Sex Modifies the Effect of Genetic Risk Scores for Polycystic Ovary Syndrome on Metabolic Phenotypes

Ky’Era V. Actkins¹,², Genevieve Jean-Pierre²,³, Melinda C. Aldrich³,⁴,⁵, Digna R. Velez Edwards⁵,⁶,⁷, and Lea K. Davis²,³

1. Department of Microbiology, Immunology, and Physiology, Meharry Medical College, Nashville, TN
2. Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN
3. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
4. Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN
5. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN
6. Vanderbilt Epidemiology Center, Institute of Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN
7. Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

Short title: Genetic sex effects of polycystic ovary syndrome

Co-Corresponding authors
Lea Davis
Associate Professor of Medicine
Vanderbilt Genetics Institute
Division of Genetic Medicine
Department of Medicine
511-A Light Hall
Vanderbilt University
2215 Garland Ave
Nashville, TN 37232
Office Tel: 615-875-9478

Ky’Era Actkins
PhD Candidate
Meharry Medical College
1005 Dr. DB Todd Blvd
Nashville, TN 37208

Keywords: polycystic ovary syndrome, sex differences, polygenic risk scores, phenome-wide association study, electronic health records, mediation analysis

Funding/Support: KVA is funded by F31HD103397. LKD is funded by U54MD010722. MCA is funded by a grant from NIH/NCI (U01CA253560). Datasets were obtained using the Vanderbilt

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
University Advanced Computing Center for Research and Education. This resource is funded by institutional, private, and federal grants, which include NIH funded Shared Instrumentation Grant S10OD017985 and S10RR025141 and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975.

Author contributions: LKD and KVA designed the study. KVA performed the statistical analyses. All authors contributed to the interpretation of the results, the revision of the manuscript, and approved the final manuscript.

Competing interests: The authors declare they have no competing interests.

Data availability statement: All GWAS summary statistics used in this study are publicly available and can be obtained from the following consortium groups and websites: International PCOS Consortium (https://doi.org/10.17863/CAM.27720), UK Biobank (http://www.nealelab.is/uk-biobank), HERMES (http://www.broadcvdi.org/), and CARDIoGRAMplusC4D (http://www.cardiogramplusc4d.org/). The summary statistics used from the Million Veteran Project can be accessed through dbGaP under accession number phs001672.v6.p1. The electronic health record data that support the findings of this study are available from Vanderbilt University Medical Center, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of Vanderbilt University Medical Center.
Abstract

Females with polycystic ovary syndrome (PCOS), the most common endocrine disorder in women, have an increased risk of developing metabolic disorders such as insulin resistance, obesity, and type 2 diabetes (T2D). Furthermore, while only diagnosable in females, males with a family history of PCOS can also exhibit a poor cardiometabolic profile. Therefore, we aimed to elucidate the role of sex in the relationship between PCOS and its comorbidities by conducting bidirectional genetic risk score analyses in both sexes. We conducted a phenome-wide association study (PheWAS) using PCOS polygenic risk scores (PCOS\textsubscript{PRS}) to understand the pleiotropic effects of PCOS genetic liability across 1,380 medical conditions in females and males recorded in the Vanderbilt University Medical Center electronic health record (EHR) database. After adjusting for age and genetic ancestry, we found that European descent males with higher PCOS\textsubscript{PRS} were significantly more likely to develop cardiovascular diseases than females at the same level of genetic risk, while females had a higher odds of developing T2D. Based on observed significant associations, we tested the relationship between PRS for comorbid conditions (e.g., T2D, body mass index, hypertension, etc.) and found that only PRS generated for BMI and T2D were associated with a PCOS diagnosis. We then further decomposed the T2D\textsubscript{PRS} association with PCOS by adjusting the model for measured BMI and BMI\textsubscript{residual} (enriched for the environmental contribution to BMI). Results demonstrated that genetically regulated BMI primarily accounted for the relationship between T2D\textsubscript{PRS} and PCOS. This was further supported in a mediation analysis, which only revealed clinical BMI measurements, but not BMI\textsubscript{residual}, as a strong mediator for both sexes. Overall, our findings show that the genetic architecture of PCOS has distinct metabolic sex differences, but these associations are only apparent when PCOS\textsubscript{PRS} is explicitly modeled. It is possible that these pathways are less explained by the direct genetic risk of metabolic traits than they are by the risk factors shared between them, which can be influenced by biological variables such as sex.
Introduction

Polycystic ovary syndrome (PCOS) is a highly heritable endocrine disorder that affects 5%-21% of females of reproductive age who are typically diagnosed by having two or more of the following features under the Rotterdam criteria: polycystic ovaries, oligo- and anovulation, or hyperandrogenism [1–3]. Although Rotterdam is the most common PCOS criteria, other criteria can be used for diagnosis, including the National Institutes of Health criteria, the Androgen Excess and PCOS Criteria, or the 2018 International Evidence Based PCOS guidelines [2]. Each criterion slightly differs in requirements in an effort to cover the range of PCOS symptoms that are exhibited in patients. However, as a result of the differing diagnostic criteria, the heterogenous presentation of symptoms, and the prevalent comorbidities that reside outside of diagnostic requirements patients often spend years seeking a diagnosis or worse, may be one of the 75% of females estimated to be undiagnosed [4,5].

Many clinicians select criteria based on their perception of the most defining PCOS feature [6]. In some cases, as with the Androgen Excess and PCOS Society criteria [7], this will mean hyperandrogenism, a symptom that typically manifests as acne, hirsutism, or alopecia [8]. Androgen excess is also hypothesized to underlie many of the comorbid metabolic dysfunctions experienced by patients such as insulin resistance, obesity, metabolic syndrome, type 2 diabetes, and cardiovascular diseases (CVD). However, as previous studies have shown, the genetic risk factors present in patients with metabolic manifestations of PCOS could differ from others who have primary presentations of reproductive dysfunction [9,10].

