Human Genetics (2019) 138:739–748
https://doi.org/10.1007/s00439-019-02024-6

ORIGINAL INVESTIGATION

Estimation of metabolic syndrome heritability in three large populations including full pedigree and genomic information

Francesca Graziano[1,2] · Ginevra Biino[3] · Maria Teresa Bonati[4] · Benjamin M. Neale[5,6] · Ron Do[7,14] · Maria Pina Concas[13] · Simona Vaccargiu[8] · Mario Pirastu[8] · Oscar Terradura-Vagnarelli[9] · Massimo Cirillo[10] · Martino Laurenzi[9] · Mario Mancini[11] · Alberto Zanchetti[4,12] · Mario Grassi[2]

Received: 7 November 2018 / Accepted: 29 April 2019 / Published online: 1 June 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Metabolic syndrome is a complex human disorder characterized by a cluster of conditions (increased blood pressure, hyperglycemia, excessive body fat around the waist, and abnormal cholesterol or triglyceride levels). Any of these conditions increases the risk of serious disorders such as diabetes or cardiovascular disease. Currently, the degree of genetic regulation of this syndrome is under debate and partially unknown. The principal aim of this study was to estimate the genetic component and the common environmental effects in different populations using full pedigree and genomic information. We used three large populations (Gubbio, ARIC, and Ogliastra cohorts) to estimate the heritability of metabolic syndrome. Due to both pedigree and genotyped data, different approaches were applied to summarize relatedness conditions. Linear mixed models (LLM) using average information restricted maximum likelihood (AIREML) algorithm were applied to partition the variances and estimate heritability (h^2) and common sib–household effect (c^2). Globally, results obtained from pedigree information showed a significant heritability (h^2: 0.286 and 0.271 in Gubbio and Ogliastra, respectively), whereas a lower, but still significant heritability was found using SNPs data (h^2_{SNP}: 0.167 and 0.254 in ARIC and Ogliastra). The remaining heritability between h^2 and h^2_{SNP} ranged between 0.031 and 0.237. Finally, the common environmental c^2 in Gubbio and Ogliastra were also significant accounting for about 11% of the phenotypic variance. Availability of different kinds of populations and data helped us to better understand what happened when heritability of metabolic syndrome is estimated and account for different possible confounding. Furthermore, the opportunity of comparing different results provided more precise and less biased estimation of heritability.

Introduction

Metabolic syndrome (MetS) is a polygenic complex disease with a strong genetic basis. MetS may be the result of the interplay between genetic and environmental factors, such as diet, physical activity, alcohol intake, and smoking that modulate complex networks of metabolic pathways (Andreassi and Botto 2003). However, the fundamental basis of the syndrome is still under investigation; in particular, due to the variability of the results, both the amount of genetic component and which set of genes—and how they contribute to its development—are yet unclear.

A traditional tool that accounts for genetic effects is the estimation of heritability. Many definitions of heritability have been proposed, but, in general, heritability represents the amount of variation in a phenotype that is influenced by genetic variation and it is defined as the proportion of variance for a phenotype that is explained by the sharing of genomic regions. Specifically, for a particular measurement, taken at a particular time or age, it is defined as the proportion of total variance in a population that is attributable to variation in additive genetic or total genetic values—termed narrow-sense heritability (or just heritability, h^2) or broad-sense heritability (H^2), respectively (Visscher et al. 2008; Visscher et al. 2006). Broad-sense heritability captures the proportion of phenotypic variation due to genetic values that may include effects due

Alberto Zanchetti: Deceased.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00439-019-02024-6) contains supplementary material, which is available to authorized users.

✉ Francesca Graziano
francesca.graziano@unimib.it

Extended author information available on the last page of the article

to dominance (interactions between alleles at the same locus) and epistasis (interactions between alleles at different loci). On the other hand, narrow-sense heritability captures only that proportion of genetic variation that is due to additive genetic values. This parameter could be biased by shared environmental effects. When close relatives have more similar environments than distant relatives, parameters could overestimate heritability, as it is difficult to distinguish between similarity due to genetic effects and environmental effects. In this paper, narrow-sense heritability (h^2) is the parameter that we have used to indicate heritability.

Recently, thanks to the huge genotype data collected in genome-wide association studies (GWAS) or in whole genome-scale sequencing, a new approach has been developed, so that heritability is inferred from the proportion of phenotypic variance explained by genotyped SNPs: h^2_{SNP} estimates how much of the resemblance is due to shared genotypes (relatedness) among unrelated subjects (Vinkhuyzen et al. 2013; Yang et al. 2010, 2011a).

The discrepancy between h^2 estimated from pedigrees and h^2_{SNP} is known as "still-missing heritability" and the interpretation of this phenomenon is under debate (Bourrat and Lu 2017; Makowsky et al. 2011; Manolio et al. 2009; Witte et al. 2014; Zuk et al. 2012).

The aim of this study was to estimate the proportion of variation in metabolic syndrome, considered as a continuous trait, captured either by genotyped SNPs or by family information, or combining both data. By comparing different estimates, it is indeed possible to appropriately describe causes of intraindividual variation and detect the relative contribution of genetic loci and environmental factors.

Materials and methods

Population cohorts

Three large cohorts, Ogliastra, Gubbio, and Atherosclerosis Risk in Communities Study (ARIC), were used to estimate the heritability of MetS.

