnosis age [44.18, 53.62]	0.2041	0.3224	0.53	1.23 (0.65 – 2.31)	
Diagnosis age [53.66, 82.67]	0	-	-	1.00	
Female	0.3465	0.1600	0.03	1.41 (1.03 – 1.93)	
Variance Parameters	**Estimate**	**SE**			
Site (σ_m^2)	0.1065	0.0820			
Ethnicity-race within site (σ_{jm}^2)	0[a]	-			
Compound Symmetry	0.0176	0.0304			
Residual	0.8687	0.0421			

Estimated parameters for the binary-normal GEE-type GLMM in (7) with a compound symmetry working correlation structure within each family were obtained using residual subject-specific pseudolikelihood. Bias-corrected robust standard errors were obtained using the Morel-Bokossa-Neerchal correction, and two-sided p-values and Wald 95% confidence intervals were calculated using the standard normal distribution. Within-site hazard ratios and their 95% confidence intervals were obtained by exponentiating corresponding estimates on the model scale. *a* and *b* are from the Weibull baseline survivor function of the form $S_0(t) = \exp[-\exp(a)\, t^b]$. 1692 first-degree relatives contributed to this analysis (1 non-Hispanic White first-degree relative was excluded due to missing covariate data).

[a] The final model excluded this random effect because convergence occurred on the boundary constraint of zero when it was included.

© 2022 American Medical Association. All rights reserved.

eTable 6. Model Fit for the Age-Specific Cumulative Risk of LVE, LVSD, or DCM in a First-Degree Relative of a Patient with DCM at a Particular US Advanced Heart Failure Program

Weibull Baseline Survivor Function Parameters	**Estimate**	**SE**	**95% CI**	
a	-4.9505	0.5662	-6.0602 – -3.8409	
b	0.8224	0.1397	0.5486 – 1.0961	
Covariates	**Estimate**	**SE**	***P***	**Hazard Ratio (95% CI)**
First-degree relative				
Female	-0.0583	0.1311	0.66	0.94 (0.73 – 1.22)
Proband				
Non-Hispanic White	0	-	-	1.00
Non-Hispanic Black	0.2403	0.1279	0.06	1.27 (0.99 – 1.63)
Hispanic White	0.0395	0.2351	0.87	1.04 (0.66 – 1.65)
Hispanic Black	-0.0706	0.6908	0.92	0.93 (0.24 – 3.61)
Diagnosis age [4.73, 34.33]	0.5109	0.1934	0.008	1.67 (1.14 – 2.44)
Diagnosis age [34.34, 44.16]	0.4690	0.1966	0.02	1.60 (1.09 – 2.35)
Diagnosis age [44.18, 53.62]	0.2445	0.1718	0.15	1.28 (0.91 – 1.79)
Diagnosis age [53.66, 82.67]	0	-	-	1.00
Female	0.2031	0.1216	0.10	1.23 (0.97 – 1.56)
Variance Parameters	**Estimate**	**SE**		
Site (σ_m^2)	0.0458	0.0345		
Ethnicity-race within site (σ_{jm}^2)	0[a]	-		
Compound Symmetry	0.0605	0.0309		
Residual	0.9130	0.0421		

Estimated parameters for the binary-normal GEE-type GLMM in (7) with a compound symmetry working correlation structure within each family were obtained using residual subject-specific pseudolikelihood. Bias-corrected robust standard errors were obtained using the Morel-Bokossa-Neerchal correction, and two-sided p-values and Wald 95% confidence intervals were calculated using the standard normal distribution. Within-site hazard ratios and their 95% confidence intervals were obtained by exponentiating corresponding estimates on the model scale. *a* and *b* are from the Weibull baseline survivor function of the form $S_0(t) = \exp[-\exp(a)\, t^b]$. 1692 first-degree relatives contributed to this analysis (1 non-Hispanic White first-degree relative was excluded due to missing covariate data).

[a] The final model excluded this random effect because convergence occurred on the boundary constraint of zero when it was included.

© 2022 American Medical Association. All rights reserved.