PCOS is multifactorial and twin studies estimate heritability at 70% [11–14]. With an underlying polygenic architecture, multiple variants are hypothesized to be involved in the development of PCOS [15,16]. Furthermore, the variation of clinical features can be partially explained by ancestry informative markers, indicating potential population specific effects [17,18]. Rare variants in genes, such as DENND1A, have also been identified in family studies for PCOS alongside many other genetic variants identified from GWAS [19]. Despite the small
effect size of individual common variants, aggregation of common risk variants together as a polygenic (or genetic) risk score (PRS) reflects the overall additive genetic liability to PCOS in individuals. This marker of disease risk is associated with PCOS diagnosis in multiple ancestries and offers many advantages to parsing out the genetic etiology of PCOS that is entangled with its comorbid presentations [11–13]. Furthermore, there is increasing evidence that a spectrum of clinical PCOS manifestations for PCOS is also correlated with the PRS for PCOS [11]. Therefore, in this study, we aimed to determine if the PRS for PCOS demonstrated pleiotropic associations with other health conditions in a hospital biobank population through a phenome-wide association study (PheWAS). By using a genetic model, we were able to determine sex-differentiated effects associated with PCOS genetic risk, revealing the impact of PCOS-associated inherited genetic variation in males, despite the fact that PCOS is only diagnosed in females. We further implemented several analyses to determine the mediating role of body mass index (BMI) on cardiometabolic comorbidities, and further identified the directionality of their effects through causal mediation analyses.

Methods

VUMC EHR-linked Biorepository

Vanderbilt University Medical Center (VUMC) is a tertiary care hospital in Nashville, Tennessee, with several outpatient clinics throughout Tennessee and the surrounding states offering primary and secondary care. Medical records have been electronically documented at VUMC since the early 1990s, resulting in a clinical research database of over 3 million EHRs referred to as the Synthetic Derivative [20]. EHRs include demographic information, health information documented through International Classification of Disease, Ninth Revision (ICD9) and Tenth Revision (ICD10) codes, procedural codes (CPT), clinical notes, medications, and laboratory values. This information is linked with a DNA biorepository known as BioVU. Use of
EHRs and genetic data for this study was approved by the Vanderbilt University Institutional Review Board (IRB #160279).

**Genetic data**

BioVU contains 94,474 individuals genotyped on the MEGA\textsuperscript{EX} platform [21]. The QC pipeline removed SNPs with low genotyping call rate (< 0.98) and individual subjects who were related (\(\hat{r} > 0.2\)), had low call rates (< 0.98), sex discrepancies, and excessive heterozygosity (\(F_{het} > 0.2\)). A principal component (PC) analysis (PCA) was performed on remaining individuals to determine genetic ancestry using FlashPCA\textsuperscript{2} [22]. BioVU genotyped samples were stratified by ancestral origin based on PCs herein referred to as the European (EUR) descent or African (AFR) descent dataset. Extended details for the quality control of these datasets have been described previously [23].

**Publicly available summary statistics**

Genome-wide association study (GWAS) summary statistics were acquired for PCOS, BMI, diastolic blood pressure, systolic blood pressure, pulse, type 2 diabetes, heart failure, and coronary artery disease (i.e., all of the significant phenotypic associations observed in the PheWAS models) to further establish the strength of associations between PCOS and its comorbidities through subsequent analyses. Each GWAS was selected based on public availability, sample size, and sample diversity.

The Genetic Investigation of ANthropometric Traits (GIANT) consortium body mass index (BMI) summary statistics included 339,224 individuals of European and non-European descent from over 125 studies. These BMI summary statistics were used as a proxy for obesity summary statistics [24]. Blood pressure traits (diastolic, systolic, and pulse) and type 2 diabetes (T2D) summary statistics were obtained from the Million Veterans Program (MVP), a large biobank consortium effort that houses biobank data from various sites in the Department of
Veterans Affairs health system [25]. Blood pressure traits were generated from a trans-ethnic sample of over 750,000 individuals from MVP [26]. T2D summary statistics were generated from a meta-analysis using data from 1.4 million participants in various biobanks and consortia groups [27]. Heart failure summary statistics were collected from 47,309 cases and 930,014 controls of European ancestry across nine studies in the Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES) consortium as a proxy for heart disease [28]. Finally, coronary artery disease (CAD) datasets generated from the Coronary Artery Disease Genome-wide Replication and Meta-analysis plus The Coronary Artery Disease (CARDIoGRAMplusC4D) consortium were used as the genetic measurement for the coronary atherosclerosis phenotype [29]. This meta-analysis assembled 60,801 cases and 123,504 controls of multiple ancestries across forty-eight study sites.

Statistical analysis

Generation of polygenic risk scores (PRS)

PCOS\textsubscript{PRS} were calculated with PRS-CS software using the weighted sums of the risk allele effects as reported in the summary statistics from the Day et al. GWAS of PCOS and applying a Bayesian continuous shrinkage parameter select SNP features and to model linkage disequilibrium [16,30]. The details of these methods have been previously described elsewhere [11]. We calculated PCOS\textsubscript{PRS} for both EUR and AFR BioVU genotyped ancestry samples which were previously shown to be associated with a PCOS diagnosis defined by a coded strict PCOS definition in our previously published EHR-based algorithm [11].