The Gubbio Population Study was a prospective epidemiological investigation on blood pressure and cardiovascular risk factors started in 1983 in Gubbio, a town in central Italy. Three surveys (the first in 1983–1985, the second in 1988–1992, and the third in 2001–2007) were conducted over about 25 years (Bonati et al. 2014; Cirillo et al. 2014; Menotti et al. 2009). At each survey, genealogical information was also registered and updated through a structured interview administered to each participant. From these data, nuclear and extended pedigrees were drawn. Nuclear pedigrees are two-generation families with first-degree relationships, that is, parent–offspring and/or sibling (Khoury et al. 1993). Nuclear pedigrees from the last survey have been considered due to the comparability with data collected in Ogliastra and ARIC studies in terms of span of years. The final Gubbio sample consisted of 2620 subjects split into 711 nuclear pedigrees.

The ARIC sample, sponsored by the National Heart, Lung and Blood Institute (NHLBI), was a prospective epidemiologic study conducted in four US communities (Investigators 1989). ARIC was designed to investigate the causes of atherosclerosis and its clinical outcomes, variation in cardiovascular risk factors, medical care, and disease by race, gender, location and date, and also includes genetic marker (SNP) information. For this study, a restricted subgroup of European-Americans was considered. The ARIC population consisted of 8451 subjects without family information. Nevertheless, a small subsample of 530 individuals had a genetic relationship ranging between 0.35 and 0.65; thus they were considered as a group of related subjects (Vattikuti et al. 2012).

Lastly, the Ogliastra Study is a large population-based epidemiologic survey carried out in the villages of the Ogliastra region in Sardinia, Italy, between 2002 and 2008 and aimed at studying complex diseases (Biino et al. 2011). Among the 12,517 subjects, 8102 (3485 men and 4617 women older than 18 years) were included in the analysis due to complete information on their MetS components. A subset of 1270 subjects was genotyped for SNPs and for almost all the individuals the genealogic information was available. Both Gubbio and Ogliastra samples have particular characteristics: no immigration and isolation (more details will be reported in "Discussion"). Thus, a final sample of 20,201 subjects was analyzed.

Data collection for each population study was approved by the appropriate institutional review boards and participants provided written informed consent.

MetS trait

A MetS score was calculated for each subject of the three cohorts according to an implemented and validated algorithm that takes into account gender and seven quantitative traits commonly used for classical Mets diagnosis (Graziano et al. 2015). Briefly, the algorithm calculated after confirmatory bifactor analysis (bCFA) summarizes waist circumference (WC), BMI, blood pressure (both systolic, SBP, and diastolic, DBP), blood glucose (GLU), HDL-cholesterol, and triglycerides (TRIG) into one quantitative phenotype.

Scores were defined by the sum of each centered and scaled MetS variable weighted by the corresponding ratio between factor loading and residual variance derived from bCFA. Then, the MetS score was rescaled to a range of 0–100. Specifically, equation for MetS score was equal to

$$\text{MetS} = 0.645 \times \text{WC} + 0.933 \times \text{BMI} + 0.059 \times \text{SBP} + 0.087$$

$\times$ DBP + 0.011 $\times$ GLU − 0.022 $\times$ HDL + 0.003 $\times$ TRIG − 63.0 for males, and MetS = 0.342 $\times$ WC + 0.636 $\times$ BMI + 0.133 $\times$ SBP + 0.146 $\times$ DBP + 0.021 $\times$ GLU − 0.027 $\times$ HDL + 0.009 $\times$ TRIG − 44.4 for females.

In this way, all clinical variables used to diagnose the disease are related to the overall continuous construct, MetS score. The syndrome becomes clinically interpretable (a degree of severity was indicated for each subject depending on his score value) and useful for investigating the genetic components (Graziano et al. 2015, 2016). About the MetS score performance, ROC curve and AUC showed a good predictive accuracy, with 0.80 specificity and 0.80 sensitivity.

Statistical analysis

To estimate heritability and family effects, we conducted three independent heritability analyses, one for each cohort. A diagram of the analytic process and cohorts' features is shown in Fig. 1.

In Gubbio and Ogliastra cohorts, we could explore additive genetic effects due to pedigrees (h^2) and common sib–household effect (c^2). Then, after a standard quality control (Online Appendix section for more details), we estimated heritability using genotyped SNPs (h^2_{SNP}) both in the ARIC cohort, by selecting the unrelated subjects, and in the Ogliastra cohort, by combining family and genotype information. Finally, in the ARIC cohort, after selection of related subjects, we estimated the related genomic heritability (h^2_{REL}).

To allow the estimation of heritability and family effects using pedigree structure and/or genotype data, we computed five relatedness matrices:

- kinship (K) and sibs–household (C) matrices refer to pedigree studies;
- genomic relationship (G_{SNP}) and genomic-related (G_{REL}) matrices refer to genomic data;
- mixed-related matrix (K_{SNP}) refers to pedigree with genomic data.

In detail, kinship (K) is a matrix that contains the pairwise relationship between all subjects derived from genealogic information or "distances" among individuals (Almasy and Blangero 1998) and sibs–household matrix (C) accounts for the common environmental effects shared by siblings (e.g., uterine diets, shared house) (Bennett et al. 1995). In our case, using sib relationship from pedigrees, we consider C as longtime-shared household effect (Xia et al. 2016).

Genomic relationship matrix (G_{SNP}) contains the estimated genomic relatedness between pairs of individuals calculated from identity-by-state marker relationships implemented by GCTA (Speed et al. 2017; Yang et al. 2010, 2011b). The G_{SNP} matrix is sensitive to linkage disequilibrium (LD), though the use of this technique before h^2 estimation is still under debate (Evans et al. 2018; Hill and Maki-Tanila 2015; Lee and Chow 2014; Speed et al. 2017; Yang et al. 2015). We chose an LD pruning with different levels assuming that most causal variants of metabolic syndrome are largely unknown (Abou Ziki and Mani 2016). Thus, a G_{SNP} matrix using different cutoffs (0.20, 0.80 and none) was estimated. No significant differences were found (see Table S1 in Online Appendix). Thus, a cutoff at 0.20 was selected for conservative and computational reasons. The G_{SNP} matrices both in ARIC and Ogliastra was performed using samples typed SNPs.