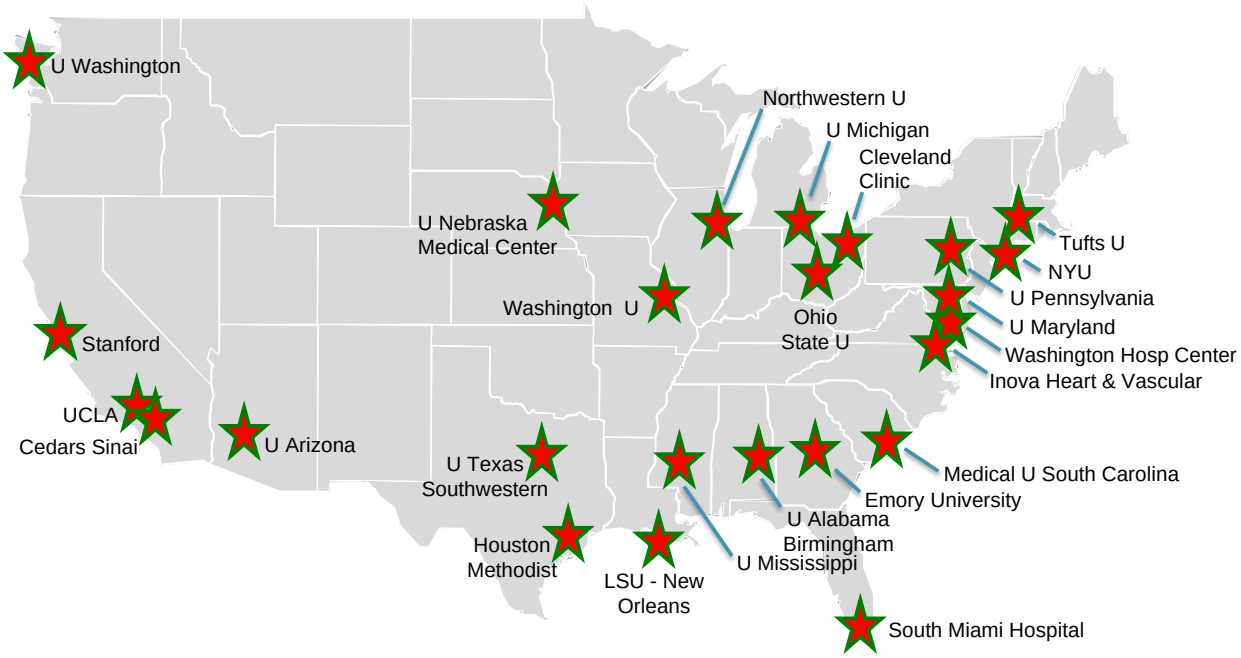
eTable 7. Vital Status of First-Degree Relatives at Proband Enrollment, by Proband Ethnicity and Race

	Hispanic N = 566		Non-Hispanic N = 6640	
	Black N = 48	White N = 518	Black N = 3135	White N = 3505
Deceased at proband enrollment, No. (%)	17 (35)	109 (21.0)	754 (24.1)[a]	800 (22.8)

[a] Actual denominator was 3133 due to 2 first-degree relatives having unknown vital status.

© 2022 American Medical Association. All rights reserved.

eFigure 1. Clinical Sites of the DCM Consortium.



Shown are the 25 clinical sites of the DCM Consortium who enrolled probands and family members for this study through the close of family member enrollment on April 1, 2021. The Ohio State University site served as an enrolling site as well as the coordinating center for this study. The table below shows the probands and first-degree relatives contributed to the final sample in eFigure 2 per clinical site. One inactivated clinical site included in the table is not shown on the map.