Phenome-wide association study (PheWAS)

Next, we were interested in identifying the pleiotropic effects of the genetic susceptibility to PCOS on the medical phenome. Therefore, in a PheWAS framework we analyzed the effects of PCOS\textsubscript{PRS} across 1,380 medical conditions. This analysis was first performed in a female
sample for our EUR and AFR ancestry datasets to validate the PRS and identify potential pleiotropic relationships (Supplementary Figures 1-5). Although males cannot be diagnosed with PCOS, they still harbor genetic risk for PCOS. Thus, we extended the analysis to males to examine sex differences. Finally, we performed a sex-combined PheWAS to increase statistical power. In the sex-combined logistic regression model, covariates included median age of individuals medical record, sex, and the first ten principal components. In the sex-stratified models, covariates included median age and the first ten PCs.

Interaction analysis

For each phenotype with evidence of a significant main effect of PCOS_{PRS} in either sex, we tested for two-way interactions (sex * PCOS_{PRS}) to determine which of the significant sex stratified PheWAS associations were influenced by biological effects of sex. Selected phenotypes of interest that met the phenome-wide false discovery rate (q < 0.05) were tested for interactions (N = 10). For these interaction analyses, sex-combined models included the main effects of PCOS_{PRS}, median age of individuals medical record, sex, and the top ten PCs.

Sensitivity analyses

Several sensitivity analyses were performed to assess the robustness of the significant phenome-wide findings. First, BMI is strongly correlated with both PCOS and its comorbidities, and thus, can influence the strength of the results [31]. Therefore, we adjusted each model for BMI (median measurement across an individual’s EHR) to observe whether the significant phenotypes associated with PCOS_{PRS} were independent of obesity-related effects. In addition to adjusting for BMI, the model was adjusted for median age, sex (only in the sex-combined sample), and the top ten PCs.

Next, we evaluated which phenotypes were dependent on a PCOS diagnosis in females by accounting for PCOS diagnosis [11]. This analysis allowed us to identify true pleiotropic
associations and provided insight into which phenotypes were exclusively correlated with genetic risk, even in the absence of PCOS.

Finally, given that BMI has strong contributions from both genetic and environmental sources of variance, we revisited the previous PheWAS analyses to specifically account for the environmental contribution of BMI. To do this, we calculated the residuals of median BMI adjusted for BMI_{PRS} \text{(residuals(medianBMI \sim BMI_{PRS})}, herein referred to as BMI_{residual}. BMI_{residual} were then used in a subsequent sensitivity analysis of the previously described PheWAS.

Genetic correlation

Linkage Disequilibrium Score Regression (LDSC) was used to calculate genetic correlation, an estimate of genetic similarity, between traits [32]. LDSC only utilizes GWAS summary statistics and is not sensitive to sample overlap, which may be present across the publicly available GWAS datasets used in this study. This method utilizes the effect estimate of each SNP and accounts for the effects of SNPs in linkage disequilibrium based on the GWAS reference population. The European reference panel was used for all analyses based on the demographic majority of each of the GWAS samples.

Logistic regression models

Public GWAS datasets were used to generate PRS for phenome-wide significant phenotypes identified in the PheWAS. PRS were then used as the independent variable in a logistic regression model against PCOS diagnosis as the dependent variable. BioVU genotyped datasets were filtered to females and contained 361 PCOS cases and 29,035 controls in the EUR dataset and 189 PCOS cases and 9,229 controls in the AFR dataset. Selection of PCOS cases and controls have been described elsewhere [11]. In brief, cases required PCOS billing codes and no exclusion codes. Controls excluded individuals who had any inclusion or exclusion codes [11]. All models were adjusted for median age of the individuals medical record,
and the top ten PCs for each ancestry. Median BMI was included as a covariate in the sensitivity analysis.

Lastly, we adjusted for BMI\textsubscript{residual} in the logistic regression models in place of BMI. Next, we performed a sensitivity analysis to test the effect of BMI\textsubscript{residual} and BMI\textsubscript{PRS} separately on PCOS diagnosis. A multiple testing correction of p < 7.35e-04 (0.05/68) was implemented to account for all statistical tests. All statistical analyses were done using R 3.6.0.

**Mediation analysis**

To test whether BMI or BMI\textsubscript{residual} mediates the pleiotropic relationships between PCOS and associated cardiometabolic conditions, a mediation analysis was employed using the mediation R package for both AFR and EUR genotyped samples [33]. For the first set of mediation analyses, we modeled BMI\textsubscript{PRS} and BMI\textsubscript{residual} as exposures, T2D diagnosis as the mediator, and PCOS case status as the outcome. Another reciprocal model was tested with the same exposures, but with PCOS diagnosis as the mediator and T2D as the outcome. This analysis tested the hypothesis that the risk conferred by BMI on one phenotype (e.g., PCOS) in turn increases risk for another other phenotype (e.g., T2D). We restricted this analysis to the female population. All models were further adjusted for median age and the top ten PCs.

In our second set of mediation analyses, we tested whether genetic risk for PCOS or T2D also influenced BMI\textsubscript{residual} through the manifestation of either clinical condition. For example, we modeled PCOSPRS as the exposure variable, T2D diagnosis as the mediator and BMI\textsubscript{residual} as the outcome. Finally, in a female only analysis, T2DPRS was modeled the exposure variable, PCOS diagnosis was included as the mediator, and BMI\textsubscript{residual} again was modeled as the outcome.

Finally, for our last set of mediation analyses, we examined the mediating effect of both measured BMI and BMI\textsubscript{residual} on the relationship between PCOS\textsubscript{PRS} (exposure) and the diagnosis of hypertension and hypertensive heart disease (i.e., outcomes that demonstrated
evidence of significant interactions with sex). T2D was also tested because it was significantly associated with PCOSPRS in both sexes. PCOSPRS was modeled as the exposure variable and T2D (phecode = 270.2) hypertension (phecode = 411), and hypertensive heart disease (phecode = 401.21) were modeled as the outcome variables. Models were considered significant when the average direct effect (ADE) and average causal mediation effect (ACME) both passed \( p < 7.35 \times 10^{-04} \). All mediation analyses were performed with 1,000 bootstrap simulations to estimate the confidence intervals and determine an empirical \( p \)-value.

