

**Metabolic health outcomes
among women with polycystic ovary syndrome**

by

Balachandran Kumarendran

**A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY**

**Institute of Applied Health Research
College of Medical and Dental Sciences
University of Birmingham
August 2023**

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Introduction: Polycystic ovary syndrome (PCOS) is the most common endocrine condition among women. Previous studies hypothesised increased risk of NAFLD, OSA, and dysglycaemia among women with PCOS, but the role of BMI and serum androgens in this association was unclear.

Objective: We aimed, 1) To determine the risk of NAFLD, OSA and dysglycaemia among women with PCOS compared to women without PCOS. 2) To determine if risk of these outcomes is independent of BMI, 3) To determine the role of serum sex hormones (serum testosterone and serum Sex Hormone Binding Globulin -SHBG) on the risk of NAFLD and dysglycaemia, 4) To determine the risk factors (including clinical androgen excess features such as hirsutism) for these outcomes among the women with PCOS.

Methods: We used the DExtER app to extract three types of retrospective cohorts from a UK primary care database (THIN) for primary analysis and additional cohorts for sensitivity analysis. Women between the ages of 18 and 50 were eligible to be included in the study. The study period was 2000 to 2016. The primary cohorts included around 63,000 women with PCOS and around 120,000 women without PCOS who were matched for general practice, age (± 1 year) and BMI (± 2 kg/m²). We also developed additional independent cohorts of around 70,000 women with serum testosterone measurements, and around 40,000 women with serum SHBG measurements for NAFLD and dysglycaemia outcomes separately. We performed Cox regression with adjustment for important covariates and performed sensitivity analysis by restricting the analysis to incident exposures and PCOS only codes. We performed sub-group analysis and Cox regression to verify if the risk is independent of BMI. We used STATA software for the data analysis.

Results and conclusion: Compared to women without PCOS, women with PCOS had nearly twice the risk for NAFLD (adjusted HR = 2.23, 95% CI 1.86 ± 2.66, $p < 0.001$), OSA (adjusted HR = 2.26, 95% CI: 1.89–2.69, $P < 0.001$), and dysglycemia (adjusted HR = 1.87, 95% CI: 1.78–1.97, $P < 0.001$). This risk was similar in sensitivity analyses. The risk was independent of BMI for all three outcomes. We also observed the risk was higher with higher levels of serum testosterone and lower levels of serum SHBG for NAFLD and dysglycaemia. Clinical androgen excess features were risk factors for all three outcomes in women with PCOS. This study has limitations that are common to primary care database studies, such as its retrospective design, the lack of detailed information on the criteria used to diagnose PCOS and NAFLD, and OSA, and the lack of data on the laboratory assays used to measure serum androgens.

Conclusion: Compared to women without PCOS, women with PCOS have higher risk for NAFLD, OSA and dysglycaemia which is independent of BMI. We recommend a lower threshold for screening for NAFLD, OSA, and dysglycaemia among women with PCOS, and exploring the possibility of clinical trials to reduce this risk by managing androgen excess.

"To ask the right question is half the battle."

- Roman statesman and philosopher Cicero

“The first step to answering a question is to understand it.”

- Aristotle

(These two quotes were the first two pieces of advice given to me by Prof. Arri Coomarasamy)

Acknowledgement

I would like to express my sincere gratitude to my supervisors, Professor Krish Nirantharakumar, Professor Arri Coomarasamy, and Dr. Dana Sumilo, for their guidance and support throughout my PhD. I am grateful for their encouragement, advice, and constructive criticism, which have helped me to develop as a researcher. I am also grateful to Professor Chandirka Wijeratne, Professor Wibke Arlt, Professor Michael W. O'Reilly, Konstantinos N. Manolopoulos, Professor Abd A Tahrani, Krishna M. Gokhale, Anuradhaa Subramanian, Konstantinos A. Toulis, Alice J. Sitch, and the other members of the THINKING group for their contributions to my research.

I would also like to thank my past and present line managers, Professor NR de Silva, Professor A. Pathmeswaran, Professor Anuradhani Kasturiratne, Professor R. Surenthirakumaran, Dr. Dinesh Coonghe, and Professor KK Cheng, for their support and encouragement. I am grateful for their guidance and advice, which have helped me to develop as an academic staff and researcher.

Finally, I would like to thank my nuclear and extended families and friends for their love and support. I am grateful for their patience and understanding during the long hours I have spent working on my PhD. I would especially like to thank my parents, my wife Malarveny, and my son, Seyon, for their constant support and encouragement.

I am grateful to everyone who has helped me directly or indirectly in the completion of my PhD. I would not have been able to achieve this without their help.

Contribution statement

Of the six chapters included in this thesis with original collaborative research, three were journal articles published in peer reviewed journals.

I wish to justify the reasons why this study should be collaborative research. I have completed an internship in gynaecology and postgraduate training in community medicine. My supervisors include a health data science professor, public health specialist, and a gynaecology professor. However, this study requires expertise from many subspecialties, namely endocrine (with a special focus on sex hormones), sleep medicine, diabetes, pharmacoepidemiology, data coding, data extraction, and advanced statistics. Prof. Wibke Arlt is a renowned researcher on sex-hormones and providence guidance throughout the study. Prof. Michael W. O'Reilly and Konstantinos N. Manolopoulos were Wibke's trainees with vast experience in sex-hormone research and helped me to understand the pathophysiological basis of androgen excess and verify the same in the discussion section. Considering their contribution to the journal article and to facilitate their career progress, Michael o' Reilly (in two publications) and Konstantinos N. Manolopoulos (in one publication) were added sharing the first authorship. Therefore, I have not added the methods or discussed the elements of the pharmaco-epidemiology study.

The journal article on dysglycaemia included both retrospective cohort design to determine the risk of dysglycaemia and a pharmacoepidemiology nested case control study to determine the effectiveness of oral contraceptive pills on the risk of dysglycaemia. I led the retrospective cohort study and led the overall writing of the manuscript. Anuradha Subramanian, a fellow PhD student, led the pharmacoepidemiology nested case control study and therefore added as equal contribution in the manuscript. I have not added the methods or discussed the elements of the pharmaco-epidemiology study.

Other co-authors were added to the articles for the following reasons: Krish Nirantharakumar, Dana Sumilo, and Arri Coomarasamy are my PhD supervisors. Chandirka Wijeratne, a professor and endocrinologist in Sri Lanka, was appointed by the University of Birmingham as local mentor of my split-site PhD. Konstantinos A. Toulis is an endocrinologist and researcher experienced in THIN data. He held an honorary visiting scholar position at the University of Birmingham and trained me in endocrine research using the THIN database. Toulis contributed to verifying the Read codes and drug codes shortlisted by me and reviewed the manuscript. Krishna M. Gokhale produced the DExtER app jointly with Krish Nirantharakumar and performed the data extraction. Alice J. Sitch, a biostatistician, verified if the statistical codes were correct in some of the complex analyses. Abd A Tahrani, an honorary Professor in the University of Birmingham, shared his sleep research expertise in study design and interpretation of clinical relevance of the results of the OSA study.

I added three additional chapters to my thesis: a general introduction, general methods, and general discussion. I independently completed these chapters to demonstrate my skills in designing the structure of a chapter, performing literature searches, scientific writing, and demonstrating my understanding of key concepts such as PCOS, androgen excess, complications of PCOS including NAFLD, OSA, and dysglycaemia, general practice databases, coding systems, defining retrospective cohorts using specific timelines, knowledge of data analysis including Cox regression, applying interaction terms in Cox regression, sensitivity analysis, subgroup analysis, and interpreting results. These three chapters enabled me to present my thesis independently by integrating the three journal articles to provide a storyline so that the attainment of the required standard can be verified. They also helped to overcome the limitation of word count in the journal articles.

Contents

Abstract.....	1
Acknowledgement.....	4
Contribution statement.....	5
List of publications arising from this PhD.....	16
Chapter 1: General Introduction	23
1.1. Definition of polycystic ovary syndrome.....	23
1.1.1. Establishing the diagnosis of Polycystic ovary syndrome.....	23
1.1.2. Clinical androgen excess.....	23
1.1.3. Biochemical androgen excess.....	24
1.1.4. Chronic anovulation	24
1.1.5. Polycystic ovaries.....	25
1.1.6. Evolution of criteria for the diagnosis of PCOS	26
1.2. Phenotypes of PCOS	27
1.3. Underlying pathophysiology	28
1.4. Epidemiology of PCOS.....	30
1.4.1. Prevalence of PCOS	30
1.4.2. Incidence of PCOS	30
1.4.3. Age related distribution and natural progression of PCOS.....	30
1.4.4. Ethnicity-related distribution of PCOS.....	31
1.5. Factors associated with PCOS or cardiometabolic conditions.....	34

1.5.1. Body mass index.....	34
1.5.2. Thyroid conditions	34
1.5.3. Social deprivation or socioeconomic condition	35
1.6. Existing evidence on PCOS and cardiometabolic conditions	36
1.6.1. PCOS and Metabolic syndrome	36
1.6.2. Non-Alcoholic Fatty Liver Disease	36
1.6.3. Obstructive Sleep Apnoea	37
1.6.4. Dysglycaemia	39
1.6.5. Other cardiometabolic conditions	40
1.7. PCOS and other health conditions	40
1.7.1. PCOS and cancers	40
1.7.2. PCOS and mental health conditions	40
1.8. Management of PCOS	41
1.8.1. Medications in the management of PCOS.....	41
1.8.2. Surgical interventions in the management of PCOS	42
1.8.3. Lifestyle interventions in the management of PCOS	42
1.9. Quality of life of women with PCOS.....	42
1.10. Health economic burden of PCOS	43
1.11. Mortality among women with PCOS.....	43
1.12. PCOS and Routine data studies	44
1.13. PCOS in the UK.....	45
1.14. Objectives	46

1.14.1. Objectives of study with NAFLD outcome	46
1.14.2. Objectives of study with OSA outcome.....	47
1.14.3. Objectives of study with dysglycaemia outcome	47
1.15. References of Chapter 1: General Introduction	48
Chapter 2: General methods.....	104
2.1. Study design.....	104
2.2. Study setting.....	104
2.3. Study data source	105
2.3.1. Real-world data or routine health data	105
2.3.2. Electronic health data	106
2.3.3. General practice electronic health data in the UK	108
2.3.4. General practice databases in the UK	108
2.3.5. The Health Improvement Network data (THIN Data)	110
2.4. Codes and code list used in the study	114
2.4.1. Read codes: The structured clinical vocabulary used in THIN	114
2.4.2. Preparation of code list.....	116
2.5. Study period.....	118
2.6. Study population	121
2.6.1. PCOS diagnosis based retrospective cohorts.....	122
2.6.2. Androgen excess cohorts.....	123
2.7. Data extraction	124
2.8. Data analysis	125

2.8.2. Study population characteristics	125
2.8.3. Primary analysis	126
2.8.4. Sensitivity analysis	127
2.8.5. Determination if the risk of outcome is independent of BMI.....	128
2.8.6. Determining the risk of outcomes among women with PCOS when compared with women without PCOS by adding PCOS diagnostic features as predictor	129
2.8.7. Determining the role of factors other than BMI in outcomes among women with PCOS and without PCOS.....	129
2.8.8. Data analysis of biochemical androgen excess cohorts	130
2.9. Role of bias	131
2.9.1. Selection bias.....	132
2.9.2. Misclassification bias	132
2.9.3. Survivor bias.....	133
2.9.4. Immortal time bias.....	133
2.9.5. Cognitive bias	134
2.10.Role of confounders	134
2.11.Ethics approval.....	134
2.12. Reporting guidelines	135
2.13. References of Chapter 2 : General Methods.....	136
Chapter 3 : PCOS and NAFLD	191
Polycystic ovary syndrome, androgen excess, and the risk of non-alcoholic fatty liver disease in women: A longitudinal study basedon a United Kingdom primary care database.....	191
Appendices of Chapter 3 (PCOS and NAFLD)	212
Chapter 4 : PCOS and OSA Increased risk of obstructive sleep apnoea in women with.....	266

Appendices of Chapter 4 (PCOS and OSA)	273
Chapter 5 : PCOS and dysglycaemia.....	284
Polycystic Ovary Syndrome, Combined Oral Contraceptives, and the Risk of Dysglycemia: A Population-Based Cohort Study With a Nested Pharmaco-epidemiological Case-Control Study...	284
Appendices of Chapter 5 (PCOS and dysglycaemia)	294
Chapter 6 : General Discussion	308
6.1. Summary of findings.....	308
6.2. Comparison with literature	309
6.2.1. The risk of NAFLD among women with PCOS	309
6.2.2. The risk of OSA among women with PCOS.....	309
6.2.3. The risk of dysglycaemia among women with PCOS.....	310
6.2.4. The role of BMI in the NAFLD, OSA and dysglycaemia among women with PCOS 311	
6.2.5. Role of androgen excess in NAFLD, and dysglycaemia among women with PCOS.....	311
6.2.6. Interrelationships between NAFLD, OSA, and Dysglycemia in PCOS and Non-PCOS Women	313
6.3. The implications of this research for practice and policy	315
6.4. The impact of this research	316
6.5. Strengths and limitations.....	317
6.6. Future investigations	323
6.7. Conclusion	324
6.8. References of Chapter 6 : General Discussion	326

List of Figures

Figure 1 Pathophysiology of PCOS with NAFLD, OSA and dysglycaemia outcomes.....	29
Fig 1. Selection of patients with polycystic ovary syndrome (PCOS) and controls.....	198
Fig 2. Impact of BMI and polycystic ovary syndrome (PCOS) on the hazard of nonalcoholic fatty liver disease (NAFLD).....	194
Fig 3. Hazard of nonalcoholic fatty liver disease (NAFLD) according to serum testosterone and sex hormone-binding globulin (SHBG)	196
Figure 2 Impact of BMI on the hazard of obstructive sleep apnoea (OSA)	210

List of tables

Table 1 Baseline characteristics of the 63,210 women with polycystic ovary syndrome (PCOS) and 121,064 control women without PCOS.	191
Table 2 Polycystic ovary syndrome (PCOS) diagnostic features and use of medication.	192
Table 3 Hazard of women with polycystic ovary syndrome (PCOS) developing nonalcoholic fatty liver disease (NAFLD) compared to women without PCOS.	192
Table 4 Factors associated with nonalcoholic fatty liver disease (NAFLD) amongst women with polycystic ovarysyndrome (PCOS) ($n = 63,210$)	195
Figure 1 :Selection of patients with polycystic ovary syndrome (PCOS).....	207
Table 1 Baseline characteristics of women with PCOS and control women without PCOS.....	208
Table 2 PCOS diagnostic features and use of medication	209
Table 3 Risk of women with PCOS developing OSA compared to women without PCOS	209

List of abbreviations

Abbreviation	Expansion
AHD	Additional Health Data
AHI	Apnoea Hypopnea Index
AMH	Anti-Müllerian hormone
AMR	Acceptable Mortality Reporting
aOR	Adjusted Odds Ratio
ASRM	American Society for Reproductive Medicine
BMI	Body Mass Index
CI	Confidence Interval
CPA/EE	Cyproterone Acetate/Ethinyl Estradiol
CPRD	Clinical Practice Research Datalink
CPT	Current Procedural Terminology
CSV	Comma Separated Values
CVD	Cardiovascular Diseases
DALY	Disability Adjusted Life Years
DExtER	Data Extraction for Epidemiological Research
DHEA	Dehydroepiandrosterone
DHEAS	Dehydroepiandrosterone Suphate
EHR	Electronic Health Record
EMR	Electronic Medical Record
ESHRE	European Society of Human Reproduction and Embryology
ETL	Extract, Transform and Load
FSH	Follicle Stimulating Hormone
FTE	Full Time Equivalent
GDM	Gestational Diabetes Mellitus
GDPR	General Data Protection Regulation
GLP-1 RA	Glucagon-like peptide-1 receptor agonists
GnRH	Gonadotropin-releasing hormone
GP	General Practitioner
HDL	High Density Lipoprotein
HES	Hospital Episodes Statistics
HOMA-IR	Homeostatic Model Assessment of Insulin Resistance
HPO	Hypothalamus-Pituitary-Ovary (axis)
HR	Hazard ratio
HRQoL	Health related quality of life
IFG	Impaired Fasting Glucose
IGT	Impaired Glucose Tolerance
INPS	In Practice Systems
IQR	Inter Quartile Range
LOINC	Logical Observation Identifiers Names and Codes
MRI	Magnetic Resonance Imaging
MVAP score	Multivariable Apnea Prediction Score
NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non Alcoholic Steato-Hepatitis

Abbreviation	Expansion
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIH	National Institutes of Health
OCP	Oral Contraceptive Pill
OR	Odds Ratio
OSA	Obstructive Sleep Apnoea
OXMIS	Oxford Medical Information System
PBCL	Pathology Bounded Code List
PCO	Polycystic Ovaries
PCOS	Polycystic Ovary Syndrome
PVI	Postcode Variable Indicators
QoF	Quality Outputs Framework
QoL	Quality of Life
RCT	Randomised Controlled Trial
RECORD	REporting of studies Conducted using Observational Routinely-collected health Data
SDB	Sleep Disordered Breathing
SDI	Socio Demographic Index
SHBG	Sex Hormone Binding Globulin
SNOMED CT	Systematized Nomenclature of Medicine – Clinical Terms
STATA	A statistical analysis software
STROBE	The Strengthening the reporting of observational studies in epidemiology
T2DM	Type 2 Diabetes Mellitus
THIN	The Health Improvement Network
TSH	Thyroid Stimulating Hormone
UK	United Kingdom
WHO	World Health Organisation

List of publications arising from this PhD

Published peer reviewed journal articles forming full chapters in this thesis.