In ARIC sample also a genomic-related matrix (G_{REL}) was computed when G_{SNP} had estimated genomic relatedness $\geq$ 0.025 (Zaitlen et al. 2013); Thus, G_{REL} matrix is define as a block matrix with G_{SNP} values for the subgroup of ARIC related subjects (first cousins or closer), zero for unrelated subjects and one in

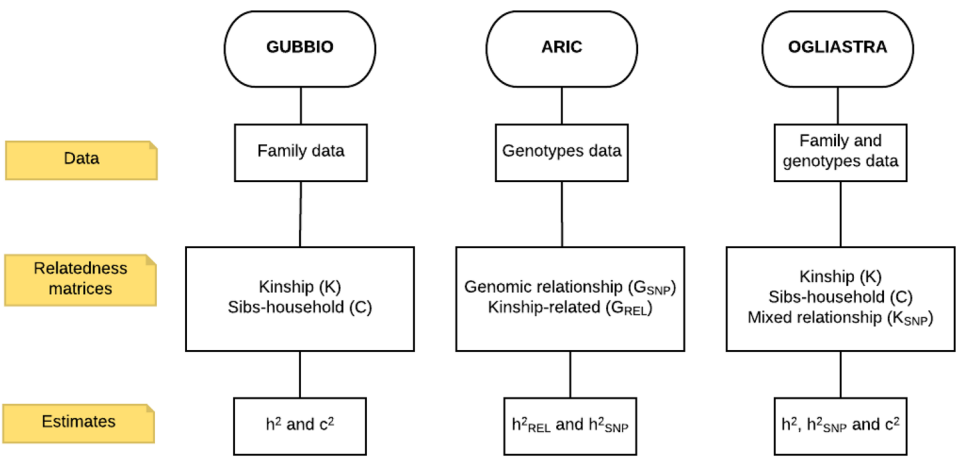

Fig. 1 Diagram of the data available, relatedness matrices, and estimates available or computed in the Gubbio, ARIC, and Ogliastra cohorts

diagonal (Online Appendix section for more details). Considering both G_{REL} and G_{SNP} matrices in the two-variance component model, still-missing heritability and heritability attributed to the common SNPs (h^2_{SNP}) were estimated.

Furthermore, in Ogliastra, thanks to the availability of both SNPs and extended pedigrees data, a mixed relationship matrix (K_{SNP}), a complex matrix including full pedigrees and genomic information, was created applying Legarra et al.'s proposal (Legarra et al. 2009; Misztal et al. 2009; Tucker et al. 2015). Thus, after partitioning subjects into several groups (including genealogic information, ungenotyped and genotyped individuals), we created a matrix that includes both pedigree-based and genomic-based relationships.

Definition of heritability estimates and corresponding relationship matrices are summarized in Table 1, while the algebra definitions of each relationship matrix are illustrated in the Online Appendix section.

Estimating MetS heritability

We used genomic and environmental matrices described above to estimate the variance components of continuous MetS score. Linear mixed models (LMM) using average information restricted maximum likelihood (AIREML) algorithm were fitted to partition the variances (Gilmour et al. 1995). The ACE (additive, common, error) variance components equation used for each model is the following:

$$y = X\beta + a + c + e \text{ with } V(y) = \sigma^2_A A + \sigma^2_C C + \sigma^2_E I,$$

where y is the vector of n individuals' MetS score, X is the matrix of observed covariates corresponding to the fixed effect in β, a is the additive polygenic effect vector, c is the common environmental effect vector, and e is the error residual effect vector. The matrices A, C, and I are the known genomic, environmental, and identity (orthogonal) relationship matrices, respectively, while σ^2_A, σ^2_C and σ^2_E are the corresponding variances.

Generally, narrow-sense heritability, both from pedigree and/or from SNPs data, is defined as the ratio σ^2_A/σ^2_T, where $\sigma^2_T = \sigma^2_A + \sigma^2_C + \sigma^2_E$ is the total variance of the quantitative phenotype, σ^2_A is the proportion of all variance due to the genetic effect and is equal to σ^2_K for pedigree data and $\sigma^2_{(REL\ -\ SNP)} + \sigma^2_{SNP}$ for genomic/mixed data. Moreover, c^2 is defined from pedigree data as the ratio σ^2_C/σ^2_T and represents the proportion of all variance attributable to shared sib–household environment.

A total of six models for all possible interesting combinations of relatedness matrices were used (Table 2). All models were adjusted for age, but not for gender because it is included in the MetS score equation. In ARIC cohort, all models were adjusted for the first ten PCs to remove some residual stratification. As shown in Table 2, model (a) and model (b) were used for the Gubbio and Ogliastra cohort, model (c) and (d) for the ARIC study and models (a), (b), (e) and (f) for the Ogliastra cohort.

We use R v.3.5.0 packages for data analysis (TeamR 2018). Pedigree matrices and genetic relationship matrices were computed with kinship2 (Sinnwell et al. 2014) and SNPRelate (Zheng et al. 2012) packages, respectively. AIREML estimates and 95% confidence intervals (95% CI) of the p fixed and random effects, and model selection statistics (log-likelihoods (LL), Akaike information criterion: $AIC = -2LL + 2 \times p$ and Bayesian information criterion, $BIC = -2LL + p \times \log(n)$) (Müller et al. 2013) for all models were performed and twofold checked using both gaston (Dandine-Roulland and Perdry 2017) and GENESIS (Conomos et al. 2017) packages.