**Results**

**PCOSPRS PheWAS results**

The PCOSPRS PheWAS was applied to 66,903 EUR individuals from BioVU. When restricted to females (N = 37,240), two phenotypes were significantly associated with PCOSPRS (Bonferroni corrected \( p \)-value = 4.56e-05) (**Figure 1a**): T2D (odds ratio [OR] = 1.10, 95% confidence interval [CI] = 1.06-1.14, \( p = 5.54 \times 10^{-07} \)) and diabetes mellitus (OR = 1.09, 95% CI = 1.05-1.13, \( p = 2.20 \times 10^{-06} \)). When the analysis was restricted to males (**Figure 1b**), hypertensive heart disease was significant (OR = 1.15, 95% CI = 1.08-1.23, \( p = 2.07 \times 10^{-05} \)) alongside a cluster of nominally (\( p < 0.05 \)) associated cardiovascular phenotypes with similar effect sizes (OR = 1.06 to 1.08). This included hypertension (OR =1.06, 95% CI = 1.03-1.10, \( p=1.30 \times 10^{-04} \)), essential hypertension (OR = 1.06, 95% CI = 1.03-1.10, \( p = 2.49 \times 10^{-04} \)), and coronary atherosclerosis (OR = 1.06, 95% CI =1.02-1.10, \( p=1.40 \times 10^{-03} \)). T2D (OR=1.07, 95% CI =1.03-1.11, \( p=7.04 \times 10^{-04} \)) and several other endocrine phenotypes were also modestly associated with PCOSPRS in males.

In the EUR sex-combined model, four phenotypes were significantly associated with PCOSPRS (**Figure 2**). This included T2D and diabetes mellitus as the top associations with the same OR of 1.08 (T2D 95% CI = 1.06-1.11, \( p = 3.15 \times 10^{-09} \); diabetes mellitus 95% CI = 1.05-1.11,
p = 1.03e-08). Following was obesity (OR = 1.07, 95% CI = 1.04-1.11, p = 9.68e-06) and hypertensive heart disease (OR = 1.11, 95% CI = 1.06-1.16, p = 3.57e-05). Three cardiovascular diseases were also nominally associated with PCOSPRS, presumably due to the addition of males. PCOSPRS yielded an OR of 1.21 (95% CI = 1.08-1.36, p = 7.82e-04) for an association with polycystic ovaries, falling just short of phenome-wide significance.

No associations passed Bonferroni correction in the AFR ancestry analysis, which included a total of 12,383 individuals (Supplementary Figure 1).

**Interaction analysis**

The sex interaction analysis demonstrated that males with a high PCOSPRS were (interaction p = 7.24e-03) more likely to be diagnosed with hypertension (ORMales = 1.06, 95% CI = 1.03-1.10, p = 1.30e-04; ORFemales = 1.03, 95% CI = 1.00-1.06, p = 0.07), essential hypertension (interaction p = 7.71e-03, ORMales = 1.06, 95% CI = 1.03-1.10, p = 2.49e-04; ORFemales = 1.03, 95% CI = 1.00-1.06, p = 0.08), and hypertensive heart disease (interaction p = 0.01, ORMales = 1.15, 95% CI = 1.08-1.23, p = 2.07e-05; ORFemales = 1.05, 95% CI = 0.98-1.13, p = 0.16) than females with the same PCOSPRS (Table 1). These sex differences were also observed when calculating the prevalence for each trait by decile of PCOSPRS, which only showed an upwards trend across deciles for traits with a significant sex effect (Figure 3). Together, these results showed that sex is an important modifier of PCOS genetic risk.

**Sensitivity analyses adjusting for BMI and PCOS case status**

In our first set of PheWAS sensitivity analyses for the EUR population, we adjusted for median BMI in all of the PRS-PheWAS models. There were no surviving associations in the sex-combined model or stratified analyses, suggesting BMI may mediate the pleiotropic effects of PCOSPRS (Supplementary Figures 2a and 3). In a separate sensitivity analysis (female only) we found that after adjusting for PCOS diagnosis, females with a high PCOSPRS still
demonstrated a significant positive phenotype-wide association with T2D and diabetes mellitus (Supplementary Figure 2b).

Genetic correlation results

We found that BMI ($r_g = 43.32\%$, $p = 8.77\times10^{-19}$) and T2D ($r_g = 29.26\%$, $p = 3.25\times10^{-09}$) had the strongest genetic correlation with PCOS (Table 2). Heart failure ($r_g = 26.51\%$, $p = 0.0025$) and pulse pressure ($r_g = 13.46\%$, $p = 0.011$) were also modestly significantly with PCOS. CAD bordered the significance threshold, but systolic and diastolic blood pressure were not significantly genetically correlated with PCOS.

Cardiometabolic PRS Analysis of PCOS diagnosis

Outside of T2D and BMI, none of the PRS built for CAD, heart failure, or blood pressure were significantly associated with a PCOS diagnosis in females of either EUR or AFR descent (Figure 4).