Kumarendran B., O'Reilly MW, Manolopoulos KN, Toulis KA, Gokhale KM, Sitch AJ, Wijeyaratne CN, Coomarasamy A, Arlt W & Nirantharakumar K. 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 2018 15 e1002542. (<https://doi.org/10.1371/journal.pmed.1002542>)

Kumarendran. B., Sumilo, D., O'Reilly, M.W., et al. (2019) Increased risk of obstructive sleep apnoea in women with polycystic ovary syndrome: A population-based cohort study. European Journal of Endocrinology, 180 (4): 265–272. <https://doi:10.1530/EJE-18-0693>.

Balachandran Kumarendran, Michael W. O'Reilly, Anuradhaa Subramanian, Dana Šumilo, Konstantinos Toulis, Krishna M. Gokhale, Chandrika N. Wijeratne, Arri Coomarasamy, Abd A. Tahrani, Laurent Azoulay, Wiebke Arlt, Krishnarajah Nirantharakumar; Polycystic Ovary Syndrome, Combined Oral Contraceptives, and the Risk of Dysglycemia: A Population-Based Cohort Study With a Nested Pharmacoepidemiological Case-Control Study. Diabetes Care 1 December 2021; 44 (12): 2758–2766. <https://doi.org/10.2337/dc21-0437>

Published peer-reviewed journal article, with one section contributing to this thesis

O'Reilly, M.W., Glisic, M., **Kumarendran. B.**, Subramanian, A., Manolopoulos, K.N., Tahrani, A.A., Keerthy, D., Muka, T., Toulis, K.A., Hanif, W., Thomas, G.N., Franco, O.H., Arlt, W., Nirantharakumar, K., 2018. Serum testosterone, sex hormone-binding globulin and sex-specific risk of incident type 2 diabetes in a retrospective primary care cohort. Clin. Endocrinol. (Oxf). <https://doi.org/10.1111/cen.13862>

Award received for one of the publications of this PhD

President's Award from His Excellency the President of Sri Lanka - 2018

Other related publications during this doctoral study

Toulis KA, Robbins T, Reddy N, **Kumarendran B**, Gokhale K, Wijesinghe H, Cheng KK, Karavitaki N, Wass J, Nirantharakumar K. (2018) Males with prolactinoma are at increased risk of incident cardiovascular disease. *Clinical Endocrinology*. 88(1):71-76

Caleyachetty, R., Thomas, G. N., Toulis, K. A., Mohammed, N., Gokhale, K. M., **Kumarendran B.**, & Nirantharakumar, K. (2017). Metabolically Healthy Obese and Incident Cardiovascular Disease Events Among 3.5 Million Men and Women. *Journal of the American College of Cardiology*, 70(12). <https://www.ncbi.nlm.nih.gov/pubmed/28911506>

Toulis KA, Willis BH, Marshall T, **Kumarendran B**, Gokhale K, Ghosh S, Thomas GN, Cheng KK, Narendran P, Hanif W, Nirantharakumar K (2017), All-cause mortality in patients with diabetes under treatment with dapagliflozin: a population-based, open-cohort study in THIN database, *J ClinEndocrinolMetab*. <https://www.ncbi.nlm.nih.gov/pubmed/28323967>

Dafoulas GE, Toulis KA, Mccorrey D, **Kumarendran B**, Thomas GN, Willis BH, Gokhale K, Gkoutos G, Narendran P, Nirantharakumar K (2017), Type 1 diabetes mellitus and risk of incident epilepsy: a population-based, open-cohort study, *Diabetologia*, 60(2):258-261. <https://www.ncbi.nlm.nih.gov/pubmed/27796422>

Toulis KA, Hanif W, Saravanan P, Willis BH, Marshall T, **Kumarendran B**, Gokhale K, Ghosh S, Cheng KK, Narendran P, Thomas GN, Nirantharakumar K (2017). All-cause mortality in patients with diabetes under glucagon-like peptide-1 agonists: A population-based, open cohort study. *Diabetes & Metabolism*, 43(3), pp. 211–216.

Other publications during this doctoral study

Santos, I. S., Lotufo, P. A., Goulart, A. C., Brant, L. C. C., Pinto, M. M., Pereira, A. C., Barreto, S. M., Ribeiro, A. L. P., Thomas, G. N., Lip, G. Y. H., Bensenor, I. M., Arasalingam, A., Beane, A., Bensenor, I. M., Brocklehurst, P., Cheng, K. K., El-Bouri, W., Feng, M., Goulart, A. C., Greenfield S, Guo Y, Guruparan M, Gusso G, Gooden TE, Haniffa R, Humphreys L, Jolly K, Jowett S, **Kumarendran B**, Lancashire E, Lane DA, Li X, Lip GYH, Li YG, Lobban T, Lotufo PA, Holland SM, Moore DJ, Nirantharakumar K, Olmos RD, Paschoal E, Pirasanth P, Powsiga U, Romagnolli C, Santos IS, Shantsila A, Sheron VA, Shribavan K, Szmigin I, Subaschandren K, Surenthirakumaran R, Tai M, Thomas GN, Varella AC, Wang H, Wang J, Zhang H, Zhong, J. (2022). Cardiovascular Health and Atrial Fibrillation or Flutter: A Cross-Sectional Study from ELSA-Brasil. *Arquivos Brasileiros de Cardiologia*. <https://doi.org/10.36660/ABC.20210970>

Sheron, V.A., Shanmugathas, S., Gooden, T.E., Guruparan, M., **Kumarendran. B.**, Lip, G.Y.H., Holland, S.M., Nirantharakumar, K., Shribavan, K., Subaschandren, K., Haniffa, R., Surenthirakumaran, R., Thomas, G.N., Uruthirakumar, P., Greenfield, S., Lane, D.A., Beane, A.. Healthcare provider and patient perspectives on access to and management of atrial fibrillation in the Northern Province, Sri Lanka: a rapid evaluation of barriers and facilitators to care. *BMC Health Serv Res* 22, 1078 (2022). <https://doi.org/10.1186/s12913-022-08440-1>

Kusuma D, Atanasova P, Anjana RM, De Silva L, Hanif A, Hasan M, Hossain M, Indrawansa S, Jayamanne D, Jha S, Kasturiratne A, Katulanda P, Khawaja KI, **Kumarendran B**, Mridha MK, Rajakaruna V, Chambers JC, Frost G, Sassi F, Miraldo Met al., 2022, Food environment and diabetes mellitus in South Asia: a geospatial analysis of health outcome data, *PLoS Medicine*, ISSN: 1549-1277

Other publications during this doctoral study (Continued)

Atanasova P, Kusuma D, Pineda E, Anjana RM, De Silva L, Hanif AAM, Hasan M, Hossain MM, Indrawansa S, Jayamanne D, Sujeet Jha, Kasturiratne A, Katulanda P, Khawaja KI, **Kumarendran B**, Mrida MK, Rajakaruna V, Chambers JC, Frost G, Sassi F, Miraldo M (2022) ‘SSM – Population Health Food environments and obesity : A geospatial analysis of the South Asia Biobank , income and sex inequalities’, SSM – Population Health, 17(December 2021), p. 101055. doi: 10.1016/j.ssmph.2022.101055

Bohren MA, Lorencatto F, Coomarasamy A, Althabe F, Devall AJ, Evans C, Oladapo OT, Lissauer D, Akter S, Forbes G, Thomas E, Galadanci H, Qureshi Z, Fawcus S, Hofmeyr GJ, Al-beity FA, Kasturiratne A, **Kumarendran B**, Mammoliti KM, Vogel JP, Gallos I & Miller S. (2021) ‘Formative research to design an implementation strategy for a postpartum hemorrhage initial response treatment bundle (E-MOTIVE): study protocol.’, Reproductive health, 18(1), p. 149. doi: 10.1186/s12978-021-01162-3.<https://reproductive-health-journal.biomedcentral.com/articles/10.1186/s12978-021-01162-3#citeas>

Kusuma, D., Pradeepa, R., Khawaja, K.I., Hasan, M., Siddiqui, S., Mahmood, S., Ali Shah, S.M., De Silva, C.K., de Silva, L., Gamage, M., Loomba, M., Rajakaruna, V.P., AM Hanif, A., Kamalesh, R.B., **Kumarendran. B**, Loh, M., Misra, A., Tassawar, A., Tyagi, A., Waghdhare, S., Burney, S., Ahmad, S., Mohan, V., Sarker, M., Goon, I.Y., Kasturiratne, A., Kooner, J.S., Katulanda, P., Jha, S., Anjana, R.M., Mridha, M.K., Sassi, F., Chambers, J.C., 2021. Low uptake of COVID-19 prevention behaviours and high socioeconomic impact of lockdown measures in South Asia: evidence from a large-scale multi-country surveillance programme. SSM – Popul. Heal. 13, 100751. <https://doi.org/10.1016/j.ssmph.2021.100751>

Other publications during this doctoral study (Continued)

Tiffany E Gooden, Jingya Wang, Alessandra Carvalho Goulart, Ana C Varella, Meihui Tai, Vethanayagan Antony Sheron, Hao Wang, Hui Zhang, Jiaoyue Zhong, **Balachandran Kumarendran**, Krishnarajah Nirantharakumar, Rajendra Surenthirakumaran, Isabela M Bensenor, Yutao Guo, Gregory Y H Lip, G Neil Thomas & Semira Manaseki-Holland (2023) Generalisability of and lessons learned from a mixed-methods study conducted in three low- and middle-income countries to identify care pathways for atrial fibrillation, Global Health Action, 16:1, <https://DOI:10.1080/16549716.2023.2231763>

Uruthirakumar, P., Surenthirakumaran, R., Gooden, T.E. G. Neil Thomas, David J. Moore, Krishnarajah Nirantharakumar, **Balachandran Kumarendran**, Kumaran Subaschandran, Shribavan Kaneshamoorthy, Vethanayagam Antony Sheron & Mahesan Guruparan. The impact of rate and rhythm control strategies on quality of life for patients with atrial fibrillation: a protocol for a systematic review. Syst Rev 12, 52 (2023). <https://doi.org/10.1186/s13643-023-02197-2>

Kanesamoorthy S, Sheron VA, Uruthirakumar P Chamira Kodippily, **Balachandran Kumarendran**, Tiffany E Gooden, Graham Neil Thomas, Krishnarajah Nirantharakumar, Gregory Y H Lip, Mahesan Guruparan, Rashan Haniffa, Rajendra Surenthirakumaran, Abi Beane, Kumaran Subaschandran Prevalence of atrial fibrillation in Northern Sri Lanka: a study protocol for a cross-sectional household survey BMJ Open 2022;12:e056480. <https://doi:10.1136/bmjopen-2021-056480>

Other publications during this doctoral study (Continued)

Santos IS, Lotufo PA, Goulart AC, Brant LCC, Pinto Filho MM, Pereira AC, Barreto SM, Ribeiro ALP, Thomas GN, Lip GYH, Bensenor IM; NIHR Global Health Research Group on Atrial Fibrillation Management; Arasalingam A, Beane A, Bensenor IM, Brocklehurst P, Cheng KK, El-Bouri W, Feng M, Goulart AC, Greenfield S, Guo Y, Guruparan M, Gusso G, Gooden TE, Haniffa R, Humphreys L, Jolly K, Jowett S, **Kumarendran B**, Lancashire E, Lane DA, Li X, Lip Co-Pi GYH, Li YG, Lobban T, Lotufo PA, Manseki-Holland S, Moore DJ, Nirantharakumar K, Olmos RD, Paschoal E, Pirasanth P, Powsiga U, Romagnolli C, Santos IS, Shantsila A, Sheron VA, Shribavan K, Szmigin I, Subaschandren K, Surenthirakumaran R, Tai M, Neil Thomas Co-Pi G, Varella AC, Wang H, Wang J, Zhang H, Zhong J. Cardiovascular Health and Atrial Fibrillation or Flutter: A Cross-Sectional Study from ELSA-Brasil. *Arq Bras Cardiol.* 2022 Sep 23;119(5):724–31. Portuguese, English. <https://doi:10.36660/abc.20210970>. Epub ahead of print. PMID: 36169453; PMCID: PMC9750207.

Rubini, S., **Kumarendran. B.**, Raguram, S., & Sriskandarajah, N. (2023). Newspaper coverage of water security issues in the water-stressed Northern Province of Sri Lanka: An explorative study of three Tamil dailies. *World Water Policy*. <https://doi.org/10.1002/wwp2.12123>

Chapter 1: General Introduction

1.1. Definition of polycystic ovary syndrome

1.1.1. Establishing the diagnosis of Polycystic ovary syndrome

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder among women. The PCOS, though this condition was not named then, was initially presented using a case series of seven women with amenorrhea and bilateral polycystic ovaries (Stein and Leventhal, 1935). Since then, the diagnosis of PCOS has evolved from clinical, morphological, and histological diagnosis and with the evolution of diagnostic facilities, it incorporated pneumoroentgenography, hormonal assay and ultrasonography (Wang and Mol, 2017). There is no single test to diagnose PCOS and a syndromic approach with a combination of more than one feature is used (Azziz et al., 2006). The most widely accepted diagnostic criteria currently is based on presence of two of three features: androgen excess (clinical or biochemical), ovulatory dysfunction and polycystic ovary (Pasquali and Gambineri, 2015; Teede et al., 2018a).

1.1.2. Clinical androgen excess

Clinical androgen excess includes hirsutism, acne, androgenic alopecia in women (female pattern hair loss), and ovulatory dysfunction (Lizneva et al., 2016a). Hirsutism is commonly graded by visual inspection of nine areas using modified Ferriman and Gallwey scoring (Yildiz et al., 2010). Adult acne is also graded by visual inspection using rating scales such as Acne Severity Index (Hacivelioglu et al., 2013), Global Evaluation Acne grading scale, Burke and Cunliffe counting method and adult female acne scoring tool (Carmina et al., 2022). Evaluation of biochemical androgen levels is recommended in women with acne irrespective of the severity of acne and treatment with oral combined estroprogestins if

androgen excess is confirmed (Carmina et al., 2022). Alopecia is classified based on various scales proposed by Hamilton, Ogata, Feit, Setty, Rook and Dawber, Bouhanna, Blanchard and Blanchard, Savin (Bouhanna, 2000), and Ludwig (Ludwig, 1977). Anovulation along with hyper-androgen is highly suggestive of PCOS (Wierman, 2021). Among 881 premenopausal women with chronic anovulation in the UK, 89% were women with PCOS (Elhassan et al., 2018).