Results

Demographic characteristics

The MetS score and its components in Gubbio, ARIC, and Ogliastra participating cohorts are described in Table 3. Sample sizes, means, standard deviations (SD), and ranges are reported by gender.

Heritability in the Gubbio, ARIC, and Ogliastra population

We estimated heritability and sib–household effects (where possible) for MetS score in the Gubbio, ARIC, and Ogliastra populations. Considering the model selection statistics

Table 1 Type, definition of estimates, and corresponding matrices used during the analyses

Type of estimate	Definition of estimate	Relatedness matrix
h^2	Additive genetic effects due to pedigrees in Gubbio or Ogliastra	Kinship (K)
c^2	Sibs household effect in Gubbio or Ogliastra	Sibs household (C)
h^2_{SNP}	Heritability using SNPs in ARIC or using mixture of pedigrees and SNPs in Ogliastra	Genomic relationship (G_{SNP}) Mixed related (K_{SNP})
h^2_{REL}	Related genomic heritability after selection of elements ≥ 0.025 in ARIC	Genomic related (G_{REL})

Table 2 Six-variance component models (a, b, c, d, e, f) and corresponding estimates (h^2, h^2_{SNP}, c^2) computed for each cohort

Model: $A+C+I$	Estimate	Gubbio	ARIC	Ogliastra
a. $K+I$	h^2	X		X
b. $K+C+I$	h^2, c^2	X		X
c. $G_{SNP}+I$	h^2_{SNP}		X	
d. $G_{REL}+G_{SNP}+I$	$(h^2_{REL}–h^2_{SNP})$, h^2_{SNP}		X	
e. $K+K_{SNP}+I$	$(h^2–h^2_{SNP})$, h^2_{SNP}			X
f. $K+K_{SNP}+C+I$	$(h^2–h^2_{SNP})$, h^2_{SNP}, c^2			X

$(h^2_{REL}–h^2_{SNP})$ or $(h^2–h^2_{SNP})$ refer to the remaining heritability. Cross (X) indicates which model was carried out in each cohort

shown in Table 4, model (b), model (c–d), and model (e) have the best fitting in the Gubbio, ARIC, and Ogliastra populations, respectively. The heritability estimates of each cohort are summarized in Table 5.

In the Gubbio cohort, from a total of 2620 split into 711 nuclear pedigrees, 2147 measured individuals were available for heritability analysis. A significant genetic component ($h^2=0.354$, 95% CI 0.255–0.454) was obtained using kinship matrix (model a). Adding sibs–household component (C), the model showed a small but significant sibs–household effect ($c^2=0.128$, 95% CI 0.03–0.28) and a decrease but still significant heritability equal to 0.286 (95% CI 0.190–0.381).

In the ARIC population, after excluding subjects with missing values, two-variance components models were fitted (model c and d) using 8330 subjects. We first estimated SNP heritability and we obtained statistically significance h^2_{SNP} equal to 0.167 (95% CI 0.091–0.245) (model c). Adding, the G_{REL} matrix of the related subjects (model d), h^2_{SNP} was equal to 0.131 (95% CI 0.045–0.215) and the remaining heritability $(h^2–h^2_{SNP})$ was equal to 0.237 (95% CI 0.031–0.444).

Combining information from pedigrees and SNPs data, we examined MetS score in the Ogliastra population, including 9130 individuals. We concurrently estimated h^2, $h^2–h^2_{SNP}$ and h^2_{SNP} and sibs–household effects fitting four-variance component models at once. As shown in Table 5, h^2 from pedigree data (model a) was significant and equal to 0.343 (95% CI 0.301–0.384). Adding the sibs–household effects (model b), h^2 was significant and equal to 0.271 (95% CI 0.223–0.319) and c^2 equal to 0.114 (95% CI 0.08–0.143). SNP heritability, h^2_{SNP}, was equal to about 0.25 in both models (95% CI 0.093–0.415 and 95% CI 0.101–0.410, respectively, for model e and f), and the remaining heritability $(h^2–h^2_{SNP})$ was equal to 0.098 (95% CI 0.009–0.255) and 0.031 (95% CI 0.003–0.174), respectively. The summed h^2 or total h^2 was equal to 0.352 and 0.276, respectively. Finally, for model (f), also c^2 was estimated ($c^2=0.116$, 95% CI 0.08–0.146).

In short, the actual additive genetic variance proportion is in the range of 0.25. The upper estimates (0.354 in Gubbio and 0.343 in Ogliastra) are biased from shared environmental variance and non-additive genetic variance shared by close relatives. If sib–household effects were considered in the Ogliastra cohort, still-missing heritability has a reduction of 0.07. This residual variance probably includes non-additive genetic variance. Thus, it could be considered part of broad-sense heritability, H^2 rather than h^2.

Finally, the high and significant remaining heritability founded in the ARIC cohort includes non-additive genetic and shared environments. The reduced h^2_{SNP} from 0.17 to 0.13 was due to related subjects in the G_{SNP} matrix.

Discussion

MetS is a complex disease and a knowledge of the underlying mechanisms may contribute to a better understanding of its pathogenesis. The focus of this study was to fill in gaps as far as genetic aspects are concerned using different approaches (Chen et al. 2015; Shetty et al. 2011). In particular, the aim of the paper was to study the heritability and environmental influence in different populations by using full pedigree and/or genetic information. The inclusion in the analyses of cohorts characterized by relatives with distant relationships allowed us to obtain a global vision of genetic effects and a less biased estimation of heritability.