BMI_{PRS} was positively associated with PCOS diagnosis for both EUR and AFR ancestry populations ($p_{EUR} = 3.71\times10^{-08}$, $p_{AFR} = 0.02$), but not independently of clinically measured BMI values. T2D_{PRS} in the EUR dataset yielded similar results with PCOS diagnosis, again losing significance upon addition of a BMI covariate in the model ($\text{OR}_{\text{unadjusted}} = 1.16$, 95% CI = 1.04-1.30, $p = 0.007$; $\text{OR}_{\text{adjusted}} = 1.09$, 95% CI = 0.97-1.22, $p = 0.16$). To determine whether this reduction in effect size was due to the genetic correlation between T2D and BMI, we also tested a model in which T2D_{PRS} was adjusted for BMI_{residual} (i.e., variance remaining in BMI after removing variance due to BMI_{PRS}) instead of BMI. This model indeed recovered the original association between T2D_{PRS} ($\text{OR}_{\text{EUR}} = 1.18$, 95% CI = 1.05-1.32, $p_{\text{EUR}} = 0.005$) and PCOS diagnosis for the EUR population, suggesting that genetically predicted BMI mediates the association between T2D_{PRS} and PCOS diagnosis (Figure 5).
Mediation analysis in females with BMI as an exposure variable

To further quantify the degree to which BMI influenced the shared genetic pathways between PCOS and T2D, a mediation analysis was performed with BMI_PRS and BMI_residual as the exposure variables, PCOS diagnosis as the outcome variable, and T2D as the mediator. The reciprocal model was then tested with T2D diagnosis as the outcome and PCOS as the mediator (Table 3). We found that both phenotypes acted as significant mediators when BMI_PRS and BMI_residual were exposure variables. PCOS diagnosis had a stronger mediating effect on T2D when BMI_residual was the exposure variable compared to BMI_PRS (13.6% vs 7.1%). However, the opposite was true for T2D, which mediated more of risk conferred by BMI_PRS than BMI_residual on the PCOS outcome (9% vs 2.1%).

PheWAS sensitivity analysis with BMI_residual

We revisited the PheWAS analysis in light of the findings for BMI_residual. By covarying for BMI_residual, (e.g., environmental estimate of BMI), we observed almost no difference from the original results in our female sample (Supplementary Figure 4a). T2D remained significant in females (OR =1.10, 95% CI = 1.06-1.15, p = 1.69e-06), as did diabetes mellitus (OR =1.09, 95% CI = 1.05-1.13, p = 9.91e-06). These associations also appeared in the combined dataset alongside obesity (Supplementary Figure 5). Although none of the phenotypes passed Bonferroni correction for males, many of the associations improved in estimate effects (Supplementary Figure 4b).

Mediation analysis in the sex-combined and sex-stratified samples

Upon finding BMI_residual as a significant risk exposure for T2D and PCOS in the first mediation analysis, but not a strong confounder in the PCOSPRS PheWAS analysis, we tested whether the genetic risk conferred by one condition (e.g., T2DPRS) was mediated by the clinical manifestation of the other (e.g., PCOS diagnosis) when examining BMI_residual as the outcome.
Here we observed a modest association that showed PCOS diagnosis mediated 33% of the risk conferred by $T2D_{PRS}$ on $BMI_{residual}$ whereas T2D diagnosis did not explain any of the variance in $BMI_{residual}$ due to PCOS$_{PRS}$ (Table 4).

Lastly, the effects of BMI on PCOS comorbidities T2D, hypertension, and hypertensive heart disease were explored (Table 5). We found that $BMI_{residual}$ was not a significant mediator for any of the tested models, even though clinical BMI was a significant mediator for T2D in females and a significant mediator for hypertensive heart disease in males. This again implicated genetically regulated BMI as the main source of the mediating effect for males since median BMI significantly mediated 14.8% of the variance in hypertensive heart disease ($p = 2e^{-16}$) and mediated 23.7% of the variance in hypertension ($p = 0.002$) due to PCOS$_{PRS}$. The lack of an association with $BMI_{residual}$ in females also suggests genetically regulated BMI as the primary mediator for T2D, where it mediated 31.5% of the variance explained by the PCOS$_{PRS}$.

### Discussion

Through a comprehensive analysis of PCOS genetic risk across multiple phenotypes, we identified sex differences in the cardiometabolic traits associated with PCOS genetic risk. Among these, the most notable difference was that males with high PCOS$_{PRS}$ were at greater risk of cardiovascular conditions than females at the same level of genetic risk. Genetically predicted BMI was revealed as a primary mediator of this risk in males, whereas both environmental and genetic BMI variance components mediated risk of T2D conferred by PCOS$_{PRS}$ in females. Furthermore, only $T2D_{PRS}$ and $BMI_{PRS}$ were associated with PCOS diagnosis, indicating that many of the associations observed in the PCOS$_{PRS}$ PheWAS were primarily driven by PCOS genetic risk and not the genetic effects of the identified comorbidities.

There is growing interest in studying PCOS related effects in males whether through their relationship to first-degree family members with PCOS or by an equivalent phenotype [34–36]. Generally, males with mothers or sisters with PCOS tend to exhibit a poorer
cardiometabolic profile which can be observed as early as infancy [35]. A previous study suggests that males who exhibit a high genetic risk for PCOS are more likely to present with morbid obesity, T2D, and diabetes mellitus, two findings which were confirmed in this study [12]. Although we also found associations with CVD phenotypes in males, these were largely influenced by genetically predicted BMI which can significantly increase the lifetime risk for CVD and mortality rates of high-risk individuals [37]. Genetic susceptibility for PCOS could be an additional catalyst for these events in males, making individuals already predisposed to adverse metabolic outcomes more vulnerable.

In an effort to determine whether genetic risk for phenotypes comorbid with PCOS could also increase risk for PCOS, we conducted separate multivariable logistic regressions with PRS for BMI, diastolic blood pressure, systolic blood pressure, pulse pressure, T2D, heart failure, and CAD on the PCOS diagnosis outcome and found no significant associations outside of T2D_{PRS} and BMI_{PRS}. This was unsurprising as the relationship between PCOS and CVD are still poorly elucidated and often debated [38]. It is possible that CVD is more prevalent in genetically high-risk males than females. The sex-difference in CVD prevalence may also be reflected in the PRS which may fail to fully capture CVD genetic risk for PCOS when stratified by sex due to fewer females in the discovery sample.