1.1.3. Biochemical androgen excess

Biochemical androgen excess is defined as presence of more than normal quantities of serum total testosterone, androstenedione (A4), dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulphate (DHEAS), 11-Oxygenated androgens, dexamethasone suppression testing, GnRH analogue test, short synacthen test and ovarian and adrenal venous sampling (Cussen et al., 2022). Serum total testosterone level is affected by other proteins such as sex hormone binding globulins (SHBG) (Pardridge, 1986). Therefore, serum SHBG is also an important biomarker for predicting PCOS risk (Deswal et al., 2018). Unlike the total serum testosterone, androstenedione (A4) is not affected by SHBG and hence a more reliable marker of androgen excess (O'Reilly et al., 2014). DHEAS/free androgen index ratio predicts a favourable metabolic profile in women with PCOS (Nadir et al., 2019). In addition to the determining androgen excess, 11-Oxygenated androgens (Tosi et al., 2022) differentiate the ovarian and adrenal-specific origin (Turcu and Auchus, 2017).

1.1.4. Chronic anovulation

Chronic anovulation is defined when a woman having fewer than eight menstrual cycle per year (Lujan et al., 2008). Diagnosis of anovulation needs to consider the age of the women as an important aspect (Pasquali and Gambineri, 2015). Adolescent women take around one year to get regular menstruation and menstrual age should be considered in interpreting

anovulation among adolescent women (Rosenfield, 2013). In adult women, prior to attributing the chronic anovulation with PCOS, we need to exclude the other causes for anovulation such as pregnancy, excess weight or excessive exercise, premature ovarian failure subclinical or overt thyroid dysfunction and hyperprolactinemia (Rebar, 2000). The use of oral contraceptive pills may mask chronic anovulation (Pasquali and Gambineri, 2015).

Table 1.1. Potential androgens in PCOS

Androgen	Role in PCOS
Testosterone (T)	Elevated levels in women with hyperandrogenic PCOS. Influences hirsutism and other androgen-related symptoms.
Androstenedione (A4)	A pro-androgen precursor. Elevated levels in hyperandrogenic PCOS.
Dehydroepiandrosterone Sulfate (DHEAS)	Another pro-androgen precursor. Elevated levels in hyperandrogenic PCOS.
11-Oxygenated Androgens	Preponderant circulating androgens in women with PCOS. Their pathophysiological significance is still being studied.
Dihydrotestosterone (DHT)	A potent androgen derived from testosterone. Associated with the development of sex characteristics, bone health, and sex drive.
Androstenediol (A5)	Used to make sex hormones estrogen and testosterone.
Androsterone	A byproduct of testosterone breakdown.

1.1.5. Polycystic ovaries

Polycystic ovaries were initially reported in laparotomies, gynaecological pan-roentgenographies (Stein and Leventhal, 1935) and autopsies (Sommers and Wadman, 1956). Introduction of ultrasound scanning greatly improved the non-invasive diagnosis of polycystic ovaries with a sensitivity of 91% and specificity of 100% when verified against laparoscopic inspection (Fox and Hull, 1993). PCO in ultrasound scan is defined as either 12 or more follicles measuring 2 ± 9 mm in diameter or increased (>10 cm³) ovarian volume

(Balen et al., 2003). The cut-off for ovarian volume and follicle number needs to decrease with increasing age (Kim et al., 2017). Although a transvaginal ultrasound scan may have some discomfort to the patient, it is preferred over a transabdominal ultrasound scan due to its accuracy, high resolution, operator friendly and no need to have full bladder (Goldstein, 1990). In an overweight population, ovarian volume was underestimated by two-dimensional ultrasound scan than three-dimensional ultrasound scan and MRI scan and antral follicle count estimated by dimensional ultrasound scan was better correlated with serum Anti-Müllerian hormone (AMH)-assays followed by MRI and two-dimensional ultrasound scan (Nylander et al., 2017).

1.1.6. Evolution of criteria for the diagnosis of PCOS

Therefore, various criteria have been proposed for the diagnosis of PCOS. The diagnosis of Stein Leventhal syndrome constituted anovulation, infertility, hirsutism, and bilateral polycystic ovaries (Stien, 1952). The first-ever dedicated conference for PCOS was held in 1990 by the National Institutes of Health–National Institute of Child Health and Human Development and developed the NIH Diagnostic Criteria for PCOS which included chronic anovulation and androgen excess while excluding the secondary causes such as adult-onset congenital adrenal hyperplasia, androgen-secreting neoplasms, and hyperprolactinemia (Legro, 2003). Notably, NIH diagnostic criteria did not include polycystic ovaries as one of the criteria. The need to include international consensus was emphasised (Balen and Michelmore, 2002).

Around 15 years later, the second conference was conducted jointly by the European Society of Human Reproduction and Embryology (ESHRE) and The American Society for Reproductive Medicine (ASRM) in Rotterdam and developed the Revised 2003 consensus on diagnostic criteria for PCOS, also commonly known as Rotterdam criteria (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004). In the Rotterdam

criteria, PCOS is diagnosed after exclusion of secondary causes and based on the presence of at least two of the three features namely, menstrual, or ovulatory dysfunction, clinical or biochemical hyperandrogenism and polycystic ovarian morphology on ultrasonography (Legro et al., 2013; Teede et al., 2018b; Conway et al., 2014).

The diagnosis of PCOS results in various consequences to the patients, their siblings and healthcare system with the follow-up care and the need for robust criteria was emphasised. Hence, the Androgen Excess Society released a position statement on the Criteria for Defining Polycystic Ovary Syndrome as a Predominantly Hyperandrogenic Syndrome (Azziz et al., 2006). Combining these three features forms four phenotypes (Conway et al., 2014).

Table 1.1. Criteria for the diagnosis of PCOS

Criteria	NIH Criteria	Rotterdam criteria	Androgen Excess and PCOS Society criteria
Year	1990	2003	2009
Hyperandrogenism	Compulsory	Two of the three	Compulsory
Oligomenorrhea	Compulsory	Two of the three	Oligomenorrhea or polycystic ovaries
Polycystic ovaries	Not required	Two of the three	Oligomenorrhea or polycystic ovaries

1.2. Phenotypes of PCOS

In 2016, phenotypes of PCOS were categorised into Phenotype A (having androgen excess, ovulatory dysfunction and polycystic ovaries), Phenotype B (having androgen excess and ovulatory dysfunction), Phenotype C (having androgen excess and polycystic ovaries) and Phenotype D (having ovulatory dysfunction and polycystic ovaries) (Lizneva et al., 2016b). A recommendation on phenotypic research among women with PCOS is also developed

(Azziz et al., 2019).

1.3. Underlying pathophysiology

The pathophysiology of PCOS involves a self-perpetuating vicious cycle (Witchel et al., 2019) with multifactorial involvement of ovarian dysfunction especially partial folliculogenesis arrest (Azziz et al., 2016), neuroendocrine axis dysfunction (Silva and Campbell, 2022) or gonadotropin derangement or aberrant gonadotropin secretion (McCartney et al., 2022), endocrine (Emanuel et al., 2022), androgen excess (Kempegowda et al., 2020), insulin resistance (Moggetti and Tosi, 2021; Chen et al., 2022c), adipose tissue dysfunction (Azziz, 2021), familial clustering (Lunde et al., 1989), polygenic origin (Khan et al., 2019), omics (Pei et al., 2021), chronic inflammation (Rudnicka et al., 2021), Oxidative stress or antioxidant imbalance (Rudnicka et al., 2022) and environmental factors such as environmental endocrine disruptors (Dunn, 2021), diet and gut microbiota (Tremellen and Pearce, 2012).

The pathophysiology of PCOS evolved greatly over the last century. Stein and Leventhal believed that the bilateral polycystic ovaries were most likely due to the hormonal origin rather than inflammation. They also believed that the mechanical congestion of the cortex of the ovary by cysts was the cause of the amenorrhea and sterility (Stein and Leventhal, 1935). Based on autopsies of 24 women and two children with bilateral polycystic ovaries, Sommers and Wadman attributed pituitary basophilism with FSH hypersecretion possibility under the hypothalamic control as the cause for the changes in the ovaries (Sommers and Wadman, 1956). Hypothalamus-Pituitary-Ovary (HPO) axis regulates the process of ovulation (Berga and Naftolin, 2012).

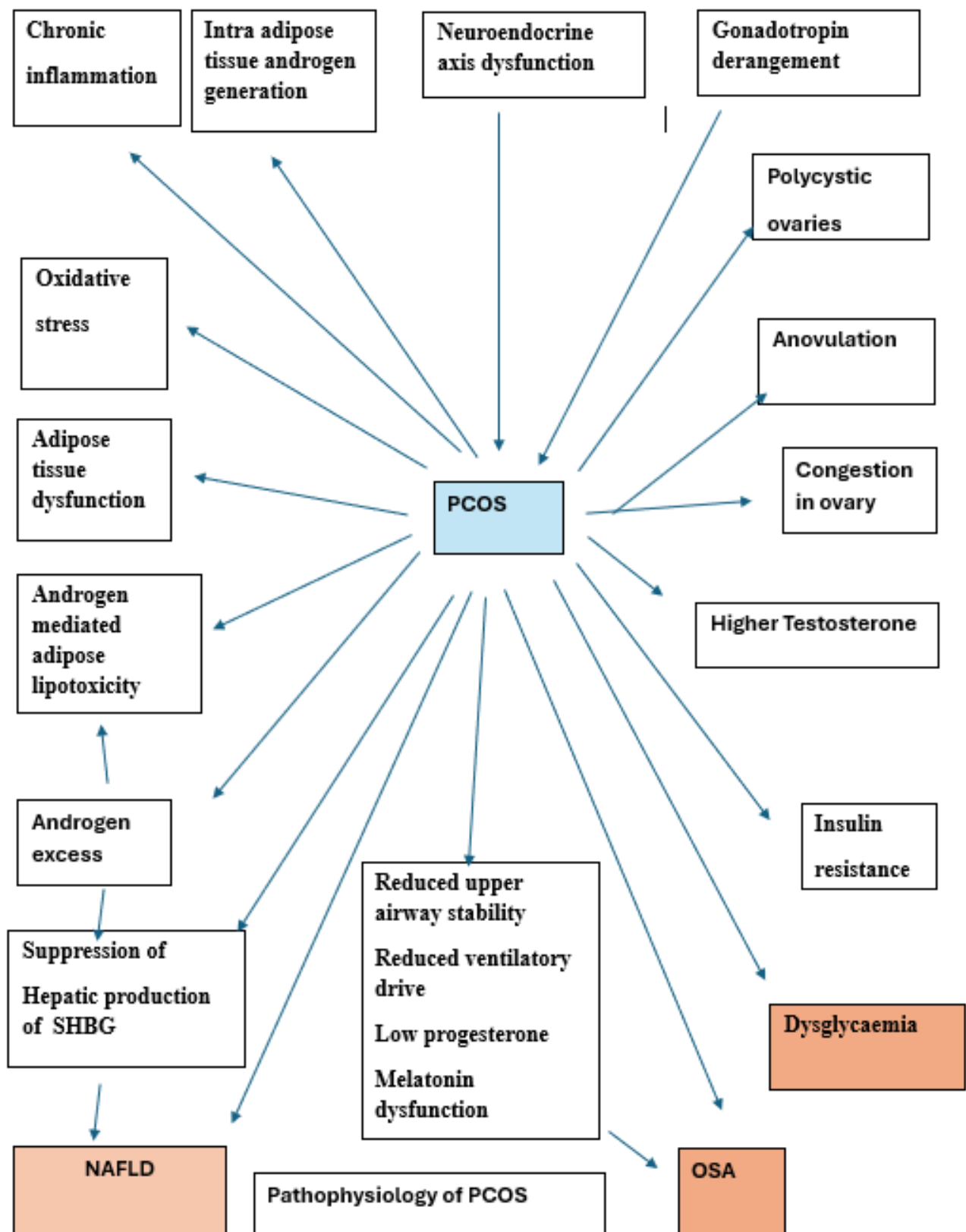


Figure 1 Pathophysiology of PCOS with NAFLD, OSA and dysglycaemia outcomes

1.4. Epidemiology of PCOS

1.4.1. Prevalence of PCOS

Reproductive aged women of all races and ethnicities are affected by PCOS (Wolf et al., 2018). Prevalence of PCOS is 6% to 10% (Bozdag et al., 2016). Prevalence of PCOS should consider the diagnostic criteria used in the estimation as the PCOS prevalence varied from 6.2% through 15% to 19.5% in a meta-analysis of studies conducted in the Europe and US using NIH 1990, ES-PCOS diagnostic criteria and ESHRE/ASRM 2003 criteria respectively (Chiaffarino et al., 2022). This has also implication for the interpretation of trend of PCOS prevalence. Repetitive surveys using same diagnostic criteria from 2010 to 2020 have also shown an increase in the prevalence of PCOS in all age groups from 5.6% to 8.6% in China (Yang et al., 2022).

1.4.2. Incidence of PCOS

In the three decades since 1990, incident cases of PCOS increased by 54% and age standardised incidence rates also 46 per 100,000 to 60 per 100,000 (Gao et al., 2023).

1.4.3. Age related distribution and natural progression of PCOS

Highest incidence of PCOS was observed among 15 to 19 age group followed by 10 to 14 years (Gao et al., 2023). Prevalence of PCOS declined over the age in the China Fertility Survey of Married Women although the findings are limited by restriction to married women (Yang et al., 2022). Diagnosis of PCOS was made at the mean age of 26.9 years in a population based retrospective cohort in the Washington, United States (Yu et al., 2023).

Phenotype of PCOS among the women visiting outpatient clinic repeatedly in a panel study varied over the age. Phenotype A (Ovulatory dysfunction, hyperandrogenism and polycystic ovaries) was the predominant phenotype in twenties, and it declined over the age sharply.

Phenotype C (Hyperandrogenism and polycystic ovaries) was the least frequent among twenties, and its proportion sharply increased to become most common phenotype among fifties (van Keizerswaard et al., 2022). Although, this study among clinic participants would have been affected by selection bias, this may reflect the change in the indication for clinic visit among the women with PCOS with increasing age (Kim and Wild, 2022).

Lean women with PCOS experienced an age dependent gradual improvement in insulin resistance, androgen levels, and lipid profiles with advancing age in a cohort of Caucasian women (Livadas et al., 2020). Observation of similar findings in a Brazilian cohort supported the need for the development of an age-adjusted diagnostic criteria for PCOS (de Medeiros et al., 2020).

1.4.4. Ethnicity-related distribution of PCOS

Comparison of PCOS distribution by ethnicity is not a straightforward process as many aspects of PCOS vary greatly by ethnicity, such as clinical and biochemical androgen excess, anovulation, ovarian morphology, and psychological conditions.

A systematic review of 35 articles which summarised the association between ethnicity and PCOS, subdivided PCOS into many ethnicity-based categories, namely black (bPCOS), East Asian (eaPCOS), Hispanic (hPCOS), Middle Eastern (mePCOS), South Asian (saPCOS) and White (wPCOS) (Sendur and Yildiz, 2021). Among the women with PCOS and when compared to white women (wPCOS), hirsutism was less among east Asian women (eaPCOS) and more among Hispanic (hPCOS), South Asian (saPCOS) and Middle Eastern women (mePCOS) (Table 2.1). A higher cut-off for modified Ferriman and Gallwey scoring is recommended for Hispanic and Middle Eastern women while a lower cut-off for East Asian women (Chan et al., 2017).

Table 1.2. Distribution of PCOS features by ethnicity

Ethnicity	Black	East Asian	Hispanic	Middle Eastern	South Asian	White
Abbreviation for ethnicity	(bPCOS)	(eaPCOS)	(hPCOS)	(mePCOS)	(saPCOS)	(wPCOS)
Hirsutism	More	Less	More	More	More	Ref*
Modified FG Scoring Cut-off	Lower	Lower	Higher	Higher	Lower	Ref
Androgen Concentrations (DHEAS)	Higher	Higher	-	Higher	Higher	Ref
SHBG Concentrations	Lower	Lower	Lower	Lower	Lower	Ref
BMI and Central Obesity	Higher central obesity	Lower BMI	-	-	Lower BMI	Ref
Insulin Resistance	Higher	Higher	Higher	Higher	Higher	Ref
Lipid Profile	Better	-	-	-	-	Ref
Blood Pressure and Cardiovascular Risk	Increased risk for higher BP	-	-	-	-	Ref

* Reference level

Higher androgen and lower sex hormone-binding globulin (SHBG) concentrations were observed among saPCOS. Compared to wPCOS, 1) DHEAS concentrations were higher among mePCOS, 2) lower SHBG and DHEAS measures were observed among hPCOS and

bPCOS. Lower BMI and higher central obesity were documented among saPCOS and eaPCOS while the BMI was higher among hPCOS and bPCOS. Compared to wPCOS, the prevalence of insulin resistance was higher among saPCOS, mePCOS, hPCOS and bPCOS. Although bPCOS had a better lipid profile, they had a higher blood pressure and increased cardiovascular risk.