The results of these analyses suggest that additive genetic component (h^2 values ranging from 0.271 to 0.286) is one of the major contributors to phenotypic variation in MetS score, whereas common sibs–household effects have moderate impact on the trait variation (about 11% both for the Gubbio and Ogliastra populations). Overall, heritability from pedigrees, h^2, or related subjects, h^2_{REL}, was significantly higher compared to the heritability from SNPs, h^2_{SNP}, ranging from 0.131 to 0.254.

One reason is that heritability from pedigrees or from related subjects accounts for common as well as rare variants, CNVs, and other structural variations with minor effects and captures information also from variants that are in LD (Vattikuti et al. 2012). Moreover, close relatives, who share SNP genotypes more often than the average, also tend to share a common environment. In particular, if sibling relationship is not considered in the variance component model, heritability could included, as well as some shared environment, some non-additive genetic variance, and non-genetic inherited factors. On the contrary, estimation of h_{SNP} by genomic relationship matrix, G_{SNP}, leaving out closer relatives reduces this type of bias. The reason is to avoid the possibility that the resemblance between close relatives could be due to non-genetic effects (i.e., shared environment). Summarizing, close relatives give more precision but

Table 3 Descriptive features of metabolic syndrome traits and MetS score for the Gubbio, Ogliastra and ARIC samples divided by gender

	Gubbio				Ogliastra				ARIC			
	Men, $N=1852$		Women, $N=2259$		Men, $N=3485$		Women, $N=4617$		Men, $N=4173$		Women, $N=4419$	
	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range
BMI (kg/m^2)	27.15 ± 3.77	13.78 –45.11	26.14 ± 4.83	15.79 –50.37	26.5 ± 3.9	15.8–48.9	25.2 ± 4.9	14.3–50.7	27.42 ± 3.97	16.10–56.26	26.65 ± 5.48	14.91–55.20
Waist circ. (cm)	93.05 ± 11.12	56–140	85.08 ± 13.96	52–135	93.0 ± 10.5	60.2–135.7	84.7 ± 13.4	52.7–146.5	99.60 ± 10.36	66–171	93.14 ± 14.81	52–169
HDL (mg/dl)	49.40 ± 12.47	22–109	61.37 ± 14.40	24–117	47.7 ± 11.5	19.1–121.0	56.6 ± 12.3	17.7–120.4	43.78 ± 12.19	9.63–128.01	57.42 ± 17.01	11.55–134.82
TRG (mg/dl)	139.4 ± 96.46	17–1377	107.4 ± 61.51	23–630	125.6 ± 96.3	21.5–1456.5	94.8 ± 55.8	15.7–870.7	147.30 ± 97.9	24–1876	128.87 ± 82.58	26–1563
Glucose (mg/dl)	94.29 ± 23.30	59–346	89.82 ± 22.08	56–388	99.7 ± 25.0	56.4–435.9	92.6 ± 23.4	54.2–426.9	105.05 ± 30.6	53.88–517.67	102.86 ± 30.85	36.97–446.47
SBP (mmHg)	130.88 ± 15.79	87–221	128.6 ± 19.12	87–217	130.9 ± 16.3	83–220	125.2 ± 18.3	80–200	120.16 ± 16.03	72–206	117.05 ± 17.45	61–203
DBP (mmHg)	77.93 ± 8.92	45–115	75.48 ± 9.53	47–111	82.9 ± 10.0	50–130	79.1 ± 10.2	40–150	73.60 ± 10.01	12–130	69.88 ± 9.66	27–129
Age (years)	53.44 ± 16.08	11–92	55.95 ± 16.57	8–93	49.1 ± 17.5	18–100	49.4 ± 17.7	18–98	54. 69 ± 5.68	44–66	54.01 ± 5.67	44–66
MetS Score	37.26 ± 10.95	0–90	30.59 ± 10.07	5–73	37.02 ± 10.90	6–87	30.09 ± 10.02	5–75	41.07 ± 10.88	9–100	31.88 ± 10.57	7–78

SD standard deviation

Table 4 Linear mixed models (LMMs) output (no. of parameters = no. of fixed effects + no. of random effects; no. of subjects; log-likelihood, AIC and BIC) for the model a, b, c, d, e, f

Cohort	Model	Number of parameters (p)	Number of subjects (n)	Log-likelihood (LL)	AIC $= -2\times LL + 2\times p$	BIC $= -2\times LL + p\times \ln(n)$
Gubbio	0. I	2 + 1	2147	−7630	15,266	15,283
	a. $K + I$	2 + 2	2147	−7602	15,213	15,235
	b. $K + C + I$	2 + 3	2147	−7590	**15,190**	**15,219**
ARIC	0. I	12 + 1	8252	−31,904	63,833	63,924
	c. $G_{SNP} + I$	12 + 2	8252	−31,896	63,820	**63,918**
	d. $G_{REL} + G_{SNP} + I$	12 + 3	8252	−31,894	**63,818**	63,923
Ogliastra	0. I	2 + 1	9130	−33,231	66,468	66,489
	a. $K + I$	2 + 2	9130	−32,986	65,980	66,008
	b. $K + C + I$	2 + 3	9130	−32,945	65,900	65,956
	e. $K + K_{SNP} + I$	2 + 3	9130	−32,976	65,961	65,998
	f. $K + K_{SNP} + C + I$	2 + 4	9130	−32,933	**65,879**	**65,920**

Bold refers to the best fitting model in each cohort

Model 0 refers to a model with only fixed covariates and random residual error

Table 5 Heritability and shared environment variance components (estimate and 95% CI) from the Gubbio, ARIC and Ogliastra cohorts

Cohort	Model	Estimated variance component (95% CI)			
		h^2	h^2–h^2_{SNP}	h^2_{SNP}	c^2
Gubbio	a	0.354 (0.255–0.454)	–	–	–
	b	0.286 (0.190–0.381)	–	–	0.128 (0.030–0.220)
ARIC	c	–	–	0.167 (0.091–0.245)	–
	d	–	0.237 (0.031–0.444)	0.131 (0.045–0.215)	–
Ogliastra	a	0.343 (0.302–0.384)	–	–	–
	b	0.271 (0.223–0.319)	–	–	0.114 (0.080–0.143)
	e	–	0.098 (0.009–0.255)	0.254 (0.093–0.415)	–
	f	–	0.031 (0.003–0.174)	0.245 (0.101–0.390)	0.116 (0.080–0.146)

Model: see Table 2 for model definition; – not computed

potentially more bias, whereas distant relatives give less precision but less bias.