This study provides novel insight into PCOS genetic etiology and continues to underscore the importance of BMI in PCOS risk. As with T2D, we found that genetically predicted BMI is primarily responsible for the phenotypic association between T2D_{PRS} and PCOS. However, PCOS_{PRS} genetic risk led to more comorbid traits in the PheWAS in which both environmental and genetically predicted BMI influenced the metabolic associations. This may mean that PCOS patients with a family history of T2D can have an increased risk for morbid outcomes, as females with PCOS are already more likely to develop T2D at an earlier age [39]. This genetic susceptibility could also explain the high prevalence of insulin resistance in PCOS patients, which can be as high as 70% across all BMI strata [40]. These effects should
be investigated further, as the genetic and biological pathways could differ in lean PCOS patients who also experience a high rate of insulin resistance [41].

This study offers many strengths. Firstly, we showed that cardiometabolic associations vary with sex and that the metabolic outcomes related to PCOS genetic architecture can be further understood by studying both males and females. Secondly, we decomposed BMI into genetically predicted (i.e., BMI_{PRS}) and environmentally enriched (i.e., BMI_{residual}) and evaluated their respective roles in mediating the cardiometabolic profiles associated with PCOS_{PRS}.

However, limitations include low power to detect any significant associations in our African descent sample. To date, there is no PCOS GWAS of African descent individuals, limiting all current similar studies to building PRS using European-based genetic variants, which do not perform as well in non-European populations [11,12]. Second, we only examined one environmental risk factor. Although many effects such as lifestyle and diet can be captured through BMI, it is not an exhaustive measurement nor does it accurately account for the full wellness of an individual [42,43]. Although other anthropometric features like hip-to-waist ratio (WHR) may be better indicators of health for some phenotypes [44], this information is not routinely collected in clinical settings or reported in EHRs. Furthermore, evidence does suggest that clinically ascertained BMI may be more informative for PCOS than WHR [45]. Finally, despite using the largest PCOS GWAS to date for this analysis, our PCOS_{PRS} still only explains a small portion of PCOS genetic variance. As these analyses expand, so too will our ability to detect the full genetic spectrum of PCOS and its subphenotypes.

PCOS is a multifaceted disorder with genetic architecture that is reflective of its heterogeneous outcomes. This polygenic structure captures a spectrum of metabolic comorbidities that is even more apparent when compared between sexes. Our findings show that males with high PCOS liability are indeed a high-risk group and those with a family history of PCOS should be closely monitored for hypertension and CVD. This is also true for females with PCOS and a family history of T2D, whose genetic risk could induce more severe comorbid
outcomes. As such, management and screening strategies should be updated to reflect advances in PCOS etiology. This call to action is paramount and requires both widespread dissemination of risk factor information to relevant stakeholders and increases in PCOS research priorities and funding. This becomes even more crucial as PCOS comorbidities are often under-recognized in clinical settings and metabolic conditions are underutilized in PCOS screening methods [46–49].
References


doi:10.1210/clinem/dgz326

doi:10.1093/humrep/dev282

doi:10.1210/jc.2005-1494


doi:10.1371/journal.pgen.1007813


doi:10.1016/j.fertnstert.2016.08.033


doi:10.1038/nature14177


43. Chrysant SG, Chrysant GS. The single use of body mass index for the obesity paradox is misleading and should be used in conjunction with other obesity indices. Postgrad Med. 2019;131: 96–102. doi:10.1080/00325481.2019.1568019


**Table 1. Hypertension is Significantly Modified by Sex for PCOS<sub>PRS</sub>**

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Estimate</th>
<th>SE</th>
<th>Z-Value</th>
<th>P</th>
<th>OR</th>
<th>LCI</th>
<th>UCI</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>-0.05</td>
<td>0.02</td>
<td>-2.69</td>
<td>7.24E-03</td>
<td>0.95</td>
<td>0.91</td>
<td>0.99</td>
<td>24,574</td>
<td>35,352</td>
</tr>
<tr>
<td>Essential hypertension</td>
<td>-0.05</td>
<td>0.02</td>
<td>-2.66</td>
<td>7.71E-03</td>
<td>0.95</td>
<td>0.91</td>
<td>0.99</td>
<td>24,159</td>
<td>35,352</td>
</tr>
<tr>
<td>Hypertensive heart disease</td>
<td>-0.11</td>
<td>0.04</td>
<td>-2.51</td>
<td>0.01</td>
<td>0.89</td>
<td>0.82</td>
<td>0.98</td>
<td>2,696</td>
<td>35,352</td>
</tr>
<tr>
<td>Coronary atherosclerosis</td>
<td>-0.04</td>
<td>0.03</td>
<td>-1.46</td>
<td>0.14</td>
<td>0.96</td>
<td>0.91</td>
<td>1.01</td>
<td>9,123</td>
<td>49,700</td>
</tr>
<tr>
<td>Ischemic Heart Disease</td>
<td>-0.03</td>
<td>0.03</td>
<td>-1.28</td>
<td>0.20</td>
<td>0.97</td>
<td>0.92</td>
<td>1.02</td>
<td>10,015</td>
<td>49,700</td>
</tr>
<tr>
<td>Loss of teeth or edentulism</td>
<td>-0.17</td>
<td>0.14</td>
<td>-1.21</td>
<td>0.23</td>
<td>0.84</td>
<td>0.64</td>
<td>1.11</td>
<td>184</td>
<td>61,061</td>
</tr>
<tr>
<td>Obesity</td>
<td>0.03</td>
<td>0.03</td>
<td>0.92</td>
<td>0.36</td>
<td>1.03</td>
<td>0.97</td>
<td>1.09</td>
<td>5,835</td>
<td>51,152</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>0.02</td>
<td>0.02</td>
<td>0.90</td>
<td>0.37</td>
<td>1.02</td>
<td>0.97</td>
<td>1.07</td>
<td>8,990</td>
<td>51,286</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>0.02</td>
<td>0.02</td>
<td>0.65</td>
<td>0.52</td>
<td>1.02</td>
<td>0.97</td>
<td>1.06</td>
<td>9,424</td>
<td>51,286</td>
</tr>
<tr>
<td>Overweight, obesity and other hyperalimentation</td>
<td>0.01</td>
<td>0.03</td>
<td>0.23</td>
<td>0.82</td>
<td>1.01</td>
<td>0.95</td>
<td>1.06</td>
<td>6,841</td>
<td>51,152</td>
</tr>
</tbody>
</table>