Although hirsutism is more prominent among South Indian and Middle Eastern women when compared to East Asian and Caucasian women (VanHise et al., 2023), circulating levels of androgen demonstrates a weaker correlation with the severity of hirsutism (Legro et al., 2010). Variation in the presentation of each clinical androgen features is believed to be mediated by tissue-specific variations in the dysregulation threshold of biochemical androgen excess (Dong & Rees, 2023).

Although polycystic ovarian morphology did not show ethnic variation in many studies, prevalence of polycystic ovarian morphology was higher among Caucasian women than Middle Eastern women (Glintborg et al., 2010) and it varied among various subgroups of Caucasian women (Chan et al., 2017).

East Asian women had more frequent menstrual irregularities (Wang et al., 2013) while South Asian women tend to have at a lower age (Wijeyaratne et al., 2002). Early onset of symptoms and delayed health seeking is observed among ethnic Indian women globally (Rao et al., 2022).

When compared with wPCOS, obesity was higher among hPCOS and bPCOS, and was lower among Asian women with PCOS; diabetes was higher among Asian women with PCOS and hPCOS; hypertension was higher among bPCOS and lower among hPCOS (Lo et al., 2006).

Diabetes risk was higher among South Asian and Filipino women than Chinese women (Guo

et al., 2021). Overall cardiometabolic risk was higher for Black women in the US than the White women (Kazemi et al., 2021).

Emotional and sexual dysfunction was relatively higher among non-white women and women born in Indian. At the same time, body image concerns and weight stigma were relatively higher among white women and women born in the UK (Sheikh et al., 2023). White women with PCOS had a relatively higher prevalence of anxiety than the black women with PCOS (Alur- Gupta et al., 2021).

1.5. Factors associated with PCOS or cardiometabolic conditions.

1.5.1. Body mass index

Obesity has been attributed to causing cardiovascular complications in women with PCOS. Overweight and obese women had reduced levels of Sex Hormone Binding Globulin (SHBG) and increased levels of total testosterone with increased insulin resistance features (Lim et al., 2013). Recent reviews confirm that non-obese women with PCO are also at an increased risk of long-term complications of cardiometabolic conditions (Zhu et al., 2019). A recent review proposes diabetes, polycystic ovarian syndrome, obstructive sleep apnoea, and obesity as an emerging theme and iterates the need to focus on the distribution of fat such as visceral adiposity and adiposity around the neck rather than using the conventional measure of BMI (Bambhroliya et al., 2022).

1.5.2. Thyroid conditions

Incidence of thyroid conditions were remarkably and significantly higher among a Danish cohort of 18,476 women with PCOS (Glintborg et al., 2019). Mean TSH levels were higher in adolescents with PCOS than those without PCOS (Garcia-Beltran et al., 2023). Higher TSH level was associated with hyperandrogenism in euthyroid women with PCOS

independent of age, BMI and thyroid autoimmunity (Cai et al., 2019).

Early evidence from nine studies showed increased levels of total cholesterol and fasting glucose in women with comorbidity of subclinical hypothyroidism and PCOS when compared with women with PCOS but euthyroid (Xing et al., 2021; De Medeiros et al., 2018). Besides the adverse impact on glucose and lipid metabolism, subclinical hypothyroidism in PCOS women increases the risk of fertility issues (Peddemul et al., 2022). Women with PCOS had a 2.9 times higher risk of having subclinical hypothyroidism when compared to women without PCOS (Ding et al., 2018b). Autoimmune thyroid disease was 3.3 times higher among women with PCOS and the risk varied based on ethnicity with descending risk among South Asian (OR=4.6), Europeans (OR=3.3) and South Americans (OR=1.9) (Romitti et al., 2018). Therefore, it is recommended not to exclude a diagnosis of PCOS in women with subclinical hypothyroidism (Freitas De-Medeiros et al., 2017).

1.5.3. Social deprivation or socioeconomic condition

Lower socioeconomic status was associated with PCOS while among the women with PCOS, women with higher BMI and women of foreign origin had lower socioeconomic status in a register-based Danish cohort (Rubin et al., 2019). The prevalence of PCOS was relatively higher among those with lower parental education (Merkin et al., 2011). Neighbourhood conditions were associated with overweight and obesity among youth in Ireland and the UK (Cronin et al., 2022). A systematic review in the UK observed a trend towards an association between socioeconomic status and cardiometabolic conditions such as cardiovascular diseases, diabetes and obesity among the migrants in the UK (Ismail et al., 2022). Obstructive sleep apnoea risk was higher among low socioeconomic group irrespective of age and sex (EtindeleSosso and Matos, 2021). Risk of NAFLD was higher among the low socioeconomic group in Korea (Cho et al., 2021). Early life economic inequalities were associated with diabetes outcome among the participants of The Maastricht Study (Derks et al., 2017).

1.6. Existing evidence on PCOS and cardiometabolic conditions

1.6.1. PCOS and Metabolic syndrome

Metabolic syndrome is commonly diagnosed using at least any three of the following five risk factors namely dysglycaemia, obesity (specifically central adiposity), hypertension, increased serum triglycerides level and low high-density lipoprotein-cholesterol levels (Nwankwo et al., 2022). Adolescents with PCOS had three times higher odds of having metabolic syndrome than their controls (Fu et al., 2023b). Metabolic syndrome demonstrating a familial clustering among the first-degree relatives of women with PCOS (Yilmaz et al., 2018) denotes the strong genetic predisposition of these two conditions.

A meta-analysis published prior to our study reported that the increased risk for metabolic syndrome was associated with obesity and metabolic features and not with the androgen excess (Lim et al., 2019). Hence, our study aimed in determining the role of androgen excess in the association between PCOS and cardiometabolic conditions.

1.6.2. Non-Alcoholic Fatty Liver Disease

The prevalence of NAFLD is 29.4% in the general population as estimated in a meta-analysis of 578 studies and the prevalence remarkably increases in overweight (69.99%) and in obese (75.27%) as reported in a meta-analysis of 151 studies (Quek et al., 2023). Although the prevalence of NAFLD was lower among women (25.6%) compared to men (39.7%) (Riazi et al., 2022), women have increased risk of advanced fibrosis than men (Balakrishnan, 2021). Incidence of NAFLD is estimated to be 46 new cases per 1000 person-years and the incidence rate has increased by three-fold between 2000 and 2015 (Le et al., 2023).

Occurrence of NAFLD is associated with tamoxifen use (Lee et al., 2020), a dose-dependent increase in consumption of sugar sweeteners (Chen et al., 2019) which might have been a confounder with fat intake, a dose dependent decrease in physical activity (Qiu et al., 2017),

NAFLD was associated with increased risk of many cardiometabolic condition, namely, all-cause, CVD and cancer related mortality in a meta-analysis of 10 million patients (Fu et al., 2023a), stroke (Tang et al., 2022), incident heart failure (Li et al., 2022), incident hypertension (Ciardullo et al., 2022), chronic kidney disease (Chen et al., 2023), gestational diabetes independent of the risk of development of T2DM (Chai et al., 2021; Lavrentaki et al., 2019), and incident diabetes (Mantovani et al., 2021).

PCOS and NAFLD are expected to have a common origin, possibly be through hyperandrogenism, as PCOS is considered an ovarian expression of metabolic syndrome.

In contrast, NAFLD is considered a hepatic expression of metabolic syndrome

(Baranova et al., 2011). NAFLD was associated with an 80% increased risk for incident metabolic syndrome (Ballestri et al., 2016).

1.6.3. Obstructive Sleep Apnoea

Obstructive sleep apnoea is apnoea during sleep due to upper airway obstruction. Although prevalence of OSA is estimated as 54% (de Araujo Dantas et al., 2023) and 55% (Aini et al., 2023) in two meta-analyses, their validity is compromised due to the incorporation of both diagnostic and screening estimates and non-standardisation of age. Globally, moderate to severe OSA is estimated to be 425 million (399–450) among adults aged 30–69 years (Benjafeld et al., 2019). The prevalence of OSA among women with PCOS is 35% (Kahal et al., 2020). A Markov model estimation in lifetables of the Australian working age (20 to 65 years) population reported that a 10% reduction of OSA prevalence would result in the prevention of 16,701 deaths, 47,542 cardiovascular events, 88,583 episodes of major depression, and 476 vehicle-related accidents with more than AU\$25 billion gain in the gross domestic product (Marquina et al., 2023).

The gold standard option to diagnose OSA is level 1 polysomnography and for convenience

and to cost-effectiveness (Natsky et al., 2022) levels II to IV, such as portable devices (Shayeb et al., 2014) and non-contact methods (Khalil et al., 2023) are used. Methods used for screening for OSA include questionnaires (Bernhardt et al., 2022) such as the Berlin questionnaire, Epworth Sleepiness Scale (Chiu et al., 2017), STOP-Bang questionnaire, noSAS score (Chen et al., 2022a), Multivariable Apnea Prediction (MVAP) score (Feltner et al., 2022), smartphone apps (Kim et al., 2022),

The performance of many screening techniques in children is improved using machine learning techniques (Gutiérrez-Tobal et al., 2022). The use of peripheral arterial tonometry is unsuitable for screening of OSA due to its lower sensitivity and specificity (Ichikawa et al., 2022).

OSA is linked with diverse conditions involving cancers, respiratory, cardiovascular, and nervous systems and head and neck conditions such as teeth, ear, and eye. OSA is associated with smoking (Zeng et al., 2023), second-hand smoking (Chang et al., 2022), lung cancer (Cheong et al., 2022), increased post-operative complications (Sun et al., 2022), risk factors for coronary atherosclerosis (Chen et al., 2022b), diastolic dysfunction independent of conventional cardiovascular risk factors (Al-Sadawi et al., 2022), atrial fibrillation (Zhang et al., 2022a), hypertension (Hou et al., 2018), higher incidence of colorectal cancers (Teo et al., 2022), various ocular manifestations (Bulloch et al., 2023), increased Floppy Eye Lid syndrome (Cheong et al., 2023), higher incidence of hearing loss (Wang et al., 2022), periodontitis (Molina et al., 2023), wider red cell distribution width (Di Lorenzo et al., 2023), increased prevalence of secondary polycythaemia (Martelli et al., 2022), risk of car accidents (Luzzi et al., 2022) and dementia (Guay-Gagnon et al., 2022).

Sleep pattern has a greater impact on the neuroendocrinology and reproduction in females (Beroukhim et al., 2022). Circadian rhythm (Cambras et al., 2021) and autonomic nervous system (Zhao and Guan, 2018) are likely to be associated with both OSA, metabolic

syndrome and type 2 diabetes. The risk of OSA was higher among women with PCOS (Helvaciet al., 2017).

A review of the link between OSA and PCOS concluded that the existing literature provides association but not causation and suggested using big data in exploring the initial steps of establishing the causation (Mokhlesi et al., 2017).

1.6.4. Dysglycaemia

WHO diagnostic criteria for diabetes (World Health Organization and International Diabetes Federation, 2006) defines diabetes as fasting plasma glucose ≥ 7.0 mmol/l or 2-hour plasma glucose ≥ 11.1 mmol/l; Impaired Fasting Glucose (IFG) by fasting plasma glucose of 6.1 to 6.9 mmol/l and (if measured) 2-hour plasma glucose of < 7.8 mmol/l; Impaired Glucose Tolerance (IGT) by fasting plasma glucose of < 7.0 mmol/l and 2-hour plasma glucose of ≥ 7.8 and < 11.1 mmol/l. When HbA1c is used, 6.5% is used as the cut-point and diabetes is not excluded by a value less than 6.5% (World Health Organisation, 2011). Prediabetes is the status in which the person has glucose level above the normal level but below the cut-point to diagnose diabetes and having a higher risk of progression to diabetes and to get other cardiovascular complications (Echouffo-Tcheugui et al., 2023). Prediabetes is denoted by Impaired Fasting Glucose (IFG) or Impaired Glucose Tolerance (IGT) or Impaired Glucose Regulation (IGR) or non-diabetic hyperglycaemia (Rooney et al., 2023). We developed a working definition of dysglycaemia by combining both diabetes and prediabetes.

While diabetes is prevalent among 9.3% and prediabetes is among 7.5%, only 50% of the patients with diabetes are aware of their disease status (Saeedi et al., 2019). Global prevalence of IGT and IFG among adults was 9.1% and 5.8% (Rooney et al., 2023). One in ten persons with prediabetes progressed to diabetes in one year (Echouffo-Tcheugui et al., 2023). Developing a combined definition of dysglycaemia is clinically justifiable due to 1) increased likelihood of progression from prediabetes to diabetes, 2) evidence to support

adverse consequences during prediabetes status rather than being a mere precursor, and 3) benefit of lifestyle and medication interventions (Lawal et al., 2020).

Diabetes is notorious for its cardiovascular (microvascular and macrovascular) and other complications. Of the total direct cost of management of diabetes, 20 to 49% is spent on managing cardiovascular complications (Einarson et al., 2018). Diabetic retinopathy is prevalent among 22.2% of the diabetes patients globally and the highest (35.9%) in Africa (Teo et al., 2021). Peripheral neuropathy is prevalent among 30% of diabetic patients (Sun et al., 2020).

1.6.5. Other cardiometabolic conditions

Reproductive-aged women with PCOS had 70% higher risk of hypertension than their healthy controls (Amiri et al., 2020). Plasma lipoproteins were higher among women with PCOS (Masson et al., 2023).

1.7. PCOS and other health conditions

1.7.1. PCOS and cancers

Compared to women without PCOS, women with PCOS had a 40% increased risk of composite of breast, endometrial and ovarian and cancers. When these cancers are considered individually, an increased risk was observed for endometrial (OR=2.2, 95%CI: 1.03 – 4.7) and ovarian cancers (OR=1.3, 95%CI: 1.0 – 1.8) while the difference was not significant for breast cancer (Barry et al., 2014).

1.7.2. PCOS and mental health conditions

When compared to women without PCOS, women with PCOS had increased odds of any depressive symptoms (OR=3.8; 95% CI: 3.0 - 4.7), moderate/severe depressive symptoms (OR=4.2; 95%CI: 2.7 – 6.5), any anxiety symptoms (OR=5.6; 95%CI: 3.2-9.8) and severe/

moderate anxiety symptoms (OR= 6.6; 95%CI:2.9 – 14.3). The odds remained higher in matched studies as well (Cooney et al., 2017). PCOS and miscellaneous health conditions

Microbial alpha diversity was relatively lower among women with PCOS (Sola-Leyva et al., 2023). Increased bone mineral density and reduced fracture risk were observed in a Mendelian randomisation analysis among the women with PCOS who had a genetic risk of excess testosterone (Deshmukh et al., 2022).

1.8. Management of PCOS

PCOS management is commonly categorised into medications and lifestyle interventions.

1.8.1. Medications in the management of PCOS

Numerous drugs are presently accessible for the treatment of PCOS. Combined oral contraceptive (COCP) perform better in improving hormonal parameters than metformin, while they are inferior to metformin in improving metabolic parameters. However, combination of both COCP and metformin performs well in terms of metabolic parameters, and COCP alone remains effective for hormonal parameters (Melin et al., 2024).

In a meta-analysis of six randomised controlled trials, Liraglutide and metformin effectively improved weight and glucose management among PCOS women (Ge et al., 2022; Tian et al., 2022). Glucagon-like peptide-1 receptor agonists (GLP-1 RA) either alone or combined with metformin were superior in the weight management among overweight/obese women with PCOS than metformin alone (Lyu et al., 2021). Resveratrol reduces the testosterone, luteinizing hormone and dehydroepiandrosterone compared to placebo (Fadlalmola et al., 2023).