In our case, we add common environmental effect (c^2) using sib relationship to pedigree models to reduce residual variance and adjust the genetic variance. Estimate of heritability from data with many close relatives (our situation for Gubbio and Ogliastra) would be free from environmental bias. Vice versa, the heritability explained by related subjects in the ARIC cohort (including relatives with close relationships) could be overestimated in one-variance component models without leaving out close relatives (model c) and in two-variance component model without sib–household adjustment (model d). In this case, the estimation of the remaining genetic variance (h^2–h^2_{SNP}) could not be free of confounding by environment. This explains why the total estimate heritability in ARIC (model d) is higher than in Gubbio and Ogliastra (0.28) after adjusting for sib–household effect (Table 3).

However, SNPs data have many advantages; for example, genotyping allows empirical estimates of genomic sharing rather than relying on theoretical distributions used in family-based study designs; population-based datasets reduce variability and provide more precise estimates of heritability and, finally, collection of large twin or family-based cohorts is difficult and not cost-effective.

The strengths of this study were the evaluation of the genetic contribution of MetS as a continuous variable, comparison of the results of different populations, and the use of different relatedness information. Presently, in most of the studies from the literature, heritability of MetS has been obtained by estimating each MetS component separately (e.g., BMI, blood pressure, HDL-cholesterol) (Bosy-Westphal et al. 2007). Vattikuti et al. (2012), in particular, using the same ARIC cohort but considering separate traits defining MetS, found heritabilities ranging between 0.23 and 0.48 in related subjects and lower ones, 0.09–0.24, in unrelated subjects. Similarly in the Ogliastra population, heritabilities

estimated for each MetS trait (Biino et al. 2015) ranged between 0.20 and 0.60. In both samples (ARIC and Ogliastra), HDL-cholesterol was the most heritable trait. Similar results were obtained in the Gubbio population (data shown in Table 2S in Online Appendix); HDL-cholesterol heritability for nuclear pedigree was the highest (0.52).

Furthermore, a lot of different criteria to define MetS and estimate genetic component were used. Despite these limitations, a moderate to high heritability has been estimated for all traits with significant differences across age and gender (Teran-Garcia and Bouchard 2007). Other results were found adopting different criteria to define MetS, as in the Jackson Heart Study, where Adult Treatment Panel III (ATPIII criteria) were used (Khan et al. 2015) and in Bellia et al. (2009) who found heritability values ranging from 0.19 to 0.38 using both International Diabetes Federation (IDF) and ATPIII criteria (Henneman et al. 2008).

Briefly, studies have shown that genetic effects influence the variability of MetS and indicate that in representative population-based samples, metabolic syndrome and its components are moderately to highly heritable. We confirmed these results focusing on the MetS score as a continuous variable and considering different relatedness matrices depending on the type of available data.

Additionally, in our study, both Ogliastra and Gubbio are isolated populations with no immigration, thus representing an ideal condition for studying complex diseases because of a reduction in background variability due to unpredictable factors (Varilo and Peltonen 2004; Wright et al. 1999). Also, these cohorts gave us the opportunity to evaluate the environment effects shared among related individuals, in particular we analyzed sibling effects; thus, c^2 could be interpreted as the environment (e.g., habits, diets) shared largely in the past by full siblings (Xia et al. 2016). Moreover, ARIC cohort gave us the possibility to compare results from unrelated large American population and to analyze also related individuals. Finally, our analyses allow us to estimate the remaining heritability, h^2–h^2_{SNP}. Specifically, by using the same data and model (f), "still-missing heritability" or discrepancy between h^2 and h^2_{SNP} could be obtained without environmental effects.

Conversely, one of the major limitations is that the MetS score does not consider new emerging risk factors for MetS, such as the inflammatory state or procoagulant variables, or intermediate phenotypes (Blanco-Gomez et al. 2016) that could interact with the other traits. Also, an LD pruning before estimation of G_{SNP} could be a limitation to estimate SNPs' heritability. As suggested by Yang et al. (2015), REML fitting by LD-MAF multicomponents is unbiased regardless of the MAF and LD properties of causal variants. Improvements can be made taking into account other risk factors, fitting dominance and epistatic models (as broad-sense heritability) using unrelated data (Müller et al. 2013), adding recent shared environment effects using pedigree data (Xia et al. 2016), and fitting LD-MAF multi-component models (Evans et al. 2018; Yang et al. 2015). Finally, more efficient and fast algorithms for LMMs could be developed (Covarrubias-Pazaran 2016; Müller et al. 2013).

