

## Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women

Kumarendran, Balachandran; O'Reilly, Michael; Manolopoulos, Konstantinos; Toulis, Konstantinos; Gokhale, Krishna; Sitch, Alice; Wijeyaratne, Chandrika N. ; Coomarasamy, Aravinthan; Arlt, Wiebke; Nirantharakumar, Krishnarajah

DOI:

[10.1371/journal.pmed.1002542](https://doi.org/10.1371/journal.pmed.1002542)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Kumarendran, B, O'Reilly, M, Manolopoulos, K, Toulis, K, Gokhale, K, Sitch, A, Wijeyaratne, CN, Coomarasamy, A, Arlt, W & Nirantharakumar, K 2018, 'Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women: a longitudinal study based on a United Kingdom primary care database', *PLoS Medicine*, vol. 15, no. 3, e1002542. <https://doi.org/10.1371/journal.pmed.1002542>

[Link to publication on Research at Birmingham portal](#)

### **Publisher Rights Statement:**

Checked for eligibility: 18/04/2018  
[10.1371/journal.pmed.1002542](https://doi.org/10.1371/journal.pmed.1002542)

### **General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### **Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

RESEARCH ARTICLE

# Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women: A longitudinal study based on a United Kingdom primary care database

Balachandran Kumarendran<sup>1,2‡</sup>, Michael W. O'Reilly<sup>3,4‡</sup>, Konstantinos N. Manolopoulos<sup>3,4‡</sup>, Konstantinos A. Toulis<sup>1</sup>, Krishna M. Gokhale<sup>1</sup>, Alice J. Sitch<sup>1</sup>, Chandrika N. Wijeyaratne<sup>5</sup>, Arri Coomarasamy<sup>3,4</sup>, Wiebke Arlt<sup>3,4</sup>\*, Krishnarajah Nirantharakumar<sup>1,3,4</sup>✉

**1** Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom, **2** Department of Public Health, Faculty of Medicine, University of Kelaniya, Kelaniya, Sri Lanka, **3** Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, United Kingdom, **4** Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, United Kingdom, **5** Department of Obstetrics and Gynaecology, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka

✉ These authors contributed equally to this work.

‡ These authors share first authorship on this work.

\* [w.arlt@bham.ac.uk](mailto:w.arlt@bham.ac.uk) (WA); [k.nirantharan@bham.ac.uk](mailto:k.nirantharan@bham.ac.uk) (KN)



## OPEN ACCESS

**Citation:** Kumarendran B, O'Reilly MW, Manolopoulos KN, Toulis KA, Gokhale KM, Sitch AJ, et al. (2018) Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women: A longitudinal study based on a United Kingdom primary care database. *PLoS Med* 15(3): e1002542. <https://doi.org/10.1371/journal.pmed.1002542>

**Academic Editor:** Jenny E Myers, University of Manchester, UNITED KINGDOM

**Received:** September 12, 2017

**Accepted:** February 21, 2018

**Published:** March 28, 2018

**Copyright:** © 2018 Kumarendran et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant metadata used to derive the results are within the paper and its supplementary files. In Tables 1 and 2 and S10 Table and S13 Table, numerators and denominators for all calculated proportions are given. The only continuous variable (age) is presented with standard deviation to work out the distribution of age in the exposed and control population. Metadata to independently calculate the hazard ratios, both in exposed and unexposed

## Abstract

### Background

Androgen excess is a defining feature of polycystic ovary syndrome (PCOS), which affects 10% of women and represents a lifelong metabolic disorder, with increased risk of type 2 diabetes, hypertension, and cardiovascular events. Previous studies have suggested an increased risk of nonalcoholic fatty liver disease (NAFLD) in individuals with PCOS and implicated androgen excess as a potential driver.

### Methods and findings

We carried out a retrospective longitudinal cohort study utilizing a large primary care database in the United Kingdom, evaluating NAFLD rates in 63,120 women with PCOS and 121,064 age-, body mass index (BMI)-, and location-matched control women registered from January 2000 to May 2016. In 2 independent cohorts, we also determined the rate of NAFLD in women with a measurement of serum testosterone ( $n = 71,061$ ) and sex hormone-binding globulin (SHBG;  $n = 49,625$ ). We used multivariate Cox models to estimate the hazard ratio (HR) for NAFLD and found that women with PCOS had an increased rate of NAFLD (HR = 2.23, 95% CI 1.86–2.66,  $p < 0.001$ ), also after adjusting for BMI or dysglycemia. Serum testosterone  $>3.0$  nmol/L was associated with an increase in NAFLD (HR = 2.30, 95% CI 1.16–4.53,  $p = 0.017$  for 3–3.49 nmol/L and HR = 2.40, 95% CI 1.24–4.66,  $p = 0.009$  for  $>3.5$  nmol/L). Mirroring this finding, SHBG  $<30$  nmol/L was associated with increased NAFLD hazard (HR = 4.75, 95% CI 2.44–9.25,  $p < 0.001$  for

groups, are given in Table 3 and S1, S2, S5, S7, S12, and S15 Tables (number of outcomes and person years of follow-up in each arm). THIN data governance does not allow us to share individual patient data, and therefore, the metadata are presented. Researchers may apply for individual patient data access at <https://www.iqvia.com/solutions/real-world-value-and-outcomes> (contact tab). Clinical codes (Read codes) used to extract data are provided in the supplementary table files (S3 and S4 Tables).

**Funding:** Wellcome Trust provided funding to MWOR (Clinical Research Training Fellowship 099909) and WA (Project Grant 092283). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: WA declares competing interests in relation to this paper, comprised of a scientific consultancy for Bayer AG and grant funding by the Wellcome Trust, a charitable foundation supporting research into human health. All other authors have declared that no competing interests exist.

**Abbreviations:** BMI, body mass index; HR, hazard ratio; IFG, impaired fasting glucose; IGR, impaired glucose regulation; IGT, impaired glucose tolerance; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PCO, polycystic ovaries; PCOS, polycystic ovary syndrome; SHBG, sex hormone-binding globulin; RECORD, Reporting of Studies Conducted Using Observational Routinely Collected Health Data; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

20–29.99 nmol/L and HR = 4.98, 95% CI 2.45–10.11,  $p < 0.001$  for  $<20$  nmol/L). Limitations of this study include its retrospective nature, absence of detailed information on criteria used to diagnosis PCOS and NAFLD, and absence of data on laboratory assays used to measure serum androgens.

## Conclusions

We found that women with PCOS have an increased rate of NAFLD. In addition to increased BMI and dysglycemia, androgen excess contributes to the development of NAFLD in women with PCOS. In women with PCOS-related androgen excess, systematic NAFLD screening should be considered.

## Author summary

### Why was this study done?

- Polycystic ovary syndrome (PCOS) affects 10% of women and is defined by androgen excess, chronic anovulation, and polycystic appearance of the ovaries on ultrasound.
- PCOS is not only a reproductive but also a lifelong metabolic disorder with increased rates of type 2 diabetes, hypertension, and cardiovascular events.
- Recent research has shown rates of nonalcoholic fatty liver disease (NAFLD) may be higher in women with PCOS and that androgens may play a causative role in its pathogenesis.

### What did the researchers do and find?

- We performed a population-based retrospective cohort study utilizing a large UK primary care database and included more than 63,000 women with PCOS and 121,000 matched controls registered between 2000 and 2016.
- We found that rates of NAFLD were increased in women with PCOS.
- We found that even normal-weight women with PCOS had an increased rate of NAFLD.
- In addition to body mass index and dysglycemia, we identified androgen excess as a potential additional contributing risk factor for NAFLD development in PCOS.
- We also studied 2 independent cohorts drawn from the same primary care database who had serum testosterone ( $n = 71,000$ ) or sex hormone-binding globulin (SHBG,  $n = 49,000$ ) measured and found that women with biochemical evidence of androgen excess (high testosterone, low SHBG) had an increased rate of NAFLD.

### What do these findings mean?

- Our findings indicate there may be an increased risk of NAFLD in women with PCOS and that androgen excess is a factor underlying that risk.
- Screening for NAFLD should be considered in women with PCOS-related androgen excess.
- Future studies will need to determine if androgen excess also drives progression to liver inflammation and fibrosis and to establish whether antiandrogen medication can reduce the risk of NAFLD.

## Introduction

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in women of reproductive age, affecting 6%–10% of women worldwide [1]. Diagnostic criteria include the presence of androgen excess, oligomenorrhea, and evidence of polycystic ovaries (PCO) on ultrasound [2]. Though conventionally perceived as a reproductive disorder, PCOS is now emerging as a lifelong metabolic disorder, with evidence of increased prevalence of obesity, insulin resistance, and metabolic syndrome [3]. However, the metabolic disease burden in patients with PCOS exceeds that observed in simple obesity [4]. Androgen excess has been implicated as a distinct risk factor, with several studies showing circulating androgen burden to correlate closely with surrogate markers of metabolic risk, independent of body mass index (BMI) [5–7].