A combination of oral contraceptive pill (OCP) and orlistat was more effective in improving the hormonal, cardiometabolic and pregnancy outcomes than OCP alone in a meta-analysis of eight studies (Chen and Cai, 2022). A network meta-analysis of 101 randomised

controlled trials demonstrated that flutamide was superior in improving both hormonal and metabolic parameters while the combination of cyproterone-acetate and ethinylestradiol showing higher efficacy improving hormonal parameters only (Abdel- Maboud et al., 2021). Combination treatment of oral contraceptive such as 2 mg cyproterone acetate plus 0.05 mg ethinylestradiol (EE/CA) or 3 mg drospirenone plus 0.03 mg ethinylestradiol (EE/DRSP) with metformin or lifestyle management was effective in reducing adverse effects on the glucose and lipid metabolism when compared to treating with oral contraceptive pills alone in a network meta-analysis (Wang et al., 2019).

1.8.2. Surgical interventions in the management of PCOS

Sleeve gastrectomy appears to reduce menstrual irregularities, total testosterone levels and BMI while increasing the SHBG and HDL (Luo et al., 2023). Bariatric surgery reduces the symptoms among obese women with PCOS (Tian et al., 2021).

1.8.3. Lifestyle interventions in the management of PCOS

Physical activity alone or combined with other interventions improves the reproductive outcomes in women with PCOS (Kelly dos Santos et al., 2020).

Green tea (Colonetti et al., 2022), and cinnamon (Keramati et al., 2022) seem to reduce the bodyweight among the women with PCOS. In a meta-analysis of six RCTs, probiotics showed a beneficial effect in improving the insulin and lipid parameters in women with PCOS (Liao et al., 2018).

1.9. Quality of life of women with PCOS

PCOS adversely impacts on various domains of Quality of Life (QoL) such as physical, psychological, sexual, social and medical aspects (Behboodi Moghadam et al., 2018). Health related quality of life (HRQoL) was significantly reduced among adolescents (Kaczmarek et

al., 2016). Hirsutism and menstruation are the most affected domains in the HRQoL (Bazarganipour et al., 2015). Women with PCOS scored lower scores for all domains of Short Form 36 (SF-36) questionnaire, with a marked reduction in the emotional role function (Li et al., 2011). Quality of life (QoL) was remarkably lower among women with PCOS and was comparable to the QoL among women with other chronic conditions (Rempert et al., 2023).

1.10. Health economic burden of PCOS

In the three decades since 1990, DALYs for PCOS increased by 91%. Although incidence rates were higher for 10 to 19 age group, DALYs were highest for 25 to 29 age group followed by 30 to 34 age group (Gao et al., 2023). Middle socio-demographic index (SDI) and high-middle SDI regions showed highest rise in the incidence rate between 2007 and 2017 (Liu et al., 2021).

In the United States, overall health-economic burden of PCOS was 7.9 billion USD in 2020 with half of this contributed by reproductive conditions and the other half by the direct cost of stroke and T2DM (Riestenberg et al., 2022). Increased risk for cardiovascular outcomes was observed among women with a history of either menstrual cycle irregularity or frequent or infrequent cycles in a UK-based cohort (Okoth et al., 2023).

1.11. Mortality among women with PCOS

Cardiovascular mortality was nearly 38% higher among the reproductive-aged women with PCOS in community-based cohort studies than among their healthy controls (Ramezani Tehrani et al., 2020). Although a Swedish cohort study with a mean age of 81 years for the participants reported no evidence of an association with overall and cardiovascular disease-specific mortality, this study included only 35 women with PCOS and likely to be under-powered (Forslund et al., 2022). A systematic review of 237,647 subjects from nine cohort

studies found no difference in the all-cause mortality rate between women with and without PCOS (Zhou et al., 2017). Anyhow, a Dutch population cohort with 62 693 person-years and 302 deaths demonstrated a 50% increased risk among the diabetic mothers of women with PCOS when compared with diabetic mothers of women without PCOS (Louwers et al., 2014). In another larger Californian cohort with 666 CVD deaths during 456,298.5 person-years of follow-up, 42% increased cardiovascular mortality was observed among the women with irregular menstrual period compared to women with regular menstrual periods (Wang et al., 2011).

1.12. PCOS and Routine data studies

The Health Improvement Data (THIN) and Clinical Practice Research Datalink are two of the major general practice data for research with major overlap (Cai et al., 2012; Carbonari et al., 2015).

UK Clinical Practice Research Datalink GOLD database has been used to study the association between weight change and 13 obesity-related complications including PCOS (MedSci et al., 2023), the likelihood of pregnancy (Haas et al., 2023), risk of adverse pregnancy outcome (Subramanian et al., 2022b), screening of T2DM among women with past GDM (Vounzoulaki et al., 2022), PCOS as an associated factor with prescription of levothyroxine among pregnant women with subclinical hypothyroidism (Yu et al., 2021), risk of cardiovascular events (Berni et al., 2021), association with autism (Cherskov et al., 2018) and adverse mental Health and neurodevelopmental outcomes (Berni et al., 2018),

U.K. Clinical Practice Research Datalink (CPRD) contributed to the study burden of hidradenitis suppurativa with PCOS as one of the associated factors (Ingram et al., 2018), response to treatment among T2DM with exclusion of PCOS (Rodgers et al., 2017), reproductive outcomes (Rees et al., 2016), adverse outcome in young women with PCOS

(Morgan et al., 2012), risk of ovarian cancer with PCOS as one of the covariates (Bodmer et al., 2011), venous thromboembolism associated with cyproterone acetate in combination with ethinyloestradiol (Dianette) (Seaman et al., 2004), risk of liver disorder in prescribing of cyproterone acetate in combination with ethinyloestradiol (Dianette) (Seaman et al., 2003b) and use of combined oral contraceptives amongst women with and without acne (Seaman et al., 2003a).

QResearch database has been used to study PCOS such as in developing and validating risk prediction algorithms to estimate the future risk of T2DM with PCOS as one of the risk factors (Hippisley-Cox and Coupland, 2017).

1.13. PCOS in the UK

The prevalence rate of PCOS in the UK was relatively lower than the average European estimate in the 2016 Global Burden of Diseases study (Miazgowski et al., 2021). The prevalence (2%) and incidence (1.84 per 1000 PYAR) of PCOS estimated using the THIN database were lower than the estimates from community-based studies (Ding et al., 2016).

Among 1797 women with PCOS extracted from a UK pooled general practice database, PCOS diagnosis was made commonly among 20-29 years. Most of those who underwent pelvic ultrasound scans, majority had it one month ahead of the diagnosis. Nearly one-third of them had serum total testosterone (30.5%) and SHBG (29.6%) measured within an year ahead of diagnosis (Heald et al., 2018).

The likelihood of getting pregnancy within three years since diagnosis was lower among women with PCOS compared to women without PCOS in a study which extracted 9955 women from UK Clinical Practice Research Datalink GOLD database (Haas et al., 2023). The risk of preterm and caesarean deliveries was higher among 27,586 deliveries with maternal PCOD when compared with 110,344 deliveries without PCOS in a study based on

Clinical Practice Research Datalink (CPRD) GOLD pregnancy register and linked Hospital Episodes Statistic (HES) data (Subramanian et al., 2022b).

Higher levels of psychological distress and difficulties with coping were observed in a qualitative study among women with PCOS recruited from ethnically diverse UK communities (Hadjiconstantinou et al., 2017).

In a THIN database study among those with their first prescription of cyproterone acetate/ethinylestradiol (CPA/EE), around 62% had hyperandrogenic conditions and around 54% had acne. Manual verification of records identified acne among 77% of the women. Use of another hormonal contraceptive along with CPA/EE was rare (Cea Soriano et al., 2017a).

A considerable proportion of other UK-based studies among women with PCOS in the UK were small-sample studies. Increased cardiovascular risk profile was observed in both 108 UK women with PCOS and 97 Qatar women compared to their 69 UK and 622 Qatar controls. BMI and blood pressure were relatively higher among UK women with PCOS while testosterone was higher among women in Qatar with PCOS (Butler et al., 2020).

1.14. Objectives

1.14.1. Objectives of study with NAFLD outcome

- 1) To determine the unadjusted risk of prevalent and incident NAFLD among women with PCOS when compared with women without PCOS and to derive the risk estimates adjusted for age, Townsend score, BMI, diabetes, or impaired glucose regulation up to end of follow-up, and hypothyroidism at baseline.
- 2) To determine if the risk is independent of their BMI status
- 3) To determine the factors associated with risk of NAFLD among the women with PCOS
- 4) To determine the risk of NAFLD by different levels of serum testosterone and SHBG in independent cohorts.

1.14.2. Objectives of study with OSA outcome

- 1) To determine the unadjusted risk of prevalent and incident OSA among women with PCOS when compared with women without PCOS and to derive the risk estimates adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation up to end of follow-up, and hypothyroidism at baseline.
- 2) To determine if the risk is independent of their BMI status
- 3) To determine the factors associated with risk of OSA among the women with PCOS

1.14.3. Objectives of study with dysglycaemia outcome

- 1) To determine the unadjusted risk of prevalent and incident dysglycaemia among women with PCOS when compared with women without PCOS and to derive the risk estimates adjusted for age, Townsend score, BMI, ethnicity, smoking status, hypothyroidism, hypothyroidism, and prescription with lipid lowering medications.
- 2) To determine if the risk is independent of their BMI status
- 3) To determine the factors associated with risk of dysglycaemia among the women with PCOS
- 4) To determine the risk of dysglycaemia by different levels of serum testosterone and SHBG in independent cohorts.

1.15. References of Chapter 1: General Introduction

Abd Elhafeez, S., D'Arrigo, G., Leonardis, D., et al. (2021) Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. *Oxidative Medicine and Cellular Longevity*, 2021. doi:10.1155/2021/1302811.

Abdel-Maboud, M., Menshawy, A., Hasabo, E.A., et al. (2021) The comparative effectiveness of 55 interventions in obese patients with polycystic ovary syndrome: A network meta-analysis of 101 randomized trials. *PloS one*, 16 (7). doi:10.1371/JOURNAL.PONE.0254412.

Agapova, S.E., Cameo, T., Sopher, A.B., et al. (2014) Diagnosis and challenges of polycystic ovary syndrome in adolescence. *Seminars in Reproductive Medicine*, 32 (3). doi:10.1055/s-0034-1371091.

Ahmadi, M. and Aslani, N. (2018) Capabilities and advantages of cloud computing in the implementation of Electronic Health Record. *Acta Informatica Medica*, 26 (1). doi:10.5455/aim.2018.26.24-28.

Aini, N., Chu, H., Banda, K.J., et al. (2023) Prevalence of sleep-related breathing disorders and associated risk factors among people with dementia: A meta-analysis. *Sleep medicine*, 103: 51–61. doi:10.1016/J.SLEEP.2023.01.020.

Alexander, M., Loomis, A.K., Fairburn-Beech, J., et al. (2018) Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Medicine*, 16 (1). doi:10.1186/s12916-018-1103-x.

Al-Sadawi, M., Theodoropoulos, K., Saeidifard, F., et al. (2022) Sleep Apnea as a Risk Factor for Diastolic Dysfunction: A Systematic Review and Meta-Analysis. *Respiration; international review of thoracic diseases*, 101 (11): 1051–1068. doi:10.1159/000525782.

Alshehri, Z., Subramanian, A., Adderley, N.J., et al. (2022) Risk of incident obstructive sleep

apnoea in patients with type 1 diabetes: a population-based retrospective cohort study. *Diabetologia*, 65 (8). doi:10.1007/s00125-022-05714-5.

Alur-Gupta, S., Lee, I., Chemerinski, A., et al. (2021) Racial differences in anxiety, depression, and quality of life in women with polycystic ovary syndrome. *F and S Reports*, 2 (2). doi:10.1016/j.xfre.2021.03.003.

Amiri, M., Ramezani Tehrani, F., Behboudi-Gandevani, S., et al. (2020) Risk of hypertension in women with polycystic ovary syndrome: a systematic review, meta-analysis and meta-regression. *Reproductive biology and endocrinology: RB&E*, 18 (1). doi:10.1186/S12958-020-00576-1.

de Araujo Dantas, A.B., Gonçalves, F.M., Martins, A.A., et al. (2023) Worldwide prevalence and associated risk factors of obstructive sleep apnea: a meta-analysis and meta-regression. *Sleep & breathing = Schlaf & Atmung*. doi:10.1007/S11325-023-02810-7.

Ashinoff, B.K., Buck, J., Woodford, M., et al. (2022) The effects of base rate neglect on sequential belief updating and real-world beliefs. *PLoS Computational Biology*, 18 (12). doi:10.1371/journal.pcbi.1010796.

Austin, L., Sharp, C.A., van der Veer, S.N., et al. (2020) Providing “the bigger picture”: Benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. *Rheumatology (United Kingdom)*, 59 (2). doi:10.1093/rheumatology/kez207.

Awaysheh, A., Wilcke, J., Elvinger, F., et al. (2018) A review of medical terminology standards and structured reporting. *Journal of Veterinary Diagnostic Investigation*. 30 (1). doi:10.1177/1040638717738276.

Azziz, R. (2021) Adiposity in polycystic ovary syndrome: excess versus dysfunction. *Fertility and Sterility*. 116 (1) pp. 87–88. doi:10.1016/j.fertnstert.2021.04.033.

Azziz, R., Carmina, E., Chen, Z., et al. (2016) Polycystic ovary syndrome. *Nature Reviews Disease Primers*. 2. doi:10.1038/nrdp.2016.57.

Azziz, R., Carmina, E., Dewailly, D., et al. (2006) Position statement: Criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: An androgen excess society guideline. *Journal of Clinical Endocrinology and Metabolism*, 91 (11): 4237–4245. doi:10.1210/jc.2006-0178.

Azziz, R., Kintziger, K., Li, R., et al. (2019) Recommendations for epidemiologic and phenotypic research in polycystic ovary syndrome: an androgen excess and PCOS society resource. *Human reproduction (Oxford, England)*, 34 (11): 2254–2265. doi:10.1093/HUMREP/DEZ185.

Balakrishnan, M. (2021) Women Have a Lower Risk of Nonalcoholic Fatty Liver Disease but Higher Risk of Nonalcoholic Fatty Liver Disease Fibrosis Than Men: Summary of the Findings of a Systematic Review and Meta-analysis. *Clinical Liver Disease*, 18 (5): 251. doi:10.1002/CLD.1139.

Balen, A., Laven, J.S.E., Tan, S.L., et al. (2003) Ultrasound assessment of the polycystic ovary: International consensus definitions. *Human Reproduction Update*. 9 (6) pp. 505–514. doi:10.1093/humupd/dmg044.

Balen, A. and Michelmore, K. (2002) What is polycystic ovary syndrome? Are national views important? *Human Reproduction*. 17 (9) pp. 2219–2227. doi:10.1093/humrep/17.9.2219.

Ballestri, S., Zona, S., Targher, G., et al. (2016) Nonalcoholic fatty liver disease is associated with an almost twofold increased risk of incident type 2 diabetes and metabolic syndrome. Evidence from a systematic review and meta-analysis. *Journal of gastroenterology and*

hepatology, 31 (5): 936–944. doi:10.1111/JGH.13264.

Bambhroliya, Z., Sandrugu, J., Lowe, M., et al. (2022) Diabetes, Polycystic Ovarian Syndrome, Obstructive Sleep Apnea, and Obesity: A Systematic Review and Important Emerging Themes. *Cureus*, 14 (6). doi:10.7759/cureus.26325.

Baranova, a, Tran, T.P., Birerdinc, a, et al. (2011) Systematic review: association of polycystic ovary syndrome with metabolic syndrome and non-alcoholic fatty liver disease. *Alimentary pharmacology & therapeutics*, 33 (7): 801–814. doi:10.1111/j.1365-2036.2011.04579.x.

Barry, J.A., Azizia, M.M. and Hardiman, P.J. (2014) Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Human reproduction update*, 20 (5): 748–758. doi:10.1093/HUMUPD/DMU012.

Basch, E. and Schrag, D. (2019) The Evolving Uses of “Real-World” Data. *JAMA*. 321 (14). doi:10.1001/jama.2019.4064.

Bazarganipour, F., Taghavi, S.A., Montazeri, A., et al. (2015) The impact of polycystic ovary syndrome on the health-related quality of life: A systematic review and meta-analysis. *Iranian Journal of Reproductive Medicine*, 13 (2).