Acknowledgements The authors thank the staff and participants of the ARIC, Gubbio, and Ogliastra studies for their important contributions. The ARIC study was carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The Gubbio study was originally funded by Merck Sharp and Dohme, Italy, and the Department of Outcome Research of Merck & Co Inc., USA. Funds were also obtained from Grant # R01HL40397 – 02 of the National Heart Lung and Blood Institute, Bethesda, Maryland, USA, and Ministero Italiano di Università e Ricerca (Grant # 068034, PRIN 2004). The present analyses were done in the context of the Istituto Auxologico Italiano participation to EU MASCARA project/EC-7th Framework Program contract no. 278249. The Ogliastra study was supported by a grant from the Italian Ministry of Education, University and Research: MERIT RBNE08NKH7_007.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Abou Ziki MD, Mani A (2016) Metabolic syndrome: genetic insights into disease pathogenesis. Curr Opin Lipidol 27(2):162–171. https://doi.org/10.1097/MOL.0000000000000276

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62(5):1198–1211. https://doi.org/10.1086/301844

Andreassi MG, Botto N (2003) DNA damage as a new emerging risk factor in atherosclerosis. Trends Cardiovasc Med 13(7):270–275

Bellia A, Giardina E, Lauro D, Tesauro M, Di Fede G, Cusumano G et al (2009) "The Linosa Study": epidemiological and heritability data of the metabolic syndrome in a Caucasian genetic isolate. Nutr Metab Cardiovasc Dis 19(7):455–461. https://doi.org/10.1016/j.numecd.2008.11.002

Bennett RL, Steinhaus KA, Uhrich SB, O'Sullivan CK, Resta RG, Lochner-Doyle D et al (1995) Recommendations for standardized human pedigree nomenclature. Pedigree Standardization Task Force of the National Society of Genetic Counselors. Am J Hum Genet 56(3):745–752

Biino G, Balduini CL, Casula L, Cavallo P, Vaccargiu S, Parracciani D, Pirastu M (2011) Analysis of 12,517 inhabitants of a Sardinian geographic isolate reveals that predispositions to thrombocytopenia and thrombocytosis are inherited traits. Haematologica 96(1):96–101. https://doi.org/10.3324/haematol.2010.029934

Biino G, Concas MP, Cena H, Parracciani D, Vaccargiu S, Cosso M, Pirastu M (2015) Dissecting metabolic syndrome components: data from an epidemiologic survey in a genetic isolate. Springerplus 4:324. https://doi.org/10.1186/s40064-015-1049-9

Blanco-Gomez A, Castillo-Lluva S, Del Mar Saez-Freire M, Hontecillas-Prieto L, Mao JH, Castellanos-Martin A, Perez-Losada J

(2016) Missing heritability of complex diseases: enlightenment by genetic variants from intermediate phenotypes. BioEssays 38(7):664–673. https://doi.org/10.1002/bies.201600084

Bonati MT, Graziano F, Monti MC, Crocamo C, Terradura-Vagnarelli O, Cirillo M, Zanchetti A (2014) Heritability of blood pressure through latent curve trajectories in families from the Gubbio population study. J Hypertens 32(11):2179–2187. https://doi.org/10.1097/hjh.0000000000000311

Bosy-Westphal A, Onur S, Geisler C, Wolf A, Korth O, Pfeuffer M, Muller MJ (2007) Common familial influences on clustering of metabolic syndrome traits with central obesity and insulin resistance: the Kiel obesity prevention study. Int J Obes (Lond) 31(5):784–790. https://doi.org/10.1038/sj.ijo.0803481

Bourrat P, Lu Q (2017) Dissolving the missing heritability problem. Philos Sci 84(5):1055–1067

Chen F, He J, Zhang J, Chen GK, Thomas V, Ambrosone CB, Stram DO (2015) Methodological considerations in estimation of phenotype heritability using genome-wide SNP data, illustrated by an analysis of the heritability of height in a large sample of african ancestry adults. PLoS One 10(6):e0131106. https://doi.org/10.1371/journal.pone.0131106

Cirillo M, Terradura-Vagnarelli O, Mancini M, Menotti A, Zanchetti A, Laurenzi M (2014) Cohort profile: the Gubbio population study. Int J Epidemiol 43(3):713–720. https://doi.org/10.1093/ije/dyt025

Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ et al (2009) Finding the missing heritability of complex diseases. Nature. https://doi.org/10.1038/nature08494

Conomos M, Thornton T, Gogarten S (2017) GENESIS: GENetic EStimation and Inference in Structured samples (GENESIS): statistical methods for analyzing genetic data from samples with population structure and/or relatedness. R package version 2.2. 7. 2017

Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One 11(6):e0156744. https://doi.org/10.1371/journal.pone.0156744

Dandine-Roulland C, Perdry H (2017) Genome-wide data manipulation, association analysis and heritability estimates in R with gaston 1.5. In: Human heredity, vol. 83, Allschwilerstrasse 10, Ch-4009 Basel, Switzerland, Karger, p. 6

Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S et al (2018) Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet 50(5):737–745. https://doi.org/10.1038/s41588-018-0108-x

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51(4):1440–1450

Graziano F, Grassi M, Sacco S, Concas MP, Vaccargiu S, Pirastu M, Biino G (2015) Probing the factor structure of metabolic syndrome in Sardinian genetic isolates. Nutr Metab Cardiovasc Dis 25(6):548–555. https://doi.org/10.1016/j.numecd.2015.02.004

Graziano F, Grassi M, Bonati MT, Zanchetti A, Biino G (2016) External validation of the MetS score, a prediction tool for metabolic syndrome. Nutr Metab Cardiovasc Dis 26(4):359–360. https://doi.org/10.1016/j.numecd.2015.12.014

Henneman P, Aulchenko YS, Frants RR, van Dijk KW, Oostra BA, van Duijn CM (2008) Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: the Erasmus Rucphen Family study. J Med Genet 45(9):572–577. https://doi.org/10.1136/jmg.2008.058388

Hill WG, Maki-Tanila A (2015) Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. J Anim Breed Genet 132(2):176–186. https://doi.org/10.1111/jbg.12140