Nonalcoholic fatty liver disease (NAFLD) is a hepatic complication of the metabolic syndrome, with global NAFLD prevalence rising to epidemic proportions in recent years [8]. It encompasses a continuous spectrum of morphological changes in the liver, from simple hepatic steatosis, continuing to the stage of nonalcoholic steatohepatitis (NASH), with prevailing inflammation, and eventually progressing to irreversible hepatic fibrosis and cirrhosis [9]. Overall, patients with NAFLD have increased mortality due to an increased risk of cardiometabolic complications and death from liver failure and hepatocellular carcinoma [10]. NAFLD is now the second most prevalent indication for liver transplantation in the United States [11]. There is a clear association between NAFLD risk, insulin resistance and obesity. Adipose tissue is thought to be the principal contributor of free fatty acids and systemic lipotoxicity underpinning the development of hepatic steatosis [12].

Given the shared risk factors between PCOS and NAFLD, there is a strong interest in exploring the relationship between the 2 conditions, both in terms of prevalence and shared pathophysiological mechanisms. A systematic review of 7 studies, involving mostly small cohorts with less than 60 patients, reported an estimated 4-fold increased risk of NAFLD among patients with PCOS compared to controls [13]. Similarly, 2 small cross-sectional studies found significant associations between NAFLD and PCOS [14,15]. Following the publication of some preliminary evidence [16], 2 recent studies and a very recent meta-analysis have implicated androgen excess as a potentially BMI-independent risk factor for NAFLD in PCOS [17–19].

Our study aimed to comprehensively investigate the excess incidence of NAFLD in patients with PCOS, assess if the excess risk was independent of their BMI status, and explore the potential role of androgen excess as an independent risk factor in the development of NAFLD, by undertaking a longitudinal study drawing from a large and diverse population base.

## Methods

The study followed the preanalysis study plan ([S1 Text](#)) and is reported as per Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines ([S2 Text](#)).

## Study design

This is a population-based retrospective cohort study to determine the association between PCOS (exposure) and NAFLD (outcome) and to assess predictors of NAFLD within the PCOS group. Independent cohorts were separately created to explore the association of biomarkers of androgen excess (serum testosterone and sex hormone-binding globulin [SHBG]) with NAFLD risk ([S2 Fig](#)).

## Setting

Data for this study were derived from The Health Improvement Network (THIN) database, a UK general practice electronic database. THIN data are generated from longitudinal data documented in electronic medical records by general practitioners for clinical and management purposes during each episode of consultations using Read codes, a hierarchical coding system for structured storage of information [20]. More than 675 practices, scattered representatively around UK, contribute data to THIN, covering 3.7 million active patients (6%–7% of UK population) [21]. THIN data are generalizable for the UK for major health conditions [22] and have been used for studies involving women with PCOS [23] and patients with NAFLD [24].

## Participants

Women aged 18 to 50 years at study entry were eligible to take part in the study. Patients with any documentation of excess alcohol use were excluded. Excess alcohol use was determined by Read codes indicative of hazardous drinking or diseases related to hazardous drinking (e.g., alcoholic liver disease). The exposed group comprised eligible women with a diagnosis of PCOS at any time during the study period. The nonexposed group comprised eligible women who did not have a documented diagnosis of PCOS and were matched within the same practice for age ( $\pm 1$  year) and BMI ( $\pm 2$  kg/m<sup>2</sup>) of each woman with PCOS. Age and BMI were used as matching variables because of their strong association with PCOS and NAFLD [25,26]. Where there were several patients who could be matched, for each exposed patient 2 were randomly selected.

## Study period

Our study period extended from 1 January 2000 (study start date) to 15 May 2016 (study end date). The follow-up start date for an indexed patient (index date) in the exposed group was set at first documentation of PCOS once a patient was eligible to take part in the study (newly diagnosed patients) or set as the date a patient became eligible to take part in the study, if they already had a diagnosis of PCOS (patients with an existing diagnosis) ([S1 Fig](#)). The same index date was assigned for their matched corresponding unexposed patients to mitigate immortality time bias [27]. The earliest of transfer date (when patient left the practice), death date, first documentation of outcome—i.e., NAFLD—(outcome date), or study end date were considered as the follow-up end date (exit date).

## PCOS definition

High similarity between “PCOS” and “polycystic ovaries (PCO)” increases the possibility of misclassification between these 2 during data entry. However, prevalence studies based on

primary care data suggest that both combined together reflect true prevalence [23]. Therefore, primary analysis included both, while sensitivity analysis was carried out with codes reflecting PCOS only.

Though limitations exist in documentation of the 3 features (ovulatory dysfunction, androgen excess, and PCO) used in the diagnosis of PCOS, these were captured where available. We considered clinical androgen excess as present if clinical features commonly accepted as indicative of clinically relevant androgen excess in women (hirsutism, alopecia, or acne) were documented.

## Outcome

Our primary outcome was incidence of NAFLD.

## Independent cohorts to explore the role of androgen excess in the development of NAFLD

Two cohorts independent of the primary PCOS cohort were constructed to test the hypothesis that androgen excess has a role in women developing NAFLD. The first cohort comprised any woman with a serum testosterone measurement, and the second one comprised any woman with a serum SHBG measurement. This was essential as androgen excess features (acne, hirsutism, and hair loss) and biomarkers of androgen excess were not sufficiently documented within the PCOS cohort and therefore did not have adequate power to detect an association (S1 and S2 Tables). Both cohorts had similar entry and exit requirements as the primary cohort study (S1 Fig).

## Selection of read codes and drug codes for the cohort exposure, outcome, and covariates

Variables for PCOS, NAFLD, and covariates were defined using published methodologies [28], definitions noted in previous publications [23,24], and our own systematic search process for codes (S3 and S4 Tables). PCOS was defined using Read codes for polycystic ovary syndrome [23] and polycystic ovaries [29]. The outcome variable NAFLD was defined using nonalcoholic fatty liver and nonalcoholic steatohepatitis Read codes [24]. Other potential confounders and covariates were informed by previous literature and included the Townsend index of deprivation [30,31], diabetes mellitus or impaired glucose regulation (IGR, includes impaired fasting glucose [IFG; defined as fasting plasma glucose 6.1–6.9 mmol/L] and impaired glucose tolerance [IGT; defined as plasma glucose 7.8–11.1 mmol/L measured 120 minutes after ingestion of 75 g glucose in the oral glucose tolerance test]), hypothyroidism, metformin use, and lipid-lowering medications [23,32]. Predictor variables were also selected for analysis restricted to women with PCOS only. These included Read codes for hirsutism, acne, and alopecia and medication codes for cyproterone acetate, drospirenone, and other antiandrogen medications.

## Grouping of quantitative variables

BMI (in kg/m<sup>2</sup>) was categorized into <25, 25 to 30, and >30 kg/m<sup>2</sup> as per the WHO recommendation for obesity measurement (normal weight, overweight, and obese).

## Statistical methods

The basic characteristics of participants, including covariates, were summarized by exposure status using appropriate summary measures. No statistical significance testing was performed



to show any difference between the patients with PCOS and the control patients at baseline, in line with good practice [33,34].

**Primary analysis and sensitivity analysis.** Crude hazard ratios (HRs) were estimated, followed by adjusted estimates using Cox-regression models. We reported the HR estimates with 95% confidence intervals. Covariates in Cox models for main analysis included age, BMI category, Townsend category, hypothyroidism, and diabetes mellitus status. Diabetes status at baseline and follow-up were both used in separate analyses. Analyses were undertaken using diabetes status during follow-up as a covariate to assess the impact of PCOS on the development of NAFLD not associated with the development of diabetes. Analyses were also undertaken by excluding patients with diabetes and IGR at baseline and by censoring when they developed diabetes or IGR subsequently during follow-up. This enabled us to assess if the hazard of developing NAFLD was independent of the subsequent development of diabetes or IGR.

Sensitivity analysis was performed to assess the possibility of bias due to case definition (PCOS and PCO versus PCOS codes only) and survival bias due to including prevalent cases with PCOS (diagnosed with PCOS before meeting eligibility criteria) in the analysis [35].

**Analysis to assess if NAFLD risk in women with PCOS was independent of BMI.** First, we conducted a subgroup analysis within each BMI category to assess if the NAFLD risk was independent of BMI status. In addition, interaction terms were used to assess possible interaction between the exposure (PCOS/PCO) and BMI categories and their combined effect on NAFLD risk.

**Analysis to explore if androgen excess in women with PCOS confers a risk for NAFLD.** Firstly, we introduced the clinical androgen excess features into the primary model to see if it attenuated the effect size. A subsequent analysis limited to exposed patients was carried out to identify predictors of NAFLD in women with PCOS with phenotypes and antiandrogenic medications as covariates in addition to previously stated covariates. These 2 analyses were limited given that only a third of patients had their androgen excess features documented.

Finally, we carried out an analysis of the 2 independent cohorts (women with a serum testosterone measurement or an SHBG measurement, respectively). Serum SHBG and testosterone categories were derived based on clinically meaningful cutoff values and by consensus among the study group to demonstrate any dose response relationship. Serum SHBG concentrations were categorized into <20, 20–29.99, 30–39.99, 40–49.99, 50–59.99, and ≥60 nmol/L, and serum testosterone concentrations were categorized into <1, 1–1.49, 1.5–1.99, 2–2.49, 2.5–2.99, 3–3.49, and ≥3.5 nmol/L. Association between these hormonal levels and NAFLD was assessed using Cox regression models adjusting for potential covariates. All data cleaning and analysis were carried out using STATA MP version 14.2.