Behboodi Moghadam, Z., Fereidooni, B., Saffari, M., et al. (2018) Measures of health-related quality of life in pcos women: A systematic review. *International Journal of Women's Health*. 10. doi:10.2147/IJWH.S165794.

Benjafield, A. V., Ayas, N.T., Eastwood, P.R., et al. (2019) Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *The Lancet Respiratory Medicine*, 7 (8). doi:10.1016/S2213-2600(19)30198-5.

Benn, M., Voss, S.S., Holmegard, H.N., et al. (2015) Extreme concentrations of endogenous

sex hormones, ischemic heart disease, and death in women. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 35 (2). doi:10.1161/ATVBAHA.114.304821.

Benson, T. (2002) Why general practitioners use computers and hospital doctors do not—Part 1: Incentives. *BMJ*, 325 (7372). doi:10.1136/bmj.325.7372.1086.

Benson, T. (2012) The history of the Read Codes: The inaugural James Read memorial lecture 2011. *Informatics in Primary Care*. 19 (3). doi:10.14236/jhi.v19i3.811.

Berga, S. and Naftolin, F. (2012) Neuroendocrine control of ovulation. *Gynecological endocrinology: the official journal of the International Society of Gynecological Endocrinology*, 28 Suppl 1 (SUPPL.1): 9–13. doi:10.3109/09513590.2012.651929.

Bernhardt, L., Brady, E.M., Freeman, S.C., et al. (2022) Diagnostic accuracy of screening questionnaires for obstructive sleep apnoea in adults in different clinical cohorts: a systematic review and meta-analysis. *Sleep & breathing = Schlaf & Atmung*, 26 (3): 1053–1078. doi:10.1007/S11325-021-02450-9.

Berni, T.R., Morgan, C.L. and Aled Rees, D. (2021) Women with Polycystic Ovary Syndrome Have an Increased Risk of Major Cardiovascular Events: A Population Study. *Journal of Clinical Endocrinology and Metabolism*, 106 (9). doi:10.1210/clinem/dgab392.

Berni, T.R., Morgan, C.L., Berni, E.R., et al. (2018) Polycystic Ovary Syndrome Is Associated With Adverse Mental Health and Neurodevelopmental Outcomes. *The Journal of Clinical Endocrinology & Metabolism*, 103 (6): 2116–2125. doi:10.1210/jc.2017-02667.

Beroukhim, G., Esencan, E. and Seifer, D.B. (2022) Impact of sleep patterns upon female neuroendocrinology and reproductive outcomes: a comprehensive review. *Reproductive biology and endocrinology : RB&E*, 20 (1). doi:10.1186/S12958-022-00889-3.

Bertrand, F. and Maumy-Bertrand, M. (2021) Fitting and Cross-Validating Cox Models to Censored Big Data With Missing Values Using Extensions of Partial Least Squares

Regression Models. *Frontiers in Big Data*, 4. doi:10.3389/fdata.2021.684794.

Betensky, R.A. (2015) Measures of Follow-up in Time-to-Event Studies: Why provide them and what should they be? *Clinical trials (London, England)*, 12 (4): 403. doi:10.1177/1740774515586176.

Blak, B.T., Thompson, M., Dattani, H., et al. (2011) Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Informatics in primary care*, 19 (4): 251–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22828580> (Accessed: 1 December 2015).

Bland, J.M. and Altman, D.G. (1995) Multiple significance tests: the Bonferroni method. *BMJ*, 310 (6973). Available at: <http://www.bmj.com/content/310/6973/170.long> (Accessed: 24 May 2017).

Bloom, B.M., Pott, J., Thomas, S., et al. (2021) Usability of electronic health record systems in UK EDs. *Emergency Medicine Journal*, 38 (6). doi:10.1136/emmermed-2020-210401.

Bodmer, M., Becker, C., Meier, C., et al. (2011) Use of metformin and the risk of ovarian cancer: A case-control analysis. *Gynecologic Oncology*, 123 (2): 200–204. doi:10.1016/j.ygyno.2011.06.038.

Boonstra, A., Jonker, T.L., van Offenbeek, M.A.G., et al. (2021) Persisting workarounds in Electronic Health Record System use: types, risks and benefits. *BMC Medical Informatics and Decision Making*, 21 (1). doi:10.1186/s12911-021-01548-0.

Booth, N. (1994) What are the Read Codes? *Health libraries review*. 11 (3) pp. 177–82. doi:10.1046/j.1365-2532.1994.1130177.x.

Bouhanna, P. (2000) Multifactorial classification of male and female androgenetic alopecia. *Dermatologic Surgery*. 26 (6) pp. 555–561. doi:10.1046/j.1524-4725.2000.00009.x.

Boursi, B., Finkelman, B., Giantonio, B.J., et al. (2022) A clinical prediction model to assess risk for pancreatic cancer among patients with prediabetes. *European Journal of Gastroenterology and Hepatology*, 34 (1). doi:10.1097/MEG.0000000000002052.

Bozdag, G., Mumusoglu, S., Zengin, D., et al. (2016) The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction*, 31 (12): 2841–2855. doi:10.1093/humrep/dew218.

Bradley, S.H., Lawrence, N.R. and Carder, P. (2018) Using primary care data for health research in England – an overview. *Future Healthcare Journal*, 5 (3). doi:10.7861/futurehosp.5-3-207.

Brankovic, M., Kardys, I., Steyerberg, E.W., et al. (2019) Understanding of interaction (subgroup) analysis in clinical trials. *European Journal of Clinical Investigation*. 49 (8). doi:10.1111/eci.13145.

Breeze, P., Thomas, C., Thokala, P., et al. (2020) The Impact of Including Costs and Outcomes of Dementia in a Health Economic Model to Evaluate Lifestyle Interventions to Prevent Diabetes and Cardiovascular Disease. *Medical Decision Making*, 40 (7). doi:10.1177/0272989X20946758.

Brown, P.J.B., O’Neil, M. and Price, C. (1998) Semantic definition of disorders in version 3 of the read codes. *Methods of Information in Medicine*. 37 (4–5). doi:10.1055/s-0038-1634556.

Bulloch, G., Seth, I., Zhu, Z., et al. (2023) Ocular manifestations of obstructive sleep apnea: a systematic review and meta-analysis. *Graefe’s archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie*. doi:10.1007/S00417-023-06103-3.

Burke, J.F., Sussman, J.B., Kent, D.M., et al. (2015) Three simple rules to ensure reasonably credible subgroup analyses. *BMJ (Online)*, 351. doi:10.1136/bmj.h5651.

Burns, M.L., Mathis, M.R., Vandervest, J., et al. (2020) Classification of current procedural terminology codes from electronic health record data using machine learning. *Anesthesiology*. doi:10.1097/ALN.0000000000003150.

Butler, A.E., Abouseif, A., Dargham, S.R., et al. (2020) Metabolic comparison of polycystic ovarian syndrome and control women in Middle Eastern and UK Caucasian populations. *Scientific Reports*, 10 (1). doi:10.1038/s41598-020-75109-2.

Cahn, A., Shoshan, A., Sagiv, T., et al. (2020) Prediction of progression from pre-diabetes to diabetes: Development and validation of a machine learning model. *Diabetes/Metabolism Research and Reviews*, 36 (2). doi:10.1002/dmrr.3252.

Cai, B., Xu, W., Bortnichak, E., et al. (2012) An algorithm to identify medical practices common to both the General Practice Research Database and The Health Improvement Network database. *Pharmacoepidemiology and drug safety*, 21 (7): 770–774. doi:10.1002/PDS.3277.

Cai, J., Zhang, Y., Wang, Y., et al. (2019) High thyroid stimulating hormone level is associated with hyperandrogenism in euthyroid polycystic ovary syndrome (PCOS) women, independent of age, BMI, and thyroid autoimmunity: A cross-sectional analysis. *Frontiers in Endocrinology*, 10 (APR). doi:10.3389/fendo.2019.00222.

Caleyachetty, R., Thomas, G.N., Toulis, K.A., et al. (2017) Metabolically Healthy Obese and Incident Cardiovascular Disease Events Among 3.5 Million Men and Women. *Journal of the American College of Cardiology*, 70 (12): 1429–1437. doi:10.1016/j.jacc.2017.07.763.

Callahan, A., Polony, V., Posada, J.D., et al. (2021) ACE: The Advanced Cohort Engine for

searching longitudinal patient records. *Journal of the American Medical Informatics Association*, 28 (7). doi:10.1093/jamia/ocab027.

Cambras, T., Romero, O., Díez-Noguera, A., et al. (2021) Circadian patterns of patients with type 2 diabetes and obstructive sleep apnea. *Journal of Clinical Medicine*, 10 (2). doi:10.3390/jcm10020244.

Carbonari, D.M., Saine, M.E., Newcomb, C.W., et al. (2015) Use of demographic and pharmacy data to identify patients included within both the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Pharmacoepidemiology and Drug Safety*, 24 (9). doi:10.1002/pds.3844.

Carmina, E., Dreno, B., Lucky, W.A., et al. (2022) Female Adult Acne and Androgen Excess: A Report From the Multidisciplinary Androgen Excess and PCOS Committee. *Journal of the Endocrine Society*, 6 (3). doi:10.1210/jendso/bvac003.

Carr, L.H., Christ, L. and Ferro, D.F. (2023) The Electronic Health Record as a Quality Improvement Tool: Exceptional Potential with Special Considerations. *Clinics in perinatology*, 50 (2): 473–488. doi:10.1016/J.CLP.2023.01.008.

Carstens, D.S., Rodriguez, W. and Wood, M.B. (2014) Task and error analysis balancing benefits over business of electronic medical records. *International Journal of Electronic Healthcare*, 7 (4). doi:10.1504/IJEH.2014.064333.

Catalog of bias (2023a) *Chronological bias* - *Catalog of Bias*. Available at: <https://catalogofbias.org/biases/chronological-bias/> (Accessed: 25 July 2023).

Catalog of bias (2023b) *Selection bias* - *Catalog of Bias*. Available at: <https://catalogofbias.org/biases/selection-bias/> (Accessed: 24 July 2023).

Cea Soriano, L., Asimwe, A. and García Rodríguez, L.A. (2017a) Prescribing of cyproterone

acetate/ethinylestradiol in UK general practice: a retrospective descriptive study using The Health Improvement Network. *Contraception*, 95 (3). doi:10.1016/j.contraception.2016.10.003.

Cea Soriano, L., Gaist, D., Soriano-Gabarró, M., et al. (2019a) The importance of validating intracranial bleeding diagnoses in The Health Improvement Network, United Kingdom: Misclassification of onset and its impact on the risk associated with low-dose aspirin therapy. *Pharmacoepidemiology and Drug Safety*, 28 (2). doi:10.1002/pds.4561.

Cea Soriano, L., López-Garcia, E., Schulze-Rath, R., et al. (2017b) Incidence, treatment and recurrence of endometriosis in a UK-based population analysis using data from The Health Improvement Network and the Hospital Episode Statistics database. *European Journal of Contraception and Reproductive Health Care*, 22 (5). doi:10.1080/13625187.2017.1374362.

Cea Soriano, L., Zong, J. and García Rodríguez, L.A. (2019b) Feasibility and validity of The Health Improvement Network database of primary care electronic health records to identify and characterise patients with small cell lung cancer in the United Kingdom. *BMC Cancer*, 19 (1). doi:10.1186/s12885-019-5305-1.

Cea-Soriano, L., García-Rodríguez, L.A., Brodovicz, K.G., et al. (2018) Real world management of pregestational diabetes not achieving glycemic control for many patients in the UK. *Pharmacoepidemiology and Drug Safety*, 27 (8). doi:10.1002/pds.4553.

Chai, T.Y.L., Rajaratnam, R.M., Deng, D., et al. (2021) The prevalence of gestational diabetes mellitus in women diagnosed with non-alcoholic fatty liver disease during pregnancy: A systematic review and meta-analysis. *Journal of diabetes and its complications*, 35 (9). doi:10.1016/J.JDIACOMP.2021.107991.

Chan, J.L., Kar, S., Vanky, E., et al. (2017) Racial and ethnic differences in the prevalence of metabolic syndrome and its components of metabolic syndrome in women with polycystic

ovary syndrome: a regional cross-sectional study. *American Journal of Obstetrics and Gynecology*, 217 (2): 189.e1-189.e8. doi:10.1016/j.ajog.2017.04.007.

Chang, C.W., Chang, C.H., Chuang, H.Y., et al. (2022) What is the association between secondhand smoke (SHS) and possible obstructive sleep apnea: a meta-analysis. *Environmental health : a global access science source*, 21 (1). doi:10.1186/S12940-022-00868-6.

Chao, W.C., Hu, H., Ung, C.O.L., et al. (2013) Benefits and challenges of electronic health record system on stakeholders: A qualitative study of outpatient physicians. *Journal of Medical Systems*, 37 (4). doi:10.1007/s10916-013-9960-5.

Chen, H., Wang, J., Li, Z., et al. (2019) Consumption of Sugar-Sweetened Beverages Has a Dose-Dependent Effect on the Risk of Non-Alcoholic Fatty Liver Disease: An Updated Systematic Review and Dose-Response Meta-Analysis. *International journal of environmental research and public health*, 16 (12). doi:10.3390/IJERPH16122192.

Chen, H., Zheng, Z., Chen, R., et al. (2022a) A meta-analysis of the diagnostic value of NoSAS in patients with sleep apnea syndrome. *Sleep & breathing = Schlaf & Atmung*, 26 (2): 519–531. doi:10.1007/S11325-021-02410-3.

Chen, L., Zou, S. and Wang, J. (2022b) Association of Obstructive Sleep Apnea Syndrome (OSA/OSAHS) with Coronary Atherosclerosis Risk: Systematic Review and Meta-Analysis. *Computational and mathematical methods in medicine*, 2022. doi:10.1155/2022/8905736.

Chen, T., Yu, Y., Jia, F., et al. (2022c) The relationship between polycystic ovary syndrome and insulin resistance from 1983 to 2022: A bibliometric analysis. *Frontiers in public health*, 10. doi:10.3389/FPUBH.2022.960965.

Chen, Y., Bai, W., Mao, D., et al. (2023) The relationship between non-alcoholic fatty liver disease and incidence of chronic kidney disease for diabetic and non-diabetic subjects: A

meta-analysis. *Advances in clinical and experimental medicine : official organ Wroclaw Medical University*, 32 (4): 0–0. doi:10.17219/ACEM/155017.

Chen, Z. and Cai, Z. (2022) Effects of oral contraceptives plus orlistat in patients with polycystic ovary syndrome and overweight/obesity: A meta-analysis. *The journal of obstetrics and gynaecology research*, 48 (6): 1399–1408. doi:10.1111/JOG.15211.

Cheong, A.J.Y., Ho, O.T.W., Wang, S.K.X., et al. (2023) Association between obstructive sleep apnea and floppy eyelid syndrome: A systematic review and metaanalysis. *Survey of ophthalmology*, 68 (2): 257–264. doi:10.1016/J.SURVOPHTHAL.2022.11.006.

Cheong, A.J.Y., Tan, B.K.J., Teo, Y.H., et al. (2022) Obstructive Sleep Apnea and Lung Cancer: A Systematic Review and Meta-Analysis. *Annals of the American Thoracic Society*. 19 (3). doi:10.1513/AnnalsATS.202108-960OC.

Cherskov, A., Pohl, A., Allison, C., et al. (2018) Polycystic ovary syndrome and autism: A test of the prenatal sex steroid theory. *Translational Psychiatry*, 8 (1). doi:10.1038/s41398-018-0186-7.

Chiaffarino, F., Cipriani, S., Dalmartello, M., et al. (2022) Prevalence of polycystic ovary syndrome in European countries and USA: A systematic review and meta-analysis. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 279. doi:10.1016/j.ejogrb.2022.10.020.

Chisholm, J. (1990) The Read clinical classification. *British Medical Journal*. 300 (6732). doi:10.1136/bmj.300.6732.1092.

Chiu, H.Y., Chen, P.Y., Chuang, L.P., et al. (2017) Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: A bivariate meta-analysis. *Sleep Medicine Reviews*. 36 pp. 57–70. doi:10.1016/j.smr.2016.10.004.