Investigators A (1989) The atherosclerosis risk in communities (ARIC) study: design and objectives. Am J Epidemiol 129(4):687–702

Khan RJ, Gebreab SY, Sims M, Riestra P, Xu R, Davis SK (2015) Prevalence, associated factors and heritabilities of metabolic syndrome and its individual components in African Americans: the Jackson heart study. BMJ Open 5(10):e008675. https://doi.org/10.1136/bmjopen-2015-008675

Khoury MJ, Beaty TH, Cohen BH (1993) Fundamentals of genetic epidemiology, vol 22. Oxford University Press, Oxford

Lee JJ, Chow CC (2014) Conditions for the validity of SNP-based heritability estimation. Hum Genet 133(8):1011–1022. https://doi.org/10.1007/s00439-014-1441-5

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. J Dairy Sci 92(9):4656–4663. https://doi.org/10.3168/jds.2009-2061

Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, Duarte CW, Allison DB, de los Campos G (2011) Beyond missing heritability: prediction of complex traits. PLoS Genet 7(4): e1002051

Menotti A, Lanti M, Angeletti M, Botta G, Cirillo M, Laurenzi M et al (2009) Twenty-year cardiovascular and all-cause mortality trends and changes in cardiovascular risk factors in Gubbio, Italy: the role of blood pressure changes. J Hypertens 27(2):266–274. https://doi.org/10.1097/HJH.0b013e32831cbb0b

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92(9):4648–4655. https://doi.org/10.3168/jds.2009-2064

Müller S, Scealy JL, Welsh AH (2013) Model selection in linear mixed models. Stat Sci 28(2):135–167

Shetty PB, Qin H, Namkung J, Elston RC, Zhu X (2011) Estimating heritability using family and unrelated individuals data. BMC Proc 5:S34

Sinnwell JP, Therneau TM, Schaid DJ (2014) The kinship2 R package for pedigree data. Hum Hered 78(2):91–93. https://doi.org/10.1159/000363105

Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ (2017) Reevaluation of SNP heritability in complex human traits. Nat Genet. https://doi.org/10.1038/ng.3865

TeamR, RC (2018) A language and environment for statistical computing. 2015. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 26 Nov

Teran-Garcia M, Bouchard C (2007) Genetics of the metabolic syndrome. Appl Physiol Nutr Metab 32(1):89–114. https://doi.org/10.1139/h06-102

Tucker G, Loh PR, MacLeod IM, Hayes BJ, Goddard ME, Berger B, Price AL (2015) Two-variance-component model improves genetic prediction in family datasets. Am J Hum Genet 97(5):677–690. https://doi.org/10.1016/j.ajhg.2015.10.002

Varilo T, Peltonen L (2004) Isolates and their potential use in complex gene mapping efforts. Curr Opin Genet Dev 14(3):316–323. https://doi.org/10.1016/j.gde.2004.04.008

Vattikuti S, Guo J, Chow CC (2012) Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. PLoS Genet 8(3):e1002637

Vinkhuyzen AA, Wray NR, Yang J, Goddard ME, Visscher PM (2013) Estimation and partition of heritability in human populations using whole-genome analysis methods. Annu Rev Genet 47:75–95. https://doi.org/10.1146/annurev-genet-111212-133258

Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, Martin NG (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2(3):e41. https://doi.org/10.1371/journal.pgen.0020041

Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era—concepts and misconceptions. Nat Rev Genet. https://doi.org/10.1038/nrg2322

Witte JS, Visscher PM, Wray NR (2014) The contribution of genetic variants to disease depends on the ruler. Nat Rev Genet 15:765–776

Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. Nat Genet 23(4):397–404. https://doi.org/10.1038/70501

Xia C, Amador C, Huffman J, Trochet H, Campbell A, Porteous D, Haley CS (2016) Pedigree- and SNP-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation. PLoS Genet 12(2):e1005804. https://doi.org/10.1371/journal.pgen.1005804

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet. https://doi.org/10.1038/ng.608

Yang J, Lee SH, Goddard ME, Visscher PM (2011a) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM et al (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43(6):519–525. https://doi.org/10.1038/ng.823

Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, Visscher PM (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet 47(10):1114–1120. https://doi.org/10.1038/ng.3390

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL (2013) Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet 9(5):e1003520

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28(24):3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA 109(4):1193–1198. https://doi.org/10.1073/pnas.1119675109

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Francesca Graziano[1,2] · Ginevra Biino[3] · Maria Teresa Bonati[4] · Benjamin M. Neale[5,6] · Ron Do[7,14] · Maria Pina Concas[13] · Simona Vaccargiu[8] · Mario Pirastu[8] · Oscar Terradura-Vagnarelli[9] · Massimo Cirillo[10] · Martino Laurenzi[9] · Mario Mancini[11] · Alberto Zanchetti[4,12] · Mario Grassi[2]

1 Center of Biostatistics for Clinical Epidemiology, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

2 Medical and Genomic Statistics Unit, Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

3 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy

4 Istituto Auxologico Italiano, IRCCS, Milan, Italy

5 Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA

6 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

7 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA

8 Institute of Genetic and Biomedical Research-Support Unity, National Research Council of Italy, Sassari, Italy

9 Centro di Studi Epidemiologici di Gubbio, Gubbio, Italy

10 Dipartimento di Sanità Pubblica, Università degli Studi di Napoli 'Federico II', Naples, Italy

11 Dipartimento di Medicina Clinica e Sperimentale, Università degli Studi di Napoli 'Federico II', Naples, Italy

12 Centro di Fisiologia Clinica e Ipertensione, Università degli Studi di Milano, Milan, Italy

13 Institute for Maternal and Child Health-IRCCS 'Burlo Garofolo', Trieste, Italy

14 Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, USA