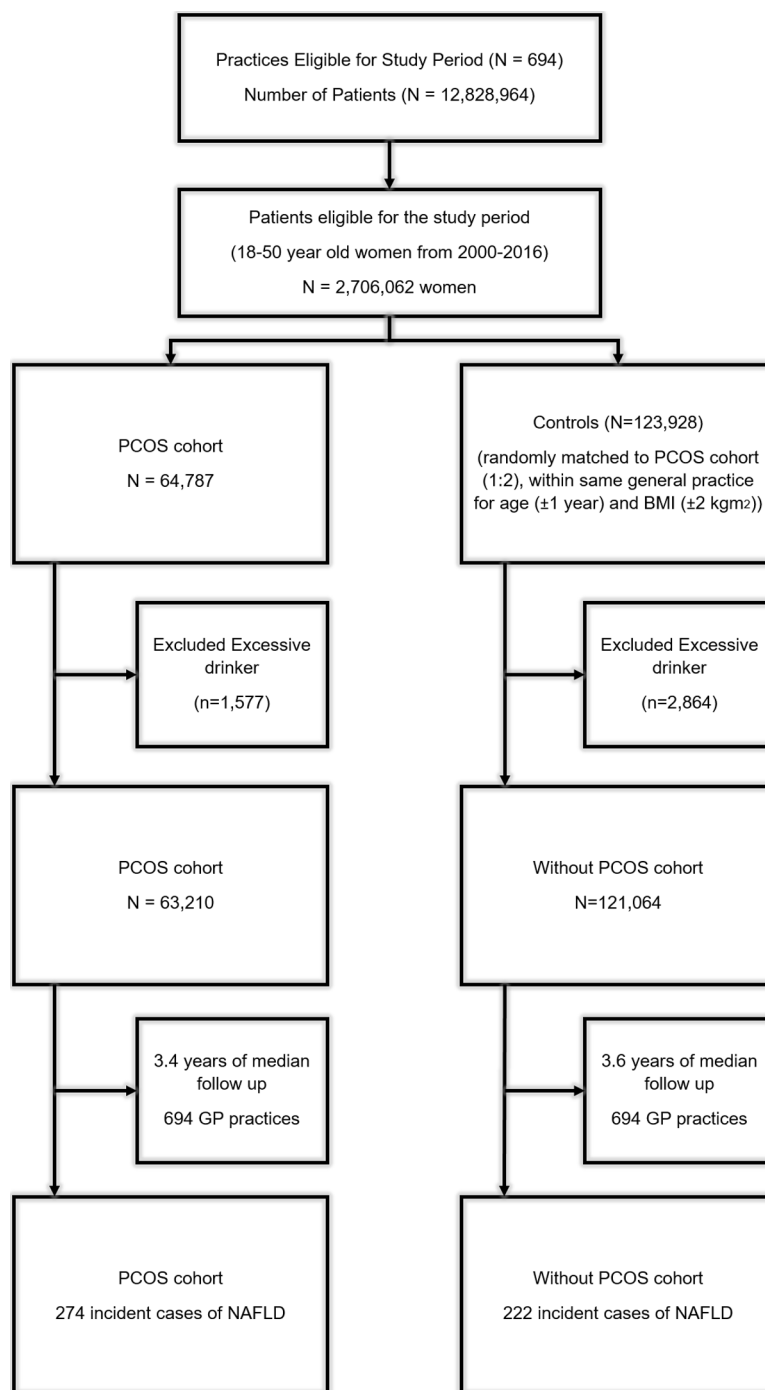
## Ethical approval

The use of the THIN data for research was approved by the South-East Multicenter Research Ethics Committee in 2003 [36] without the need for informed consent. As per requirement of the ethical approval, further registration and authorization for this project were obtained from the relevant Scientific Review Committee (17THIN026).

## Results

### Study population characteristics

The study population included 63,210 women with PCOS and 121,064 women without PCOS, matched for age, BMI, and general practice (Fig 1). Both cohorts were followed up for a



**Fig 1. Selection of patients with polycystic ovary syndrome (PCOS) and controls.** A total of 694 general practices were eligible for inclusion in the study, with a total of 2,706,062 women for the study period 2000–2016. A total of 64,787 women were listed as having PCOS according to the criteria defined in the “Methods” section. Controls were randomly matched to women with PCOS at a ratio of 2:1 within each practice, generating a random control cohort of 123,928 women. Participants with a documented history of alcohol excess were excluded; therefore, the final cohort consisted of 63,120 women with PCOS and 121,064 controls.

<https://doi.org/10.1371/journal.pmed.1002542.g001>

median of 3.5 years (IQR 1.4–7.1 years). The median age of participants was 30 years (IQR 25.2–35.4 years), with no apparent imbalance in distribution of age, Townsend index, and



smoking status between groups. At baseline, women with PCOS were more likely to have diabetes (2.2% versus 1.4% in controls), hypertension (3.1% versus 2.4%), and hypothyroidism (3.9% versus 2.5%) (Table 1). During follow-up, 2.1% of women in the PCOS cohort developed diabetes in comparison to 1.1% of women in the unexposed cohort.

Among the women with PCOS, 19,425 were defined based on the PCOS Read code, while the remainder ( $n = 43,785$ ) were defined based on a PCO Read code. In total, 20,162 women had a diagnosis after joining the general practice and were categorized as incident cases. Among women with PCOS, 34.5% of women had documented clinical features suggestive of androgen excess (acne 21.7%, hirsutism 12.9%, and androgenetic alopecia 6.2%), and 29.3% had features suggestive of oligoanovulation (Table 2). Before study entry or during follow-up, 29.4% of women with PCOS ( $n = 18,583$ ) took an oral contraceptive with an antiandrogenic progestin component (cyproterone acetate 22.6% and drospirenone 12.1%), and 24% ( $n = 15,152$ ) received metformin treatment (Table 2).

## Primary analysis

Women with PCOS were found to have an increased rate of NAFLD in comparison to matched controls without PCOS (HR = 2.38, 95% CI 1.99–2.84,  $p < 0.001$ ). The incidence of NAFLD in women with PCOS was 9.2 per 10,000 person years, while in the matched control

**Table 1. Baseline characteristics of the 63,210 women with polycystic ovary syndrome (PCOS) and 121,064 control women without PCOS.**

Characteristics	PCOS	Controls
Age mean (SD)	30.6 (7.1)	30.8 (7.1)
Townsend score* $n$ (%)		
1	12,410 (19.6)	23,540 (19.4)
2	11,224 (17.8)	21,126 (17.4)
3	13,238 (20.9)	25,629 (21.3)
4	12,634 (20.0)	24,930 (20.6)
5	8,737 (13.8)	17,596 (14.5)
Missing data	4,967 (7.9)	8,243 (6.8)
BMI (kg/m <sup>2</sup> ) categories $n$ (%)		
<25	21,689 (34.3)	44,742 (37.0)
25–30	12,412 (19.6)	24,656 (20.4)
>30	19,499 (30.8)	32,627 (27.0)
Missing or implausible data	9,610 (15.2)	19,039 (15.7)
Smoking status $n$ (%)		
Nonsmokers	46,133 (73.0)	88,519 (73.1)
Smokers	14,191 (22.4)	29,272 (24.2)
Missing or implausible data	2,886 (4.6)	3,273 (2.7)
Medical conditions at baseline $n$ (%)		
Diabetes	1,378 (2.2)	1,693 (1.4)
Hypertension	1,953 (3.1)	2,839 (2.4)
Hypothyroidism	2,462 (3.9)	3,009 (2.5)
IGR*	348 (0.6)	384 (0.3)

\*Townsend score of social deprivation: presented as quintiles with 1 as the least deprived and 5 as the most deprived.

\*IGR, impaired glucose regulation (includes impaired fasting glucose [IFG; defined as fasting plasma glucose 6.1–6.9 mmol/L] and impaired glucose tolerance [IGT; defined as plasma glucose 7.8–11.1 mmol/L measured 120 minutes after ingestion of 75 g glucose in the oral glucose tolerance test]).

<https://doi.org/10.1371/journal.pmed.1002542.t001>

**Table 2. Polycystic ovary syndrome (PCOS) diagnostic features and use of medication.**

Characteristics	All cases and control	
	PCOS ( <i>n</i> = 63,210)	Controls ( <i>n</i> = 121,064)
	Number (%)	Number (%)
<b>PCOS features</b>		
Anovulation	18,514 (29.29)	13,466 (11.12)
Any androgen excess	21,786 (34.47)	25,341 (20.93)
Hirsutism	8,142 (12.88)	1,921 (1.59)
Acne	13,708 (21.69)	19,968 (16.49)
Alopecia	3,895 (6.16)	5,472 (4.52)
Polycystic ovaries	48,369 (76.5)	(0.00)
<b>Medications</b>		
OCP including an antiandrogenic progestin component*	18,583 (29.40)	15,564 (12.86)
Cyproterone	14,282 (22.59)	8,767 (7.24)
Drospirenone	7,648 (12.10)	9,302 (7.68)
Other antiandrogen	155 (0.25)	13 (0.01)
Metformin	15,152 (23.97)	2,966 (2.45)
Lipid-modifying drugs	2,722 (4.31)	3,432 (2.83)

OCP, oral contraceptive pill.

\*Use of either cocypindiol or cyproterone acetate or drospirenone. Since some patients were prescribed both cyproterone and drospirenone during the period covered, the number of OCP users is slightly smaller than the arithmetic sum of the numbers of cyproterone and drospirenone users.