Cho, J., Lee, I., Park, D.H., et al. (2021) Relationships between socioeconomic status, handgrip strength, and non-alcoholic fatty liver disease in middle-aged adults. *International Journal of Environmental Research and Public Health*, 18 (4). doi:10.3390/ijerph18041892.

Ciardullo, S., Grassi, G., Mancina, G., et al. (2022) Nonalcoholic fatty liver disease and risk of incident hypertension: a systematic review and meta-analysis. *European journal of gastroenterology & hepatology*, 34 (4): 365–371. doi:10.1097/MEG.0000000000002299.

Clark, R. V., Wald, J.A., Swerdloff, R.S., et al. (2019) Large divergence in testosterone concentrations between men and women: Frame of reference for elite athletes in sex-specific competition in sports, a narrative review. *Clinical Endocrinology*. 90 (1). doi:10.1111/cen.13840.

Cleves, M.A., Gould, W. and Marchenko, Y. V. (2016) *An Introduction to Survival Analysis Using Stata*. STATA.

Colonetti, L., Grande, A.J., Toreti, I.R., et al. (2022) Green tea promotes weight loss in women with polycystic ovary syndrome: Systematic review and meta-analysis. *Nutrition Research*. 104. doi:10.1016/j.nutres.2022.03.009.

Conway, G., Dewailly, D., Diamanti-Kandarakis, E., et al. (2014) The polycystic ovary syndrome: a position statement from the European Society of Endocrinology. *European journal of endocrinology / European Federation of Endocrine Societies*, 171 (4): P1-29. doi:10.1530/EJE-14-0253.

Cooney, L.G., Lee, I., Sammel, M.D., et al. (2017) High prevalence of moderate and severe depressive and anxiety symptoms in polycystic ovary syndrome: a systematic review and meta-analysis. *Human reproduction (Oxford, England)*, 32 (5): 1075–1091. doi:10.1093/HUMREP/DEX044.

CPRD (2022) *CPRD GOLD Data Specification*.

Cronin, F.M., Hurley, S.M., Buckley, T., et al. (2022) Mediators of socioeconomic differences in overweight and obesity among youth in Ireland and the UK (2011–2021): a systematic review. *BMC Public Health*, 22 (1). doi:10.1186/s12889-022-14004-z.

Cusick, M.M., Sholle, E.T., Davila, M.A., et al. (2020) A Method to Improve Availability and Quality of Patient Race Data in an Electronic Health Record System. *Applied Clinical Informatics*, 11 (5). doi:10.1055/s-0040-1718756.

Cussen, L., McDonnell, T., Bennett, G., et al. (2022) Approach to androgen excess in women: Clinical and biochemical insights. *Clinical Endocrinology*. 97 (2). doi:10.1111/cen.14710.

Dafoulas, G.E., Toulis, K.A., Mccorry, D., et al. (2017) Type 1 diabetes mellitus and risk of incident epilepsy: a population-based, open-cohort study. *Diabetologia*, 60 (2): 258–261. doi:10.1007/s00125-016-4142-x.

Danaei, G., García Rodríguez, L.A., Cantero, O.F., et al. (2018) Electronic medical records can be used to emulate target trials of sustained treatment strategies. *Journal of Clinical Epidemiology*, 96. doi:10.1016/j.jclinepi.2017.11.021.

Davé, S. and Petersen, I. (2009) Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiology and Drug Safety*, 18 (8): 704–707. doi:10.1002/pds.1770.

Delgado-Rodríguez, M. and Llorca, J. (2004) Bias. *Journal of epidemiology and community health*, 58 (8): 635–641. doi:10.1136/jech.2003.008466.

Denburg, M.R., Haynes, K., Shults, J., et al. (2011) Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease.

Pharmacoepidemiology and drug safety, 20 (11): 1138–49. doi:10.1002/pds.2203.

Deo, S.V., Deo, V. and Sundaram, V. (2021) Survival analysis—part 2: Cox proportional hazards model. *Indian Journal of Thoracic and Cardiovascular Surgery*, 37 (2). doi:10.1007/s12055-020-01108-7.

Derks, I.P.M., Koster, A., Schram, M.T., et al. (2017) The association of early life socioeconomic conditions with prediabetes and type 2 diabetes: results from the Maastricht study. *International Journal for Equity in Health*, 16 (1). doi:10.1186/s12939-017-0553-7.

Deshmukh, H., Shah, N., Papageorgiou, M., et al. (2022) Genetic risk for the polycystic ovary syndrome, bone mineral density and fractures in women and men: A UK Biobank Mendelian randomisation study. *Bone*, 155. doi:10.1016/j.bone.2021.116285.

Deswal, R., Yadav, A. and Dang, A.S. (2018) Sex hormone binding globulin - an important biomarker for predicting PCOS risk: A systematic review and meta-analysis. *Systems Biology in Reproductive Medicine*, 64 (1): 12–24. doi:10.1080/19396368.2017.1410591.

Dijkman, B., Kooistra, B., Bhandari, M., et al. (2009) How to work with a subgroup analysis. *Canadian Journal of Surgery*, 52 (6).

Ding, T., Baio, G., Hardiman, P.J., et al. (2016) Diagnosis and management of polycystic ovary syndrome in the UK: a retrospective cohort study. *BMJ Open*, 6: 5–8. doi:10.1136/bmjopen-2016-012461.

Ding, T., Hardiman, P.J., Petersen, I., et al. (2018a) Incidence and prevalence of diabetes and cost of illness analysis of polycystic ovary syndrome: a Bayesian modelling study. *Human Reproduction*, 33 (7): 1299–1306. doi:10.1093/humrep/dey093.

Ding, X., Yang, L., Wang, J., et al. (2018b) Subclinical Hypothyroidism in Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis. *Frontiers in Endocrinology*. 9.

doi:10.3389/fendo.2018.00700.

Dong, J., & Rees, D. A. (2023). Polycystic ovary syndrome: pathophysiology and therapeutic opportunities. *BMJ Medicine*, 2(1), e000548. <https://doi.org/10.1136/bmjmed-2023-000548>

Duchesneau, E.D., Jackson, B.E., Webster-Clark, M., et al. (2022) The Timing, the Treatment, the Question: Comparison of Epidemiologic Approaches to Minimize Immortal Time Bias in Real-World Data Using a Surgical Oncology Example. *Cancer Epidemiology Biomarkers and Prevention*, 31 (11). doi:10.1158/1055-9965.EPI-22-0495.

Dunn, A.J. (2021) Environmental Influences and Polycystic Ovarian Syndrome. *Clinical obstetrics and gynecology*, 64 (1): 33–38. doi:10.1097/GRF.0000000000000562.

Dupont, W.D. (2021) Review of Michael N. Mitchell's Data Management Using Stata: A Practical Handbook, Second Edition. *Stata Journal*, 21 (3). doi:10.1177/1536867X211045581.

Eberhardt, J., Bright, C.J., Gildea, C., et al. (2022) Impact of the third national 'Be Clear on Cancer' Breast Cancer in Women over 70 Campaign on general practitioner attendance and referral, diagnosis rates and prevalence awareness. *European Journal of Cancer Care*, 31 (3). doi:10.1111/ecc.13583.

Echouffo-Tcheugui, J.B., Perreault, L., Ji, L., et al. (2023) Diagnosis and Management of Prediabetes: A Review. *JAMA*, 329 (14): 1206–1216. doi:10.1001/jama.2023.4063.

Einarson, T.R., Acs, A., Ludwig, C., et al. (2018) Economic Burden of Cardiovascular Disease in Type 2 Diabetes: A Systematic Review. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 21 (7): 881–890. doi:10.1016/J.JVAL.2017.12.019.

Elhassan, Y.S., Idkowiak, J., Smith, K., et al. (2018) Causes, patterns, and severity of

androgen excess in 1205 consecutively recruited women. *Journal of Clinical Endocrinology and Metabolism*, 103 (3): 1214–1223. doi:10.1210/jc.2017-02426.

Ellis, N.T. and Kidd, M.R. (2000) General practice computerisation: Lessons from the United Kingdom. *Medical Journal of Australia*. 172 (1). doi:10.5694/j.1326-5377.2000.tb123873.x.

Von Elm, E., Altman, D.G., Egger, M., et al. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *PLOS Medicine*, 4 (10): e296. doi:10.1371/JOURNAL.PMED.0040296.

von Elm, E., Altman, D.G., Egger, M., et al. (2007) The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*, 370 (9596): 1453–1457. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25046131>.

Emanuel, R.H.K., Roberts, J., Docherty, P.D., et al. (2022) A review of the hormones involved in the endocrine dysfunctions of polycystic ovary syndrome and their interactions. *Frontiers in Endocrinology*, 13. doi:10.3389/fendo.2022.1017468.

Etindele Sossio, F. and Matos, E. (2021) Socioeconomic disparities in obstructive sleep apnea: a systematic review of empirical research. *Sleep and Breathing*. 25 (4). doi:10.1007/s11325-020-02274-z.

Fadlalmola, H.A., Elhusein, A.M., Al-Sayaghi, K.M., et al. (2023) Efficacy of resveratrol in women with polycystic ovary syndrome: a systematic review and meta-analysis of randomized clinical trials. *Pan African Medical Journal*, 44. doi:10.11604/pamj.2023.44.134.32404.

Farrokhyar, F., Skorzewski, P., Phillips, M.R., et al. (2022) When to believe a subgroup analysis: revisiting the 11 criteria. *Eye (Basingstoke)*. 36 (11). doi:10.1038/s41433-022-

01948-0.

Feltner, C., Wallace, I.F., Aymes, S., et al. (2022) Screening for Obstructive Sleep Apnea in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 328 (19) pp. 1951–1971. doi:10.1001/jama.2022.18357.

FitzSimmons, J. and Davidson, S. (1990) Ascertainment bias in diabetic outcome data. *American Journal of Obstetrics and Gynecology*. 163 (2). doi:10.1016/0002-9378(90)91235-5.

Fletcher, J. (2007) Subgroup analyses: How to avoid being misled. *British Medical Journal*, 335 (7610). doi:10.1136/bmj.39265.596262.AD.

Flory, J.H., Roy, J., Gagne, J.J., et al. (2017) Missing laboratory results data in electronic health databases: Implications for monitoring diabetes risk. *Journal of Comparative Effectiveness Research*, 6 (1). doi:10.2217/cer-2016-0033.

Forslund, M., Landin-Wilhelmsen, K., Brännström, M., et al. (2023) No difference in morbidity between perimenopausal women with PCOS with and without previous wedge resection. *European journal of obstetrics, gynecology, and reproductive biology*, 285: 74–78. doi:10.1016/J.EJOGRB.2023.04.008.

Forslund, M., Schmidt, J., Brännström, M., et al. (2022) Morbidity and mortality in PCOS: A prospective follow-up up to a mean age above 80 years. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 271. doi:10.1016/j.ejogrb.2022.02.020.

Fox, R. and Hull, M. (1993) Ultrasound Diagnosis of Polycystic Ovaries. *Annals of the New York Academy of Sciences*, 687 (1): 217–223. doi:10.1111/j.1749-6632.1993.tb43868.x.

Freitas De-Medeiros, S., Yamamoto, M.M.W., Souto De-Medeiros, M.A., et al. (2017) Should subclinical hypothyroidism be an exclusion criterion for the diagnosis of polycystic

ovary syndrome? *Journal of Reproduction and Infertility*, 18 (2).

Fu, C.E., Ng, C.H., Yong, J.N., et al. (2023a) A Meta-analysis on Associated Risk of Mortality in Nonalcoholic Fatty Liver Disease. *Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists*, 29 (1): 33–39. doi:10.1016/J.EPRAC.2022.10.007.

Fu, L., Xie, N., Qu, F., et al. (2023b) The Association Between Polycystic Ovary Syndrome and Metabolic Syndrome in Adolescents: a Systematic Review and Meta-analysis. *Reproductive Sciences*. 30 (1). doi:10.1007/s43032-022-00864-8.

Gallagher, A.M., Dedman, D., Padmanabhan, S., et al. (2019) The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiology and Drug Safety*, 28 (5). doi:10.1002/pds.4747.

Gao, Y., Liu, H., Qiao, L., et al. (2023) Study of Burden in Polycystic Ovary Syndrome at Global, Regional, and National Levels from 1990 to 2019. *Healthcare (Switzerland)*, 11 (4). doi:10.3390/healthcare11040562.

Garcia-Beltran, C., Bassols, J., Carreras-Badosa, G., et al. (2023) Raised Thyroid Stimulating Hormone in Girls with Polycystic Ovary Syndrome: Effects of Randomized Interventions. *Hormone Research in Paediatrics*. doi:10.1159/000529183.

Gardner, M.P., Wang, J., Hazlehurst, J.M., et al. (2023) Risk of progression from pre-diabetes to type 2 diabetes in a large UK adult cohort. *Diabetic Medicine*, 40 (3). doi:10.1111/dme.14996.

Ge, J.J., Wang, D.J., Song, W., et al. (2022) The effectiveness and safety of liraglutide in treating overweight/obese patients with polycystic ovary syndrome: a meta-analysis. *Journal*

of endocrinological investigation, 45 (2): 261–273. doi:10.1007/S40618-021-01666-6.

George, A., Stead, T.S. and Ganti, L. (2020) What's the Risk: Differentiating Risk Ratios, Odds Ratios, and Hazard Ratios? *Cureus*. doi:10.7759/cureus.10047.

Glintborg, D., Mumm, H., Hougaard, D., et al. (2010) Ethnic differences in Rotterdam criteria and metabolic risk factors in a multiethnic group of women with PCOS studied in Denmark. *Clinical endocrinology*, 73 (6): 732–8. doi:10.1111/j.1365-2265.2010.03873.x.

Glintborg, D., Rubin, K.H., Nybo, M., et al. (2019) Increased risk of thyroid disease in danish women with polycystic ovary syndrome: A cohort study. *Endocrine Connections*, 8 (10). doi:10.1530/EC-19-0377.

Gokhale, K.M., Chandan, J.S., Toulis, K., et al. (2021) Data extraction for epidemiological research (DExtER): a novel tool for automated clinical epidemiology studies. *European Journal of Epidemiology*, 36 (2): 165–178. doi:10.1007/S10654-020-00677-6/TABLES/6.

Goldstein, S.R. (1990) Incorporating endovaginal ultrasonography into the overall gynecologic examination. *American journal of obstetrics and gynecology*, 162 (3): 625–632. doi:10.1016/0002-9378(90)90971-9.

Greengross, P., Grant, K. and Collini, E. (1999) *The history and development of the UK National Health Service 1948 - 1999*. Available at: www.healthsystemsrc.org.

Guay-Gagnon, M., Vat, S., Forget, M.F., et al. (2022) Sleep apnea and the risk of dementia: A systematic review and meta-analysis. *Journal of sleep research*, 31 (5). doi:10.1111/JSR.13589.

Guest, J.F., Panca, M., Sladkevicius, E., et al. (2014) Clinical outcomes and cost-effectiveness of continuous positive airway pressure to manage obstructive sleep apnea in patients with type 2 diabetes in the U.K. *Diabetes Care*, 37 (5). doi:10.2337/dc13-2539.

- Gulliford, M.C., Sun, X., Anjuman, T., et al. (2020) Comparison of antibiotic prescribing records in two UK primary care electronic health record systems: Cohort study using CPRD GOLD and CPRD Aurum databases. *BMJ Open*, 10 (6). doi:10.1136/bmjopen-2020-038767.
- Guo, L., Gordon, N.P., Chandra, M., et al. (2021) The risks of polycystic ovary syndrome and diabetes vary by ethnic subgroup among young asian women. *Diabetes Care*. 44 (6). doi:10.2337/dc21-0373.
- Gutiérrez-Tobal, G.C., Álvarez, D., Kheirandish-Gozal, L., et al. (2022) Reliability of machine learning to diagnose pediatric obstructive sleep apnea: Systematic review and meta-analysis. *Pediatric Pulmonology*, 57 (8): 1931–1943. doi:10.1002/ppul.25423.
- Haas, C.L., Varbo, A., Laursen, P.N., et al. (2023) Association between body mass index, weight loss and the chance of pregnancy in women with polycystic ovary syndrome and overweight or obesity: a retrospective cohort study in the UK. *Human reproduction (Oxford, England)*, 38 (3). doi:10.1093/humrep/deac267.
- Hacivelioglu, S., Gungor, A.N.C., Gencer, M., et al. (2013) Acne severity and the Global Acne Grading System in polycystic ovary syndrome. *International Journal of Gynecology & Obstetrics*, 123 (1): 33–36. doi:10.1016/j.ijgo.2013.05.005.
- Hadjiconstantinou, M., Mani, H., Patel, N., et al. (2017) Understanding and supporting women with polycystic ovary syndrome: A qualitative study in an ethnically diverse UK sample. *Endocrine Connections*, 6 (5). doi:10.1530/EC-17-0053.
- Hanson, M.S. (2005) Stata tip 26: Maximizing compatibility between macintosh and windows. *Stata Journal*. 5 (4). doi:10.1177/1536867x0500500411.
- Heald, A.H., Livingston, M., Holland, D., et al. (2018) Polycystic ovarian syndrome: Assessment of approaches to diagnosis and cardiometabolic monitoring in UK primary care.