We used a logistic regression model to test the interaction of effect between PCOS<sub>PRS</sub> and sex. Models were adjusted for age, sex, genetic ancestry, and the interaction term. Estimate effects are in the direction of females.
Table 2. PCOS Shares Genetic Architecture with Cardiometabolic Traits

<table>
<thead>
<tr>
<th>Traits</th>
<th>Genetic Correlation</th>
<th>SE</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diastolic Blood Pressure</td>
<td>-3.60%</td>
<td>0.07</td>
<td>-0.54</td>
<td>0.59</td>
</tr>
<tr>
<td>Systolic Blood Pressure</td>
<td>9.33%</td>
<td>0.05</td>
<td>1.79</td>
<td>0.07</td>
</tr>
<tr>
<td>Pulse Pressure</td>
<td>13.46%</td>
<td>0.05</td>
<td>2.54</td>
<td>0.01</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>29.26%</td>
<td>0.05</td>
<td>5.92</td>
<td>3.25e-09</td>
</tr>
<tr>
<td>Heart Failure</td>
<td>26.51%</td>
<td>0.09</td>
<td>3.02</td>
<td>0.003</td>
</tr>
<tr>
<td>Coronary Artery Diseases</td>
<td>17.35%</td>
<td>0.09</td>
<td>1.95</td>
<td>0.05</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>43.32%</td>
<td>0.05</td>
<td>8.85</td>
<td>8.77e-19</td>
</tr>
</tbody>
</table>
**Table 3. BMI Genetic Variance is a Causal Risk Factor for PCOS and T2D**

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Mediator</th>
<th>Outcome</th>
<th>ADE $\beta$ (95% CI)</th>
<th>ACME $\beta$ (95% CI)</th>
<th>Proportion Mediated (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{BMI}_{\text{PRS}}$</td>
<td>PCOS</td>
<td>T2D</td>
<td>0.025 (0.014-0.037)</td>
<td>0.002 (6.92E-04-0.004)</td>
<td>0.071 (0.025-0.144)</td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>PCOS</td>
<td>0.025 (0.013-0.037)</td>
<td>0.002 (0.001-0.004)</td>
<td>0.090 (0.039-0.191)</td>
</tr>
<tr>
<td>$\text{BMI}_{\text{residual}}$</td>
<td>PCOS</td>
<td>T2D</td>
<td>0.037 (0.023-0.052)</td>
<td>0.006 (0.002-0.010)</td>
<td>0.136 (0.049-0.254)</td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>PCOS</td>
<td>0.127 (0.104-0.150)</td>
<td>0.003 (8.12-04- 0.005)</td>
<td>0.021 (0.006-0.042)</td>
</tr>
</tbody>
</table>