<https://doi.org/10.1371/journal.pmed.1002542.t002>

group, it was 3.9 per 10,000 person years (Table 3). The estimated hazard ratio remained similar after adjustment for age, Townsend score, BMI, diabetes or IGR, and hypothyroidism at baseline (HR = 2.23, 95% CI 1.86–2.66,  $p < 0.001$ ) and when adjusted for dysglycemia developing during follow-up (HR = 2.14, 95% CI 1.79–2.56,  $p < 0.001$ ) (Table 3).

Analysis excluding women with diabetes or IGR at baseline and by censoring when they developed diabetes or IGR did not alter our estimates (HR = 2.22, 95% CI 1.85–2.68,  $p < 0.001$ ) (S5 Table).

**Table 3. Hazard of women with polycystic ovary syndrome (PCOS) developing nonalcoholic fatty liver disease (NAFLD) compared to women without PCOS.**

	Primary analysis		Sensitivity analysis (Incident cases)		Sensitivity analysis (PCOS-specific Read codes only)	
	PCOS (Exposed)	Controls (Unexposed)	PCOS (Exposed)	Controls (Unexposed)	PCOS (Exposed)	Controls (Unexposed)
Total number of participants	63,210	121,064	20,162	37,544	19,425	35,967
Person years	297,842	576,065	94,778	167,009	84,041	149,852
Incident NAFLD <i>n</i> (%)	274 (0.43)	222 (0.18)	80 (0.40)	65 (0.17)	83 (0.43)	64 (0.18)
Incidence rates per 10,000 person years	9.2	3.9	8.4	3.9	9.9	4.3
Hazard ratio (95% CI)	2.38 (1.99–2.84)		2.15 (1.55–2.99)		2.24 (1.62–3.11)	
<i>p</i> -value	<0.001		<0.001		<0.001	
Baseline adjusted hazard ratio (95% CI)*	2.23 (1.86–2.66)		2.07 (1.49–2.87)		2.10 (1.52–2.93)	
<i>p</i> -value	<0.001		<0.001		<0.001	
Follow-up adjusted hazard ratio (95% CI)†	2.14 (1.79–2.56)		1.98 (1.43–2.75)		2.04 (1.47–2.84)	
<i>p</i> -value	<0.001		<0.001		<0.001	

\*Adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation, and hypothyroidism.

†Adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation up to end of follow-up, and hypothyroidism at baseline.

<https://doi.org/10.1371/journal.pmed.1002542.t003>

Increasing age (HR = 1.05, 95% CI 1.03–1.06,  $p < 0.001$ ), BMI (BMI 25–30 kg/m<sup>2</sup>, HR = 3.37, 95% CI 2.36–4.83,  $p < 0.001$ ; and BMI >30 kg/m<sup>2</sup>, HR = 6.98, 95% CI 5.07–9.60,  $p < 0.001$ ), and diabetes or IGR (HR = 2.39, 95% CI 1.76–3.25,  $p < 0.001$ ) at baseline were also found to be significant independent factors associated with an increased rate of NAFLD (S6 Table), with the impact of PCOS on NAFLD rate estimated to be of a similar magnitude to the impact of diabetes and IGR.

### Sensitivity analysis

Findings were robust when analysis was limited to PCOS codes only (adjusted HR = 2.10, 95% CI 1.52–2.93,  $p < 0.001$ ) and when limited to incident cases only (HR = 2.07, 95% CI 1.49–2.87,  $p < 0.001$ ) (Table 3).

### Analysis to assess if NAFLD risk in women with PCOS was independent of BMI

In the subgroup analysis, women with PCOS were at increased risk of NAFLD in all 3 BMI subgroups (HR = 1.85, 95% CI 1.02–3.34,  $p = 0.043$  in <25 kg/m<sup>2</sup>; HR = 2.05, 95% CI 1.37–3.07,  $p < 0.001$  for 25–30 kg/m<sup>2</sup>; and HR = 2.06, 95% CI 1.65–2.57,  $p < 0.001$  for >30 kg/m<sup>2</sup>) (S7 Table). In the analysis with interaction terms for PCOS and BMI, women with PCOS and a BMI of <25 kg/m<sup>2</sup> had twice the risk of developing NAFLD (HR = 1.92, 95% CI 1.06–3.46,  $p < 0.031$ ), while women with PCOS with a BMI of 25–30 kg/m<sup>2</sup> or >30 kg/m<sup>2</sup> were, respectively, 7 times (HR = 6.75, 95% CI 4.10–11.10,  $p < 0.001$ ) and 14 times (HR = 14.12, 95% CI 9.11–21.87,  $p < 0.001$ ) more at risk compared to control women with a BMI less than 25 kg/m<sup>2</sup> (Fig 2 and S8 Table).

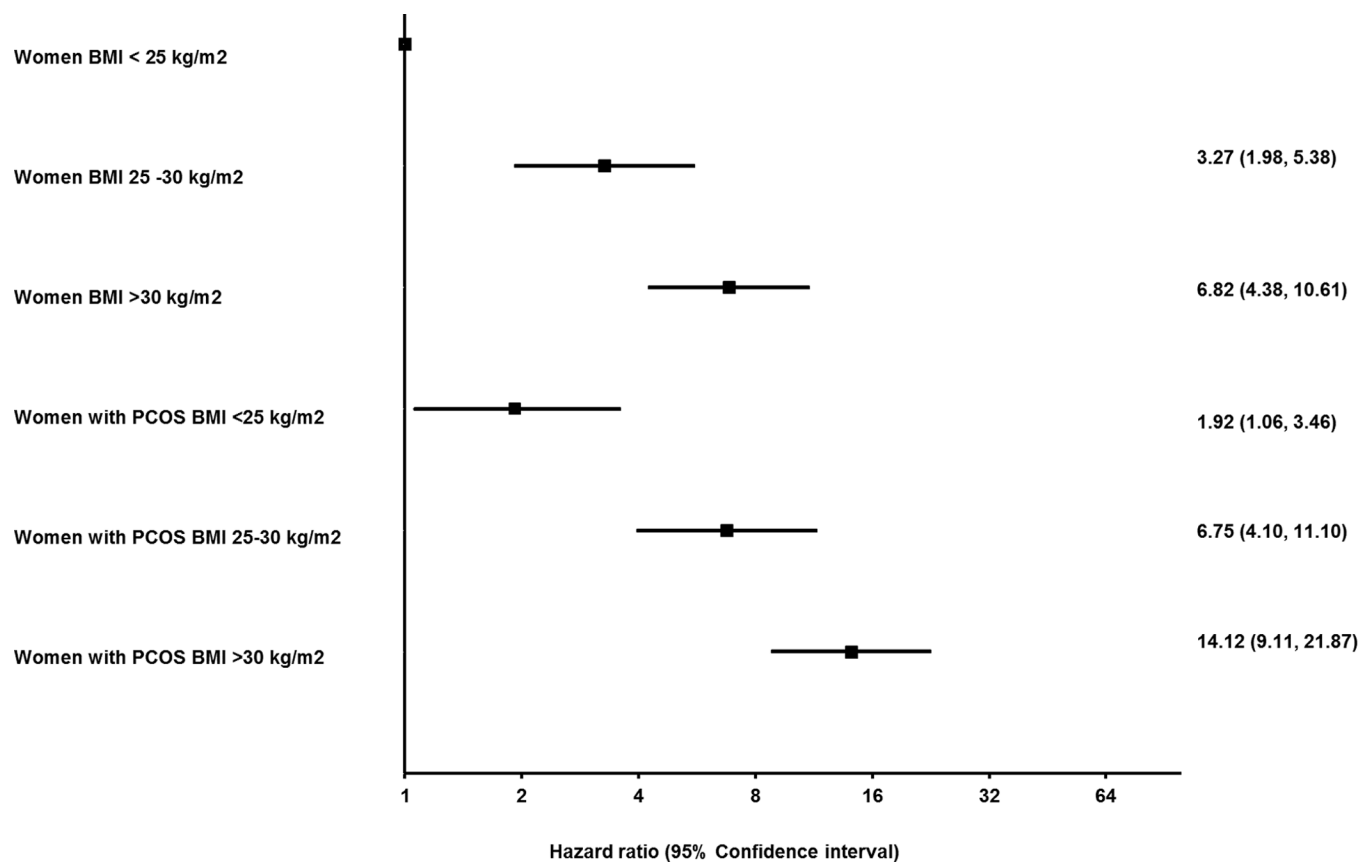
### Analysis to explore if androgen excess in women with PCOS confers a higher risk of NAFLD compared to patients with PCOS alone. Future studies should aim to replicate our findings in others for NAFLD

Introducing androgen excess features into the primary model suggested a potential reduction in the HR from 2.23 (95% CI 1.86–2.66) to 2.07 (95% CI 1.72–2.50) (S9 Table).

**Clinical features associated with NAFLD in women with PCOS.** Along with increasing age, higher BMI and dysglycemia, anovulation (HR = 1.36, 95% CI 1.06–1.756,  $p = 0.02$ ) and clinical androgen excess features, hirsutism (HR = 1.41, 95% CI 1.06–1.870,  $p = 0.01$ ) and alopecia (HR = 1.74, 95% CI 1.23–2.45,  $p = 0.02$ ), but not acne (HR = 0.87, 95% CI 0.62–1.22,  $p = 0.87$ ), were found to be significantly associated with NAFLD within the women with PCOS (Table 4). Antiandrogen therapy was associated with reduced hazard, but this did not reach statistical significance (HR = 0.80, 95% CI 0.59–1.097,  $p = 0.16$ ) (Table 4). Increasing age, Townsend index of deprivation, higher BMI, hypothyroidism, and anovulation (HR = 1.60, 95% CI 1.16–2.20,  $p = 0.004$ ) were found to be significantly associated with NAFLD within women without PCOS (S10 Table).