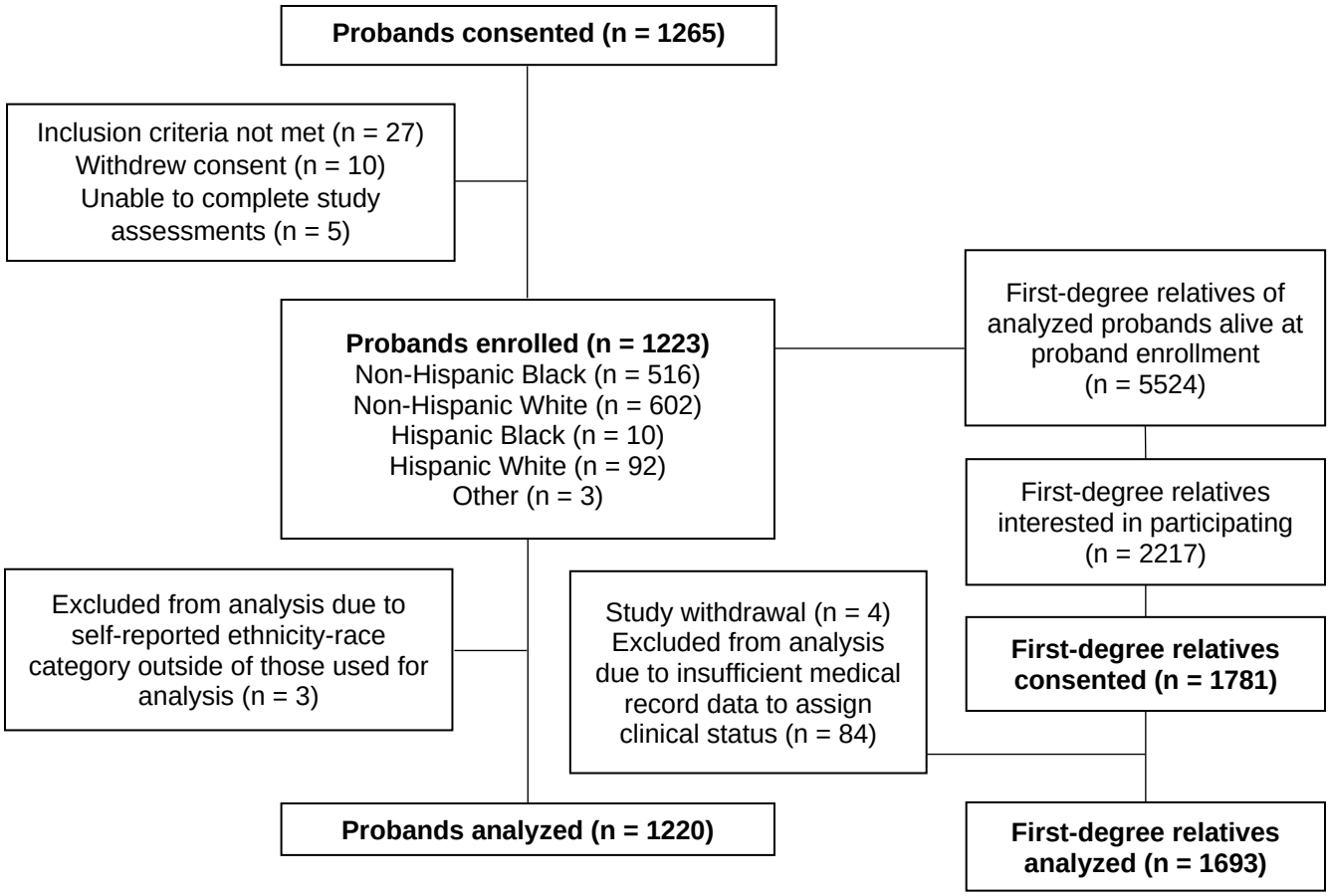
	Probands		First-Degree Relatives	
Clinical Site	**No.**	**%**	**No.**	**%**
Cedars-Sinai Medical Center	49	4.0	130	7.7
Cleveland Clinic	110	9.0	153	9.0
Emory University	45	3.7	48	2.8
Houston Methodist	60	4.9	56	3.3
Inova Heart and Vascular Institute	53	4.3	50	3.0
Louisiana State University	25	2.1	18	1.1
Medical University of South Carolina	8	0.7	9	0.5
Medstar Washington Hospital Center	149	12.2	192	11.3
New York University Langone Medical Center	26	2.1	39	2.3
Northwestern University	34	2.8	35	2.1
Ohio State University	168	13.8	318	18.8
South Miami Hospital	32	2.6	38	2.2
Stanford University	29	2.4	54	3.2
Tufts University	49	4.0	42	2.5
University of California Los Angeles Medical Center	16	1.3	14	0.8
University of Texas Southwestern Medical Center	28	2.3	40	2.4
University of Alabama Birmingham	32	2.6	43	2.5

© 2022 American Medical Association. All rights reserved.