ADE = average direct effect, ACME = average causal mediation effect, CI = confidence interval
Table 4. Genetic effects of T2D Contributes to Environmental BMI.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Exposure</th>
<th>Mediator</th>
<th>Outcome</th>
<th>ADE β (95% CI)</th>
<th>ACME β (95% CI)</th>
<th>Proportion Mediated (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex-Combined</td>
<td>PCOSPRS</td>
<td>T2D</td>
<td>BMIresidual</td>
<td>0.041 (-0.055-0.147) P = 0.44</td>
<td>0.059 (0.033-0.088) P = &lt;2e-16</td>
<td>0.589 (-1.112-3.248) P = 0.06</td>
</tr>
<tr>
<td>Females</td>
<td>PCOSPRS</td>
<td>T2D</td>
<td>BMIresidual</td>
<td>0.024 (-0.102-0.172) P = 0.78</td>
<td>0.059 (0.041-0.121) P = &lt;2e-16</td>
<td>0.652 (-4.638-6.128) P = 0.18</td>
</tr>
<tr>
<td>Males</td>
<td>T2DPRS</td>
<td>PCOS</td>
<td>BMIresidual</td>
<td>0.040 (-0.101-0.176) P = 0.58</td>
<td>0.038 (0.004-0.074) P = 0.02</td>
<td>0.342 (-2.354-4.320) P = 0.31</td>
</tr>
<tr>
<td>Females</td>
<td>T2DPRS</td>
<td>PCOS</td>
<td>BMIresidual</td>
<td>0.268 (0.018-0.519) P = 0.04</td>
<td>0.139 (0.046-0.242) P = 0.002</td>
<td>0.338 (0.122-0.863) P = 0.004</td>
</tr>
</tbody>
</table>
Table 5. Genetically Regulated BMI is a Stronger Mediator in PCOS<sub>PRS</sub> Shared Metabolic Pathways.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Mediator</th>
<th>Outcome</th>
<th>ADE β (95% CI)</th>
<th>ACME β (95% CI)</th>
<th>Proportion Mediated (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex-Combined</td>
<td>BMI</td>
<td>T2D</td>
<td>0.009 (0.004-0.014)</td>
<td>0.005 (0.004-0.007)</td>
<td>0.367 (0.250 - 0.585)</td>
</tr>
<tr>
<td></td>
<td>BMI Residuals</td>
<td></td>
<td>0.0134 (0.008- 0.019)</td>
<td>0.001 (-1.76E-04- 0.002)</td>
<td>0.072 (-0.011- 0.155)</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>Hypertension</td>
<td>0.007 (-2.13e-04-0.013)</td>
<td>0.005 (0.004-0.006)</td>
<td>0.435 (0.269-1.043)</td>
</tr>
<tr>
<td></td>
<td>BMI Residuals</td>
<td></td>
<td>0.012 (0.005-0.0178)</td>
<td>8.89e-04 (-3.13E-04-0.002)</td>
<td>0.072 (-0.036-0.207)</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>Hypertensive Heart Disease</td>
<td>0.008 (0.003-0.013)</td>
<td>0.002 (0.001-0.003)</td>
<td>0.237 (0.144-0.457)</td>
</tr>
<tr>
<td></td>
<td>BMI Residuals</td>
<td></td>
<td>0.009 (0.004-0.014)</td>
<td>0.0006 (-0.0001-0.001)</td>
<td>0.062 (-0.011-0.152)</td>
</tr>
<tr>
<td>Females</td>
<td>BMI</td>
<td>T2D</td>
<td>0.012 (0.005-0.019)</td>
<td>0.005 (0.004-0.007)</td>
<td>0.315 (0.208-0.535)</td>
</tr>
<tr>
<td></td>
<td>BMI Residuals</td>
<td></td>
<td>0.016 (0.009-0.023)</td>
<td>9.64e-04 (-6.00e-04-0.003)</td>
<td>0.058 (-0.037-0.154)</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>Hypertension</td>
<td>0.002 (-0.007-0.010)</td>
<td>0.006 (0.004-0.007)</td>
<td>0.742 (-2.165-5.363)</td>
</tr>
<tr>
<td></td>
<td>BMI Residuals</td>
<td>Hypertensive Heart Disease</td>
<td>Hypertension</td>
<td>T2D</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>-----------------------------</td>
<td>--------------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.007</td>
<td>0.002</td>
<td>0.014</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.002-0.015)</td>
<td>(0.001-0.003)</td>
<td>(0.004-0.024)</td>
<td>(0.008-0.027)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.11</td>
<td>P = 0.31</td>
<td>P = 0.06</td>
<td>P = 0.03</td>
<td></td>
</tr>
<tr>
<td>BMI Residuals</td>
<td>0.003</td>
<td>0.005</td>
<td>0.018</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.002-0.009)</td>
<td>(0.003-0.007)</td>
<td>(0.009-0.027)</td>
<td>(0.006-0.025)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.31</td>
<td>P = 0.13</td>
<td>P = &lt;2e-16</td>
<td>P = &lt;2e-16</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.005</td>
<td>0.004</td>
<td>0.003</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.004-0.024)</td>
<td>(0.003-0.006)</td>
<td>(0.001-0.004)</td>
<td>(0.008-0.027)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.03</td>
<td>P = &lt;2e-16</td>
<td>P = &lt;2e-16</td>
<td>P = &lt;2e-16</td>
<td></td>
</tr>
<tr>
<td>Hypertensive Heart Disease</td>
<td>0.007</td>
<td>0.002</td>
<td>0.003</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.002-0.015)</td>
<td>(0.001-0.003)</td>
<td>(0.001-0.004)</td>
<td>(-0.063-0.139)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.11</td>
<td>P = &lt;2e-16</td>
<td>P = &lt;2e-16</td>
<td>P = 0.43</td>
<td></td>
</tr>
</tbody>
</table>

PCOS<sub>PRS</sub> was analyzed as the exposure variable. P < 0.05 was considered statistically significant. ADE = average direct effect, ACME = average causal mediation effect, CI = confidence interval.
Figure 1. Metabolic Associations Vary by Sex for Genetically High-Risk Individuals. PCOSPRS phenome-wide association study results are displayed in the Manhattan plots for (A) females and (B) males. The red line represents the Bonferroni correction of $P = 4.56e-05$ and $P = 5.07e-05$, respectively. Arrows in the upward direction represent increased risk.
The phenome-wide association study (PheWAS) was performed on the full dataset (females and males) to increase power. Significant results that passed Bonferroni correction ($P = 3.73 \times 10^{-05}$) are annotated in the Manhattan plot.

**Figure 2.** Sex Combined PCOS$^{\text{PRS}}$ PheWAS Analysis Increases Detection of Associations.
Figure 3. The Prevalence of Hypertensive Metabolic Traits Differs Across the PCOS\textsubscript{PRS} Strata. (A) Top Figure: Hypertension, (B) Bottom Figure: Hypertensive Heart Disease. The prevalence of the significant phenotypes from the sex interaction analysis are shown. The prevalence for hypertension in our dataset was 48% for males and 36% for females. The prevalence for hypertensive heart disease was 10% for males and 5% for females.
Figure 4. T2D_{PRS} and BMI_{PRS} are Associated with PCOS Case Status. The results from the logistic regression analysis between PCOS diagnosis and polygenic risk scores (PRS) generated for significant metabolic traits from the PheWAS analysis are displayed in the forest plot. One model was unadjusted for body mass index (unadjusted) and the other model was adjusted for body mass index (adjusted). * indicates P < 0.05.
Figure 5. **T2DPRS is Associated with PCOS After Accounting for BMI_{residual}**. The logistic regression model between T2DPRS and PCOS diagnosis was covaried for age, the top ten principal components for ancestry, and BMI_{residual}. BMI_{residual} represents BMI after the removal of BMI genetic variance from the variable. * indicates P < 0.05.