### Analysis of the cohorts with serum testosterone and SHBG measurements

There were 71,061 eligible women with a measurement of serum testosterone (median age 31.8 years, IQR 25.9–38.1 years) (S2 Fig). Within this cohort, serum testosterone results (nmol/l) were distributed as follows (concentration, %): <1, 33.7%; 1–1.49, 24.9%; 1.5–1.99, 19.0%; 2–2.49, 11.5%; 2.5–2.99, 5.7%; 3–3.49, 2.6%; and ≥3.5, 2.6% (S11 Table). Patients in both the 3–3.49 nmol/l (adjusted HR = 2.30, 95% CI 1.16–4.53,  $p = 0.017$ ) and ≥3.5 nmol/l



**Fig 2. Impact of BMI and polycystic ovary syndrome (PCOS) on the hazard of nonalcoholic fatty liver disease (NAFLD).** PCOS and control groups were subcategorized into lean, overweight, and obese groups according to BMI, with lean control women designated as the reference category (unexposed, BMI <25 kg/m<sup>2</sup>). The analysis is adjusted for age, Townsend score of social deprivation, diabetes or impaired glucose regulation, and hypothyroidism at baseline. The hazard of NAFLD increased with BMI in both control women and women with PCOS. Lean women with PCOS had an almost 2-fold higher hazard of NAFLD compared to lean controls. The highest hazard of developing NAFLD was observed in the obese PCOS cohort.

<https://doi.org/10.1371/journal.pmed.1002542.g002>

(adjusted HR = 2.40, 95% CI 1.24–4.66,  $p = 0.009$ ) groups had increased rates of NAFLD (Fig 3A and S12, S13 and S14 Tables).

There were 49,625 eligible women with a measurement of serum SHBG (median age 30.4 years, IQR 24.9–36.6 years) (S2 Fig). Within this cohort, there were 9.2% in the <20 nmol/L group, 15.3% in the 20–29.99 group, 16.0% in the 30–39.99, 14.4% in the 40–49.99, 11.7% in the 50–59.99 and 33.6% in the ≥60 nmol/L groups (S15 Table). Patients with lower SHBG (<30 nmol/L: adjusted HR = 4.98, 95% CI 2.45–10.11,  $p < 0.001$ ; <20 nmol/L; adjusted HR = 4.75, 95% CI 2.44–9.25,  $p < 0.001$ ) had increased rates of NAFLD (Fig 3B, S16 and S17 Tables).

## Discussion

The results of this study provide, to our knowledge, the most conclusive evidence to date that women with PCOS have an increased rate of NAFLD above that conferred by simple obesity. A previous meta-analysis of 7 studies with a combined sample size of less than 600 patients suggested an almost 4-fold increase in the risk of NAFLD in women with PCOS compared with matched controls [13]. A very recent meta-analysis that combined data of 2,700 patients

**Table 4. Factors associated with nonalcoholic fatty liver disease (NAFLD) amongst women with polycystic ovary syndrome (PCOS) (*n* = 63,210).**

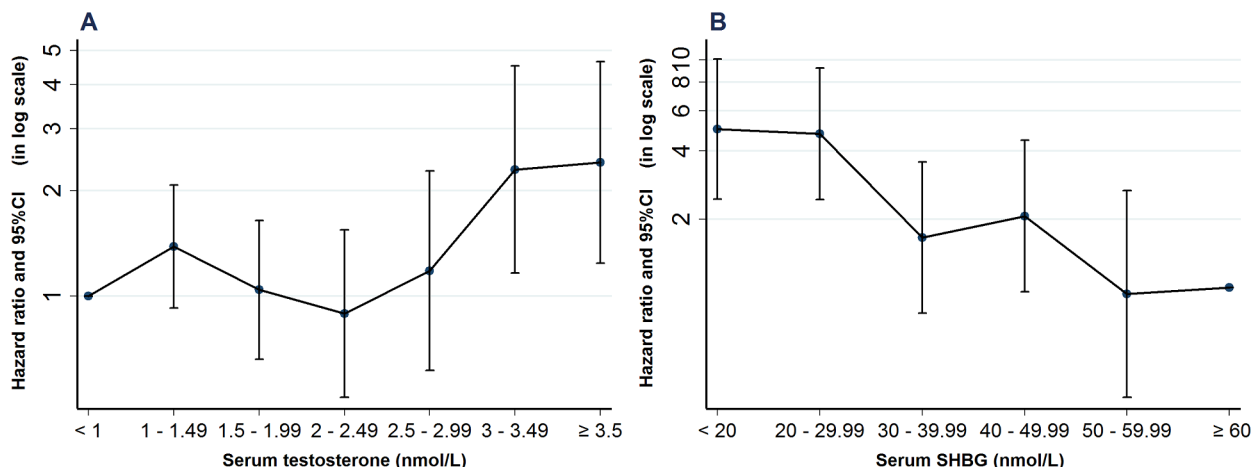
Covariates	Hazard ratio	95% CI	<i>p</i> -value
Age (years)	1.04	(1.02–1.06)	<0.001
Townsend			
1	1.00		
2	0.97	(0.66–1.43)	0.88
3	0.86	(0.59–1.28)	0.46
4	1.25	(0.87–1.79)	0.23
5	1.16	(0.78–1.74)	0.47
Missing	1.18	(0.68–2.07)	0.56
BMI (kg/m <sup>2</sup> ) category			
<25	1.00		
25–30	3.46	(2.07–5.78)	<0.001
>30	6.91	(4.35–10.98)	<0.001
Missing	3.05	(1.76–5.30)	<0.001
Diabetes or IGR*	2.57	(1.65–4.01)	<0.001
Hypothyroidism	1.04	(0.63–1.74)	0.87
Anovulation	1.36	(1.06–1.75)	0.02
Androgen excess feature			
Hirsutism	1.41	(1.06–1.87)	0.01
Acne	0.87	(0.62–1.22)	0.87
Alopecia	1.74	(1.23–2.45)	0.02
Polycystic ovaries	1.08	(0.79–1.49)	0.61
Lipid-modifying drugs	1.26	(0.71–2.23)	0.43
Metformin	1.01	(0.74–1.39)	0.94
Antiandrogen drug	0.80	(0.59–1.09)	0.16

\*IGR, impaired glucose regulation (includes impaired fasting glucose [IFG; fasting plasma glucose 6.1–6.9 mmol/L] and impaired glucose tolerance [IGT; plasma glucose 7.8–11.1 mmol/L measured 120 minutes after ingestion of 75 g glucose in the oral glucose tolerance test]).

<https://doi.org/10.1371/journal.pmed.1002542.t004>

from 17 studies found a 2.5-fold increased risk of NAFLD in women with PCOS compared with age- and BMI-matched controls [19]. To our knowledge, our population-based longitudinal cohort study is the first longitudinal and by far the largest study to examine the association between PCOS and NAFLD, analyzing over 63,000 women with PCOS against 121,000 matched controls. A diagnosis of PCOS was consistently associated with a 2.0–2.4-fold increase in rate of NAFLD in multiple adjusted analyses, and this increased hazard persisted even when restricting to patients with PCOS. Intriguingly, even women with PCOS with a normal BMI had a significantly increased rate of NAFLD.

Within our PCOS cohort, androgen excess, anovulation, and dysglycemia independently increased the hazard of developing NAFLD. The prevalence of androgen excess in our study is likely to be an underestimate, as multiple studies have shown prevalence estimates as high as 80% in PCOS [5]. To mitigate this, we further probed the association between androgens and NAFLD hazard by examining the impact of serum testosterone and SHBG levels on rate of NAFLD in separate independent cohorts. The analysis of these 2 cohorts demonstrated that female testosterone levels in excess of 3 nmol/L and SHBG levels below 30 nmol/L were associated with an increased rate of NAFLD.



**Fig 3. Hazard of nonalcoholic fatty liver disease (NAFLD) according to serum testosterone and sex hormone-binding globulin (SHBG) levels.** (A) Serum testosterone (nmol/L) was available in 71,061 unselected women. Using women with testosterone levels below 1.0 nmol/L as a reference cohort, the hazard of NAFLD was 2.4-fold higher in those with serum testosterone levels of  $>3$  nmol/L. (B) Serum SHBG was available in 49,625 unselected women. Using a reference cohort with SHBG levels of  $\geq 60$  nmol/L, the hazard of NAFLD was almost 5-fold elevated in patients with serum levels of  $<30$  nmol/L. Both analyses are adjusted for age, Townsend score, diabetes or impaired glucose regulation, and hypothyroidism at baseline.