	Probands		First-Degree Relatives	
Clinical Site	**No.**	**%**	**No.**	**%**
University of Arizona	52	4.3	52	3.1
University of Maryland	26	2.1	20	1.2
University of Michigan	18	1.5	24	1.4
University of Mississippi	10	0.8	12	0.7
University of Nebraska	38	3.1	69	4.1
University of Pennsylvania	72	5.9	85	5.0
University of Washington	39	3.2	36	2.1
Washington University	49	4.0	113	6.7
Inactivated Site	3	0.3	3	0.2
Total	**1220**	**100**	**1693**	**100**

© 2022 American Medical Association. All rights reserved.

eFigure 2. DCM Precision Medicine Study Participant Recruitment



© 2022 American Medical Association. All rights reserved.

eReferences

1. Kinnamon DD, Morales A, Bowen DJ, Burke W, Hershberger RE and for the DCM Consortium. Toward Genetics-Driven Early Intervention in Dilated Cardiomyopathy: Design and Implementation of the DCM Precision Medicine Study. *Circ Cardiovasc Genet*. 2017;10:e001826.
2. Petretta M, Pirozzi F, Sasso L, Paglia A and Bonaduce D. Review and metaanalysis of the frequency of familial dilated cardiomyopathy. *Am J Cardiol*. 2011;108:1171-6.
3. Prentice RL. Binary Regression Using an Extended Beta-Binomial Distribution, with Discussion of Correlation Induced by Covariate Measurement Errors. *J Am Stat Assoc*. 1986;81:321-327.
4. Stroup W. *Generalized linear mixed models: modern concepts, methods and applications.* New York: CRC Press; 2012.
5. Austin PC and Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. *Stat Med*. 2017;36:3257-3277.
6. Molenberghs G and Verbeke G. *Models for discrete longitudinal data*. New York: Springer; 2005.
7. Larsen K, Petersen JH, Budtz-Jorgensen E and Endahl L. Interpreting parameters in the logistic regression model with random effects. *Biometrics*. 2000;56:909-14.
8. Koepf W. *Hypergeometric Summation: An Algorithmic Approach to Summation and Special Function Identities.* 2nd ed. London: Springer; 2014.
9. Lawless JF. *Statistical Models and Methods for Lifetime Data.* 2nd ed. Hoboken: John Wiley & Sons; 2011.
10. Jewell NP and van der Laan MJ. Current Status Data: Review, Recent Developments and Open Problems. *UC Berkeley Division of Biostatistics Working Paper Series*. 2002; Working Paper 113.
11. Vonesh E. *Generalized Linear and Nonlinear Models for Correlated Data: Theory and Applications Using SAS.* 1st ed. Cary, NC: SAS Institute; 2012.
12. Morel JG, Bokossa MC and Neerchal NK. Small sample correction for the variance of GEE estimators. *Biometrical J*. 2003;45:395-409.
13. Falconieri N, Van Calster B, Timmerman D and Wynants L. Developing risk models for multicenter data using standard logistic regression produced suboptimal predictions: A simulation study. *Biom J*. 2020;62:932-944.
14. Wynants L, Vergouwe Y, Van Huffel S, Timmerman D and Van Calster B. Does ignoring clustering in multicenter data influence the performance of prediction models? A simulation study. *Stat Methods Med Res*. 2018;27:1723-1736.
15. Skrondal A and Rabe-Hesketh S. Prediction in multilevel generalized linear models. *J R Stat Soc A Stat*. 2009;172:659-687.
16. Muller CJ and MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *Int J Epidemiol*. 2014;43:962-70.
17. U.S. Census Bureau Population Division. Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States: April 1, 2010 to July 1, 2019 (NC-EST2019-SR11H). Jun 2020.
18. van der Vaart AW. *Asymptotic statistics*. New York: Cambridge University Press; 1998.

© 2022 American Medical Association. All rights reserved.