<https://doi.org/10.1371/journal.pmed.1002542.g003>

The mechanisms underpinning the increased NAFLD rate in PCOS are likely to represent a complex interplay between androgen excess, insulin resistance, and obesity [37]. However, our data provide convincing evidence for an independent contribution from androgen excess. Several smaller studies have previously implicated androgen excess as associated with metabolic dysfunction in women with PCOS [16,38,39]. Increasing free serum testosterone levels in premenopausal women was associated with NAFLD in midlife in a recent prospective study, independent of traditional risk factors such as insulin resistance, BMI, and dyslipidemia [40]. Similarly, a meta-analysis looking at 2,700 patients from 17 studies found that serum testosterone was higher in those women with PCOS who had NAFLD [19].

We have recently provided mechanistic evidence to the association between PCOS, androgen excess, and NAFLD by showing that intra-adipose androgen generation in PCOS drives systemic lipotoxicity by increasing adipocyte hypertrophy and fatty acid overspill [41]. We found that PCOS patients with androgen excess had increased circulating glycerophospholipids and lysoglycerophospholipids, recently identified as potential biomarkers of NASH [42]. We also observed that systemic lipotoxicity increased further after an acute androgen challenge in patients with PCOS, but not in BMI-matched controls [41]. Androgen-mediated adipose lipotoxicity could therefore represent an important mechanism conveying liver injury in hyperandrogenemic PCOS.

Previous *in vitro* data have suggested a direct effect of androgens on hepatic lipid metabolism in women, with testosterone increasing lipogenic gene expression and *de novo* lipogenesis in primary human hepatocytes from female, but not male, donors [43]. The sexually dimorphic role of androgens in metabolic disease is an emerging topic, with female androgen excess and male hypogonadism sharing an overlapping metabolic phenotype characterized by abdominal obesity, dyslipidemia, insulin resistance, and NAFLD [44]. Mirroring our data, a recent meta-analysis of 13,721 men and 5,840 women demonstrated a significant sexually dimorphic association between serum testosterone and NAFLD risk [45]. Higher testosterone levels reduced the likelihood of NAFLD in men but conversely increased the risk in women. The same meta-analysis concluded that increasing SHBG concentrations in both sexes independently lowered the



risk of hepatic steatosis. In our study, serum SHBG levels below 30 nmol/L, as typically found in androgen excess, conferred an almost 5-fold increase in the hazard of NAFLD. The majority of circulating testosterone is bound to SHBG in men and women, and it is generally accepted that only unbound or “free” testosterone is able to enter metabolic target tissues [46]. Hepatic SHBG output is suppressed by insulin [5], and therefore, decreased circulating SHBG is a surrogate marker of hyperinsulinemia in the setting of systemic insulin resistance [47]. SHBG metabolism in women therefore sits at the interface between insulin resistance and androgen excess, both key players in PCOS-related metabolic dysfunction [48]. However, it should be noted that interpreting the impact of decreased serum SHBG concentrations on NAFLD risk in women is limited by the fact that both factors known to be associated with decreased SHBG, androgen excess and insulin resistance, are highly prevalent in women with PCOS [49,50].

Insulin resistance is implicated as a major player in the pathogenesis of NAFLD, and impaired suppression of adipose tissue lipolysis by insulin may lead directly to systemic free fatty acid overspill from adipocytes, with accumulation of intrahepatic diacylglycerol and triacylglycerol, as well as likely direct hepatocyte injury [51]. Parameters of insulin resistance were not available in our cohort, but we were able to phenotype patients according to glycemic status. The presence of dysglycemia was a significant factor associated with increased hazard of NAFLD within the PCOS population (HR = 2.56), but the elevated rate of NAFLD persisted in our PCOS cohort compared to control women even after correction for glycemic status at baseline and during follow-up.

The effect of obesity on NAFLD rate in this study also merits discussion. Multivariate analysis consistently showed that a diagnosis of PCOS conferred an increased rate of NAFLD independent of body weight. The observation of a significantly increased adjusted HR of NAFLD in lean PCOS patients is notable, as at present these patients are not perceived to be at high risk of progression to metabolic dysfunction [52]. As lean women with PCOS exhibit insulin resistance much more rarely than those with obesity [53], it is very reasonable to assume that androgen excess is the major factor driving the increased NAFLD hazard in women of normal weight with PCOS. This lends to the argument that androgen excess is a major causative factor in the development of NAFLD.

The limitations of this large population study include its retrospective nature, limited documentation of the diagnostic criteria, clinical features and investigations used for identifying PCOS and NAFLD, absence of any information on the stage of NAFLD, and the fact that no data were available on laboratory assays employed to measure serum testosterone concentrations. It is likely that a proportion of serum testosterone values were not measured by gold-standard liquid chromatography-tandem mass spectrometry techniques. Furthermore, data on the serum androgen precursor androstenedione, previously highlighted as the most sensitive biomarker of androgen burden and PCOS-related metabolic dysfunction [5], were not available. In addition, patients who underwent serum testosterone and SHBG measurements are not representative of the general population, as they were likely investigated for clinical indications such as suspected PCOS. It is possible that women in our 2 independent cohorts with biochemical results for serum testosterone and SHBG indicative of androgen excess were indeed individuals with underlying PCOS who may not have been evaluated further and given a diagnosis of PCOS. As expected in a study of this nature, there were no indices of insulin resistance available in the patient cohorts.

In conclusion, we found an increased rate of NAFLD in PCOS, and this extended to lean women with PCOS, indicative of androgen excess as a likely major causal factor. We could show that increased circulating androgen burden determined by serum testosterone and reduced SHBG levels independently confer an increased rate of NAFLD. Of note, antiandrogen

treatment was associated with a reduction in NAFLD rate, though this failed to reach statistical significance, likely due to the limited number of women on antiandrogens in our cohort.

Our data support that androgen excess is an independent adverse metabolic risk factor in PCOS and that metabolic risk stratification in PCOS should take androgen excess into account. We suggest that women with PCOS should be considered for systematic screening for NAFLD, if biochemical evidence of androgen excess is present; the diagnostic work-up should be carried out in accordance with clinical guidelines for the assessment of individuals deemed at increased risk of NAFLD [54]. Future studies need to clarify whether reducing androgen burden will convey long-term metabolic benefits in both lean and obese women with PCOS with androgen excess, including a reduction in their risk of developing NAFLD.

## Supporting information

**S1 Fig. Visual representation of study timeline and selection process (identification of the exposed group [individuals with polycystic ovary syndrome (PCOS)/polycystic ovaries (PCO) read codes] and their controls [without PCOS/PCO] matched 1:2 by age at index date, BMI, and practice of origin).**

(PNG)

**S2 Fig. Venn diagram showing the overlap between polycystic ovary syndrome (PCOS)/polycystic ovaries (PCO), testosterone, and sex hormone-binding globulin (SHBG) cohorts.**

(JPG)

**S1 Table. NAFLD among women with polycystic ovary syndrome (PCOS)/polycystic ovaries (PCO) by serum testosterone concentration category ( $n = 11,251$ ).**

(DOCX)

**S2 Table. NAFLD among women with polycystic ovary syndrome (PCOS)/polycystic ovaries (PCO) by serum sex hormone-binding globulin (SHBG) concentration category ( $n = 10,827$ ).**

(DOCX)

**S3 Table. Read codes used in data extraction.**

(DOCX)

**S4 Table. Drug codes used to identify antiandrogens, lipid-modifying drugs, and metformin.**

(DOCX)

**S5 Table. Hazard of women with polycystic ovary syndrome (PCOS) developing nonalcoholic fatty liver disease (NAFLD) compared to women without PCOS using diabetes censored data.**

(DOCX)

**S6 Table. Regression model estimates for hazard of women with polycystic ovary syndrome (PCOS) developing nonalcoholic fatty liver disease (NAFLD) compared to women without PCOS ( $n = 184,274$ ).**

(DOCX)

**S7 Table. Subgroup analysis for hazard of women with polycystic ovary syndrome (PCOS) developing nonalcoholic fatty liver disease (NAFLD) compared to women without PCOS stratified by BMI category.**

(DOCX)

**S8 Table. Regression model estimates for impact of BMI and polycystic ovary syndrome (PCOS) on hazard of women with PCOS developing nonalcoholic fatty liver disease (NAFLD) compared to women without PCOS ( $n = 184,274$ ).**

(DOCX)

**S9 Table. Regression model estimates for hazard of women with polycystic ovary syndrome (PCOS) developing nonalcoholic fatty liver disease (NAFLD) compared to women without PCOS with addition of PCOS diagnostic features as predictor ( $n = 184,274$ ).**

(DOCX)

**S10 Table. Factors associated with nonalcoholic fatty liver disease (NAFLD) amongst women without polycystic ovary syndrome (PCOS) ( $n = 121,064$ ).**

(DOCX)

**S11 Table. Baseline characteristics in cohort of women with available serum testosterone measurement ( $n = 71,061$ ).**

(DOCX)

**S12 Table. Factors associated with nonalcoholic fatty liver disease (NAFLD) amongst the cohort of women with available serum testosterone measurement ( $n = 71,061$ ).**

(DOCX)

**S13 Table. Hazard of women with a serum testosterone level of  $\geq 1$  nmol/L developing nonalcoholic fatty liver disease (NAFLD) compared to women with a serum testosterone level of  $< 1$  nmol/L ( $n = 71,061$ ).**

(DOCX)

**S14 Table. Risk of nonalcoholic fatty liver disease (NAFLD) in women with available serum testosterone measurement following exclusion of participants with PCOS/PCO codes ( $n = 58,606$ ).**

(DOCX)

**S15 Table. Baseline characteristics in the cohort of women with available serum sex hormone-binding globulin (SHBG) measurement ( $n = 49,625$ ).**

(DOCX)

**S16 Table. Factors associated with nonalcoholic fatty liver disease (NAFLD) amongst the cohort of women with available serum sex hormone-binding globulin (SHBG) measurement ( $n = 49,625$ ).**

(DOCX)

**S17 Table. Hazard of women with a serum sex hormone-binding globulin (SHBG) level of  $< 60$  nmol/L developing NAFLD compared to women with a serum sex hormone-binding globulin (SHBG) level of  $\geq 60$  nmol/L ( $n = 49,625$ ).**

(DOCX)

**S18 Table. Risk of nonalcoholic fatty liver disease (NAFLD) based on available serum sex hormone-binding globulin (SHBG) measurement following exclusion of participants with polycystic ovary syndrome (PCOS)/polycystic ovaries (PCO) codes ( $n = 38,784$ ).**

(DOCX)

**S1 Text. Analysis plan extracted from the approved protocol.**

(DOCX)

**S2 Text. The reporting of studies conducted using observational routinely collected health data (RECORD) statement—checklist of items, extended from the strengthening the reporting of observational studies in epidemiology (STROBE) statement, that should be reported in observational studies using routinely collected health data.**  
(DOCX)

## Author Contributions

**Conceptualization:** Balachandran Kumarendran, Wiebke Arlt, Krishnarajah Nirantharakumar.

**Data curation:** Balachandran Kumarendran, Michael W. O'Reilly, Konstantinos N. Manolopoulos, Konstantinos A. Toulis, Krishna M. Gokhale, Alice J. Sitch, Chandrika N. Wijeyaratne, Arri Coomarasamy, Wiebke Arlt, Krishnarajah Nirantharakumar.

**Formal analysis:** Balachandran Kumarendran, Alice J. Sitch, Krishnarajah Nirantharakumar.

**Methodology:** Balachandran Kumarendran, Michael W. O'Reilly, Konstantinos N. Manolopoulos, Wiebke Arlt, Krishnarajah Nirantharakumar.

**Software:** Balachandran Kumarendran, Krishna M. Gokhale, Krishnarajah Nirantharakumar.

**Supervision:** Chandrika N. Wijeyaratne, Arri Coomarasamy, Wiebke Arlt, Krishnarajah Nirantharakumar.

**Writing – original draft:** Balachandran Kumarendran, Michael W. O'Reilly, Konstantinos N. Manolopoulos, Wiebke Arlt, Krishnarajah Nirantharakumar.

**Writing – review & editing:** Balachandran Kumarendran, Michael W. O'Reilly, Konstantinos N. Manolopoulos, Konstantinos A. Toulis, Krishna M. Gokhale, Alice J. Sitch, Chandrika N. Wijeyaratne, Arri Coomarasamy, Wiebke Arlt, Krishnarajah Nirantharakumar.

## References

1. Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod.* 2016; 31(12):2841–55. <https://doi.org/10.1093/humrep/dew218> PMID: 27664216.
2. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril* 2004 Jan; 81(1): 19–25 PMID: 14711538
3. Wild RA, Carmina E, Diamanti-Kandarakis E, Dokras A, Escobar-Morreale HF, Futterweit W, et al. Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. *J Clin Endocrinol Metab.* 2010; 95(5):2038–49. <https://doi.org/10.1210/jc.2009-2724> PMID: 20375205.
4. Fauser BC, Tarlatzis BC, Rebar RW, Legro RS, Balen AH, Lobo R, et al. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertil Steril.* 2012; 97(1):28–38 e25. <https://doi.org/10.1016/j.fertnstert.2011.09.024> PMID: 22153789.
5. O'Reilly MW, Taylor AE, Crabtree NJ, Hughes BA, Capper F, Crowley RK, et al. Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *J Clin Endocrinol Metab.* 2014; 99(3):1027–36. <https://doi.org/10.1210/jc.2013-3399> PMID: 24423344; PubMed Central PMCID: PMC3955250.
6. Dewailly D, Catteau-Jonard S, Reyss AC, Leroy M, Pigny P. Oligoanovulation with polycystic ovaries but not overt hyperandrogenism. *J Clin Endocrinol Metab.* 2006; 91(10):3922–7. <https://doi.org/10.1210/jc.2006-1054> PMID: 16849400.

7. O'Reilly MW, Kempegowda P, Jenkinson C, Taylor AE, Quanson JL, Storbeck KH, et al. 11-Oxygenated C19 Steroids Are the Predominant Androgens in Polycystic Ovary Syndrome. *J Clin Endocrinol Metab.* 2017; 102(3):840–8. <https://doi.org/10.1210/jc.2016-3285> PMID: 27901631.
8. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology.* 2016; 64(1):73–84. <https://doi.org/10.1002/hep.28431> PMID: 26707365.
9. Dowman JK, Tomlinson JW, Newsome PN. Pathogenesis of non-alcoholic fatty liver disease. *QJM.* 2010; 103(2):71–83. <https://doi.org/10.1093/qjmed/hcp158> PMID: 19914930; PubMed Central PMCID: PMC2810391.
10. Armstrong MJ, Adams LA, Canbay A, Syn WK. Extrahepatic complications of nonalcoholic fatty liver disease. *Hepatology.* 2014; 59(3):1174–97. <https://doi.org/10.1002/hep.26717> PMID: 24002776.
11. Wong RJ, Aguilar M, Cheung R, Perumpail RB, Harrison SA, Younossi ZM, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology.* 2015; 148(3):547–55. <https://doi.org/10.1053/j.gastro.2014.11.039> PMID: 25461851.
12. Jelenik T, Kaul K, Sequaris G, Fogel U, Phielix E, Kotzka J, et al. Mechanisms of Insulin Resistance in Primary and Secondary Non-Alcoholic Fatty Liver. *Diabetes.* 2017. <https://doi.org/10.2337/db16-1147> PMID: 28490610.
13. Ramezani-Binabaj M, Motalebi M, Karimi-Sari H, Rezaee-Zavareh MS, Alavian SM. Are women with polycystic ovarian syndrome at a high risk of non-alcoholic Fatty liver disease; a meta-analysis. *Hepat Mon.* 2014; 14(11):e23235. <https://doi.org/10.5812/hepatmon.23235> PMID: 25598791; PubMed Central PMCID: PMC4286712.
14. Vassilatou E, Vassiliadi DA, Salambasis K, Lazaridou H, Koutsomitopoulos N, Kelekis N, et al. Increased prevalence of polycystic ovary syndrome in premenopausal women with nonalcoholic fatty liver disease. *Eur J Endocrinol.* 2015; 173(6):739–47. <https://doi.org/10.1530/EJE-15-0567> PMID: 26340970.
15. Romanowski MD, Parolin MB, Freitas AC, Piazza MJ, Basso J, Urbanetz AA. Prevalence of non-alcoholic fatty liver disease in women with polycystic ovary syndrome and its correlation with metabolic syndrome. *Arq Gastroenterol.* 2015; 52(2):117–23. <https://doi.org/10.1590/S0004-28032015000200008> PMID: 26039829.
16. Vassilatou E, Lafoyianni S, Vryonidou A, Ioannidis D, Kosma L, Katsoulis K, et al. Increased androgen bioavailability is associated with non-alcoholic fatty liver disease in women with polycystic ovary syndrome. *Hum Reprod.* 2010; 25(1):212–20. <https://doi.org/10.1093/humrep/dep380> PMID: 19887498.
17. Cai J, Wu CH, Zhang Y, Wang YY, Xu WD, Lin TC, et al. High-free androgen index is associated with increased risk of non-alcoholic fatty liver disease in women with polycystic ovary syndrome, independent of obesity and insulin resistance. *Int J Obes (Lond).* 2017. <https://doi.org/10.1038/ijo.2017.116> PMID: 28487551.
18. Kim JJ, Kim D, Yim JY, Kang JH, Han KH, Kim SM, et al. Polycystic ovary syndrome with hyperandrogenism as a risk factor for non-obese non-alcoholic fatty liver disease. *Aliment Pharmacol Ther.* 2017; 45(11):1403–12. <https://doi.org/10.1111/apt.14058> PMID: 28370150.
19. Rocha ALL, Faria LC, Guimaraes TCM, Moreira GV, Candido AL, Couto CA, et al. Non-alcoholic fatty liver disease in women with polycystic ovary syndrome: systematic review and meta-analysis. *J Endocrinol Invest.* 2017; 40(12): 1279–1288 <https://doi.org/10.1007/s40618-017-0708-9> PMID: 28612285.
20. Booth N. What are the Read Codes? *Health Libr Rev.* 1994; 11(3):177–82. PMID: 10139676.
21. Sammon CJ, Petersen I. Backdating of events in electronic primary health care data: should one censor at the date of last data collection. *Pharmacoepidemiol Drug Saf.* 2016; 25(4):378–84. <https://doi.org/10.1002/pds.3981> PMID: 26887352.
22. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care.* 2011; 19(4):251–5. PMID: 22828580.
23. Ding T, Baio G, Hardiman PJ, Petersen I, Sammon C. Diagnosis and management of polycystic ovary syndrome in the UK (2004–2014): a retrospective cohort study. *BMJ Open.* 2016; 6(7):e012461. <https://doi.org/10.1136/bmjopen-2016-012461> PMID: 27401369; PubMed Central PMCID: PMC4947736.
24. Loomis AK, Kabadi S, Preiss D, Hyde C, Bonato V, St Louis M, et al. Body Mass Index and Risk of Non-alcoholic Fatty Liver Disease: Two Electronic Health Record Prospective Studies. *J Clin Endocrinol Metab.* 2016; 101(3):945–52. <https://doi.org/10.1210/jc.2015-3444> PMID: 26672639; PubMed Central PMCID: PMC4803162.
25. Lim SS, Norman RJ, Davies MJ, Moran LJ. The effect of obesity on polycystic ovary syndrome: a systematic review and meta-analysis. *Obes Rev.* 2013; 14(2):95–109. <https://doi.org/10.1111/j.1467-789X.2012.01053.x> PMID: 23114091.



26. Li L, Liu DW, Yan HY, Wang ZY, Zhao SH, Wang B. Obesity is an independent risk factor for non-alcoholic fatty liver disease: evidence from a meta-analysis of 21 cohort studies. *Obes Rev*. 2016; 17(6):510–9. <https://doi.org/10.1111/obr.12407> PMID: 27020692.
27. Karim ME, Gustafson P, Petkau J, Tremlett H, Long-Term Benefits Adverse Effects of Beta-Interferon for Multiple Sclerosis Study Group, Karim ME, et al. Comparison of Statistical Approaches for Dealing With Immortal Time Bias in Drug Effectiveness Studies. *Am J Epidemiol*. 2016; 184(4):325–35. <https://doi.org/10.1093/aje/kwv445> PMID: 27455963; PubMed Central PMCID: PMC4983651.
28. Dave S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf*. 2009; 18(8):704–7. <https://doi.org/10.1002/pds.1770> PMID: 19455565.
29. Vinogradova Y, Coupland C, Hippisley-Cox J. Use of combined oral contraceptives and risk of venous thromboembolism: nested case-control studies using the QResearch and CPRD databases. *BMJ*. 2015; 350:h2135. <https://doi.org/10.1136/bmj.h2135> PMID: 26013557; PubMed Central PMCID: PMC4444976.
30. Townsend P. Deprivation. *Journal of Social Policy*. 1987; 16(2):125–46.
31. Keane MG, Horsfall LJ, Rait G, Pereira SP. Sociodemographic trends in the incidence of pancreatic and biliary tract cancer in UK primary care. *PLoS ONE*. 2014; 9(9):e108498. <https://doi.org/10.1371/journal.pone.0108498> PMID: 25268478; PubMed Central PMCID: PMC4182488.
32. Glintborg D, Hass Rubin K, Nybo M, Abrahamsen B, Andersen M. Morbidity and medicine prescriptions in a nationwide Danish population of patients diagnosed with polycystic ovary syndrome. *Eur J Endocrinol*. 2015; 172(5):627–38. <https://doi.org/10.1530/EJE-14-1108> PMID: 25656495.
33. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995; 310(6973):170. PMID: 7833759; PubMed Central PMCID: PMC2548561.
34. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015; 12(10):e1001885. <https://doi.org/10.1371/journal.pmed.1001885> PMID: 26440803; PubMed Central PMCID: PMC4595218.
35. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004; 58(8):635–41. <https://doi.org/10.1136/jech.2003.008466> PMID: 15252064; PubMed Central PMCID: PMC1732856.
36. Petersen I, Collings SL, McCrea RL, Nazareth I, Osborn DP, Cowen PJ, et al. Antiepileptic drugs prescribed in pregnancy and prevalence of major congenital malformations: comparative prevalence studies. *Clin Epidemiol*. 2017; 9:95–103. <https://doi.org/10.2147/CLEP.S118336> PMID: 28243149; PubMed Central PMCID: PMC5317245.
37. Hazlehurst JM, Tomlinson JW. Non-alcoholic fatty liver disease in common endocrine disorders. *Eur J Endocrinol*. 2013; 169(2):R27–37. <https://doi.org/10.1530/EJE-13-0296> PMID: 23653455.
38. Chen MJ, Chiu HM, Chen CL, Yang WS, Yang YS, Ho HN. Hyperandrogenemia is independently associated with elevated alanine aminotransferase activity in young women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2010; 95(7):3332–41. <https://doi.org/10.1210/jc.2009-2698> PMID: 20427499.
39. Jones H, Sprung VS, Pugh CJ, Daousi C, Irwin A, Aziz N, et al. Polycystic ovary syndrome with hyperandrogenism is characterized by an increased risk of hepatic steatosis compared to nonhyperandrogenic PCOS phenotypes and healthy controls, independent of obesity and insulin resistance. *J Clin Endocrinol Metab*. 2012; 97(10):3709–16. <https://doi.org/10.1210/jc.2012-1382> PMID: 22837189.
40. Sarkar M, Wellons M, Cedars MI, VanWagner L, Gunderson EP, Ajmera V, et al. Testosterone Levels in Pre-Menopausal Women are Associated With Nonalcoholic Fatty Liver Disease in Midlife. *Am J Gastroenterol*. 2017; 112(5):755–62. <https://doi.org/10.1038/ajg.2017.44> PMID: 28291240.
41. O'Reilly MW, Kempegowda P, Walsh M, Taylor AE, Manolopoulos KN, Allwood JW, et al. AKR1C3-mediated adipose androgen generation drives lipotoxicity in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2017. <https://doi.org/10.1210/jc.2017-00947> PMID: 28645211.
42. Anjani K, Lhomme M, Sokolovska N, Poitou C, Aron-Wisniewsky J, Bouillot JL, et al. Circulating phospholipid profiling identifies portal contribution to NASH signature in obesity. *J Hepatol*. 2015; 62(4):905–12. <https://doi.org/10.1016/j.jhep.2014.11.002> PMID: 25450212.
43. Nasiri M, Nikolaou N, Parajes S, Krone NP, Valsamakis G, Mastorakos G, et al. 5alpha-Reductase Type 2 Regulates Glucocorticoid Action and Metabolic Phenotype in Human Hepatocytes. *Endocrinology*. 2015; 156(8):2863–71. <https://doi.org/10.1210/en.2015-1149> PMID: 25974403; PubMed Central PMCID: PMC4511138.
44. Schiffer L, Kempegowda P, Arlt W, O'Reilly MW. MECHANISMS IN ENDOCRINOLOGY: The sexually dimorphic role of androgens in human metabolic disease. *Eur J Endocrinol*. 2017. <https://doi.org/10.1530/EJE-17-0124> PMID: 28566439.



45. Jaruvongvanich V, Sanguankeo A, Rianguwat T, Upala S. Testosterone, Sex Hormone-Binding Globulin and Nonalcoholic Fatty Liver Disease: a Systematic Review and Meta-Analysis. *Ann Hepatol*. 2017; 16(3):382–94. <https://doi.org/10.5604/16652681.1235481> PMID: 28425408.
46. Rosner W, Hryb DJ, Kahn SM, Nakhla AM, Romas NA. Interactions of sex hormone-binding globulin with target cells. *Molecular and cellular endocrinology*. 2010; 316(1):79–85. <https://doi.org/10.1016/j.mce.2009.08.009> PMID: 19698759.
47. Semple RK, Savage DB, Cochran EK, Gorden P, O'Rahilly S. Genetic syndromes of severe insulin resistance. *Endocr Rev*. 2011; 32(4):498–514. <https://doi.org/10.1210/er.2010-0020> PMID: 21536711.
48. Franks S. Polycystic ovary syndrome. *N Engl J Med*. 1995; 333(13):853–61. <https://doi.org/10.1056/NEJM199509283331307> PMID: 7651477.
49. Randeve HS, Tan BK, Weickert MO, Lois K, Nestler JE, Sattar N, et al. Cardiometabolic aspects of the polycystic ovary syndrome. *Endocr Rev*. 2012; 33(5):812–41. <https://doi.org/10.1210/er.2012-1003> PMID: 22829562; PubMed Central PMCID: PMC3461136.
50. Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. *Endocr Rev*. 1997; 18(6):774–800. <https://doi.org/10.1210/edrv.18.6.0318> PMID: 9408743.
51. Tilg H, Moschen AR, Roden M. NAFLD and diabetes mellitus. *Nat Rev Gastroenterol Hepatol*. 2017; 14(1):32–42. <https://doi.org/10.1038/nrgastro.2016.147> PMID: 27729660.
52. Li S, Chu Q, Ma J, Sun Y, Tao T, Huang R, et al. Discovery of Novel Lipid Profiles in PCOS: Do Insulin and Androgen Oppositely Regulate Bioactive Lipid Production? *J Clin Endocrinol Metab*. 2017; 102(3):810–21. <https://doi.org/10.1210/jc.2016-2692> PMID: 27886515.
53. Cassar S, Teede HJ, Harrison CL, Joham AE, Moran LJ, Stepto NK. Biomarkers and insulin sensitivity in women with Polycystic Ovary Syndrome: Characteristics and predictive capacity. *Clin Endocrinol (Oxf)*. 2015; 83(1):50–8. <https://doi.org/10.1111/cen.12619> PMID: 25262763.
54. Byrne CD, Targher G. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease: is universal screening appropriate? *Diabetologia*. 2016; 59(6):1141–4. <https://doi.org/10.1007/s00125-016-3910-y> PMID: 27053232.