International Journal of Clinical Practice, 72 (1). doi:10.1111/ijcp.13046.

Healy, M.L., Gibney, J., Pentecost, C., et al. (2014) Endocrine profiles in 693 elite athletes in the postcompetition setting. *Clinical Endocrinology*, 81 (2). doi:10.1111/cen.12445.

Helvaci, N., Karabulut, E., Demir, A.U., et al. (2017) Polycystic ovary syndrome and the risk of obstructive sleep apnea: A meta-analysis and review of the literature. *Endocrine Connections*, 6 (7): 437–445. doi:10.1530/EC-17-0129.

Herrett, E., Gallagher, A.M., Bhaskaran, K., et al. (2015) Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International journal of epidemiology*, 44 (3): 827–836. doi:10.1093/IJE/DYV098.

Hippisley-Cox, J. and Coupland, C. (2017) Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. *BMJ (Clinical research ed.)*, 359. doi:10.1136/bmj.j5019.

Hippisley-Cox, J., Stables, D. and Pringle, M. (2004) QRESEARCH: A new general practice database for research. *Informatics in Primary Care*. 12 (1). doi:10.14236/jhi.v12i1.108.

Hobbs, F.D.R., Bankhead, C., Mukhtar, T., et al. (2016) Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *The Lancet*, 387 (10035). doi:10.1016/S0140-6736(16)00620-6.

Horsfall, L., Walters, K. and Petersen, I. (2013) Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiology and drug safety*, 22 (1): 64–9. doi:10.1002/pds.3368.

Hou, H., Zhao, Y., Yu, W., et al. (2018) Association of obstructive sleep apnea with hypertension: A systematic review and meta-analysis. *Journal of Global Health*, 8 (1). doi:10.7189/jogh.08.010405.

House of Parliament (2016) *Government Proposals for Electronic Health Records*. Available at: www.parliament.uk/post.

Humbert, M., Sitbon, O., Yaïci, A., et al. (2010) Survival in incident and prevalent cohorts of patients with pulmonary arterial hypertension. *European Respiratory Journal*, 36 (3). doi:10.1183/09031936.00057010.

Ichikawa, M., Akiyama, T., Tsujimoto, Y., et al. (2022) Diagnostic accuracy of home sleep apnea testing using peripheral arterial tonometry for sleep apnea: A systematic review and meta-analysis. *Journal of sleep research*, 31 (6). doi:10.1111/JSR.13682.

IMS Health (2015) *THIN Data Guide for Researchers*.

Ingram, J.R., Jenkins-Jones, S., Knipe, D.W., et al. (2018) Population-based Clinical Practice Research Datalink study using algorithm modelling to identify the true burden of hidradenitis suppurativa. *British Journal of Dermatology*, 178 (4). doi:10.1111/bjd.16101.

Isakari, M., Sanchez, A., Conic, R., et al. (2023) Benefits and Challenges of Transitioning Occupational Health to an Enterprise Electronic Health Record. *Journal of Occupational & Environmental Medicine*, 65 (7): 615–620. doi:10.1097/JOM.0000000000002864.

Ismail, S.U., Asamane, E.A., Osei-Kwasi, H.A., et al. (2022) Socioeconomic Determinants of Cardiovascular Diseases, Obesity, and Diabetes among Migrants in the United Kingdom: A Systematic Review. *International journal of environmental research and public health*, 19 (5). doi:10.3390/IJERPH19053070.

Jacobs, M., Boersma, L.J., Swart, R., et al. (2019) Electronic Health Record implementation in a large academic radiotherapy department: Temporarily disruptions but long-term benefits. *International Journal of Medical Informatics*, 129. doi:10.1016/j.ijmedinf.2019.07.008.

Jarada, T.N., O’Sullivan, D.E., Brenner, D.R., et al. (2023) Selection Bias in Real-World

Data Studies Used to Support Health Technology Assessments: A Case Study in Metastatic Cancer. *Current Oncology*, 30 (2). doi:10.3390/curroncol30020151.

Joham, A.E., Piltonen, T., Lujan, M.E., et al. (2022) Challenges in diagnosis and understanding of natural history of polycystic ovary syndrome. *Clinical Endocrinology*. 97 (2). doi:10.1111/cen.14757.

Jones, M. (2022) Benefits of using SNOMED CT in the UK National Health Service (NHS). *Roadmap to Successful Digital Health Ecosystems: A Global Perspective*, pp. 489–497. doi:10.1016/B978-0-12-823413-6.00023-9.

Kaczmarek, C., Haller, D.M. and Yaron, M. (2016) Health-Related Quality of Life in Adolescents and Young Adults with Polycystic Ovary Syndrome: A Systematic Review. *Journal of Pediatric and Adolescent Gynecology*. 29 (6). doi:10.1016/j.jpog.2016.05.006.

Kahal, H., Kyrou, I., Uthman, O.A., et al. (2020) The prevalence of obstructive sleep apnoea in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Sleep and Breathing*, 24 (1): 339–350. doi:10.1007/s11325-019-01835-1.

Kazemi, M., Kim, J.Y., Parry, S.A., et al. (2021) Disparities in cardio metabolic risk between Black and White women with polycystic ovary syndrome: a systematic review and meta-analysis. *American Journal of Obstetrics and Gynecology*. 224 (5). doi:10.1016/j.ajog.2020.12.019.

van Keizerswaard, J., Dietz de Loos, A.L.P., Louwers, Y. V., et al. (2022) Changes in individual polycystic ovary syndrome phenotypical characteristics over time: a long-term follow-up study. *Fertility and sterility*, 117 (5): 1059–1066. doi:10.1016/J.FERTNSTERT.2022.01.014.

Kelly dos Santos, I., Ashe, M.C., Cobucci, R.N., et al. (2020) The effect of exercise as an

intervention for women with polycystic ovary syndrome: A systematic review and meta-analysis. *Medicine*, 99 (16): E19644. doi:10.1097/MD.00000000000019644.

Kempegowda, P., Melson, E., Manolopoulos, K.N., et al. (2020) Implicating androgen excess in propagating metabolic disease in polycystic ovary syndrome. *Therapeutic advances in endocrinology and metabolism*, 11. doi:10.1177/2042018820934319.

Keramati, M., Musazadeh, V., Malekhamdi, M., et al. (2022) Cinnamon, an effective anti-obesity agent: Evidence from an umbrella meta-analysis. *Journal of Food Biochemistry*. 46 (8). doi:10.1111/jfbc.14166.

Khadilkar, S. V. and Khadilkar, S.S. (2020) Bias in Clinical Practice. *Journal of Obstetrics and Gynecology of India*. 70 (1). doi:10.1007/s13224-019-01304-5.

Khalil, C., Zarabi, S., Kirkham, K., et al. (2023) Validity of non-contact methods for diagnosis of Obstructive Sleep Apnea: a systematic review and meta-analysis. *Journal of clinical anesthesia*, 87. doi:10.1016/J.JCLINANE.2023.111087.

Khan, M.J., Ullah, A. and Basit, S. (2019) Genetic Basis of Polycystic Ovary Syndrome (PCOS): Current Perspectives. *The application of clinical genetics*, 12: 249–260. doi:10.2147/TACG.S200341.

Khan, N.F., Harrison, S.E. and Rose, P.W. (2010) Validity of diagnostic coding within the General Practice Research Database: A systematic review. *British Journal of General Practice*. 60 (572). doi:10.3399/bjgp10X483562.

Kim, A.S. and Wild, R.A. (2022) The natural history of polycystic ovary syndrome: commentary on a longitudinal study evaluating changes in phenotype with age. *Fertility and Sterility*, 117 (5): 1067–1068. doi:10.1016/J.FERTNSTERT.2022.03.003.

Kim, D.H., Kim, S.W. and Hwang, S.H. (2022) Diagnostic value of smartphone in

obstructive sleep apnea syndrome: A systematic review and meta-analysis. *PloS one*, 17 (5). doi:10.1371/JOURNAL.PONE.0268585.

Kim, H.J., Adams, J.M., Gudmundsson, J.A., et al. (2017) Polycystic ovary morphology: age-based ultrasound criteria. *Fertility and Sterility*, 108 (3): 548–553. doi:10.1016/j.fertnstert.2017.07.005.

Kim, J., Kaufman, J.S. and Bang, H. (2018) Graphing Ratio Measures on Forest Plot. *Journal of the American College of Cardiology*. 71 (5). doi:10.1016/j.jacc.2017.10.098.

King, J., Patel, V., Jamoom, E.W., et al. (2014) Clinical benefits of electronic health record use: National findings. *Health Services Research*, 49 (1 PART 2). doi:10.1111/1475-6773.12135.

Koletsis, D. and Pandis, N. (2017) Survival analysis, part 3: Cox regression. *American Journal of Orthodontics and Dentofacial Orthopedics*. 152 (5). doi:10.1016/j.ajodo.2017.07.009.

Kontopantelis, E., Stevens, R.J., Helms, P.J., et al. (2018) Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: A cross-sectional population study. *BMJ Open*, 8 (2). doi:10.1136/bmjopen-2017-020738.

Kumarendran, B., Sumilo, D., O'Reilly, M.W., et al. (2019) Increased risk of obstructive sleep apnoea in women with polycystic ovary syndrome: A population-based cohort study. *European Journal of Endocrinology*, 180 (4): 265–272. doi:10.1530/EJE-18-0693.

Lambert, P.C. (2017) The estimation and modeling of cause-specific cumulative incidence functions using time-dependent weights. *Stata Journal*, 17 (1). doi:10.1177/1536867x1701700110.

Langan, S.M., Schmidt, S.A., Wing, K., et al. (2018) The reporting of studies conducted

using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *The BMJ*. 363. doi:10.1136/bmj.k3532.

Lavrentaki, A., Thomas, T., Subramanian, A., et al. (2019) Increased risk of non-alcoholic fatty liver disease in women with gestational diabetes mellitus: A population-based cohort study, systematic review and meta-analysis. *Journal of diabetes and its complications*, 33 (10). doi:10.1016/J.JDIACOMP.2019.06.006.

Lawal, Y., Bello, F., & Kaoje, Y. S. (2020). Prediabetes deserves more attention: A review. In *Clinical Diabetes* (Vol. 38, Issue 4, pp. 328–338). American Diabetes Association Inc. <https://doi.org/10.2337/cd19-0101>

Le, M.H., Le, D.M., Baez, T.C., et al. (2023) Global incidence of non-alcoholic fatty liver disease: a systematic review and meta-analysis of 63 studies and 1,201,807 persons. *Journal of hepatology*. doi:10.1016/J.JHEP.2023.03.040.

Lee, B., Jung, E.A., Yoo, J.J., et al. (2020) Prevalence, incidence and risk factors of tamoxifen-related non-alcoholic fatty liver disease: A systematic review and meta-analysis. *Liver international: official journal of the International Association for the Study of the Liver*, 40 (6): 1344–1355. doi:10.1111/LIV.14434.

Lee, W.T.N., Ling, Y., Sheares, K.K., et al. (2012) Predicting survival in pulmonary arterial hypertension in the UK. *European Respiratory Journal*, 40 (3). doi:10.1183/09031936.00196611.

Legro, R.S. (2003) Diagnostic Criteria in Polycystic Ovary Syndrome. *Seminars in Reproductive Medicine*, 21 (3): 267–275. doi:10.1055/s-2003-43304.

Legro, R.S., Arslanian, S.A., Ehrmann, D.A., et al. (2013) Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *The Journal of clinical endocrinology and metabolism*, 98 (12): 4565–92. doi:10.1210/jc.2013-2350.

Legro, R.S., Schlaff, W.D., Diamond, M.P., et al. (2010) Total testosterone assays in women with polycystic ovary syndrome: Precision and correlation with hirsutism. *Journal of Clinical Endocrinology and Metabolism*, 95 (12): 5305–5313. doi:10.1210/jc.2010-1123.

Lévesque, L.E., Hanley, J.A., Kezouh, A., et al. (2010) Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes. *BMJ (Online)*, 340 (7752): 907–911. doi:10.1136/bmj.b5087.

Lewis, J.D., Schinnar, R., Bilker, W.B., et al. (2007) Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiology and drug safety*, 16 (4): 393–401. doi:10.1002/pds.1335.

Li, W., Wen, W., Xie, D., et al. (2022) Association between non-alcoholic fatty liver disease and risk of incident heart failure: a meta-analysis of observational studies. *Therapeutic advances in chronic disease*, 13. doi:10.1177/20406223221119626.

Li, Y., Li, Y., Yu Ng, E.H., et al. (2011) Polycystic ovary syndrome is associated with negatively variable impacts on domains of health-related quality of life: Evidence from a meta-analysis. *Fertility and Sterility*, 96 (2). doi:10.1016/j.fertnstert.2011.05.072.

Liao, D., Zhong, C., Li, C., et al. (2018) Meta-analysis of the effects of probiotic supplementation on glycemia, lipidic profiles, weight loss and C-reactive protein in women with polycystic ovarian syndrome. *Minerva Medica*. 109 (6). doi:10.23736/S0026-4806.18.05728-2.

Lim, S.S., Kakoly, N.S., Tan, J.W.J., et al. (2019) Metabolic syndrome in polycystic ovary syndrome: a systematic review, meta-analysis and meta-regression. *Obesity Reviews*. 20 (2). doi:10.1111/obr.12762.

Lim, S.S., Norman, R.J., Davies, M.J., et al. (2013) The effect of obesity on polycystic ovary syndrome: a systematic review and meta-analysis. *Obesity reviews : an official journal of the*

International Association for the Study of Obesity, 14 (2): 95–109. doi:10.1111/j.1467-789X.2012.01053.x.

Lis, Y. and Mann, R.D. (1995) The VAMP research multi-purpose database in the U.K. *Journal of Clinical Epidemiology*, 48 (3). doi:10.1016/0895-4356(94)00137-F.

Liu, J., Wu, Q., Hao, Y., et al. (2021) Measuring the global disease burden of polycystic ovary syndrome in 194 countries: Global Burden of Disease Study 2017. *Human Reproduction*, 36 (4): 1108–1119. doi:10.1093/humrep/deaa371.

Livadas, S., Macut, D., Bothou, C., et al. (2020) Insulin resistance, androgens, and lipids are gradually improved in an age-dependent manner in lean women with polycystic ovary syndrome: insights from a large Caucasian cohort. *Hormones*, 19 (4). doi:10.1007/s42000-020-00211-z.

Lizneva, D., Gavrilova-Jordan, L., Walker, W., et al. (2016a) Androgen excess: Investigations and management. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 37: 98–118. doi:10.1016/j.bpobgyn.2016.05.003.

Lizneva, D., Suturina, L., Walker, W., et al. (2016b) Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertility and Sterility*, 106 (1): 6–15. doi:10.1016/j.fertnstert.2016.05.003.

Lo, J.C., Feigenbaum, S.L., Yang, J., et al. (2006) Epidemiology and Adverse Cardiovascular Risk Profile of Diagnosed Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology & Metabolism*, 91 (4): 1357–1363. doi:10.1210/jc.2005-2430.

Di Lorenzo, B., Pau, M.C., Zinellu, E., et al. (2023) Association between Red Blood Cell Distribution Width and Obstructive Sleep Apnea Syndrome: A Systematic Review and Meta-Analysis. *Journal of clinical medicine*, 12 (9). doi:10.3390/JCM12093302.

Lourie, E.M., Utidjian, L.H., Ricci, M.F., et al. (2021) Reducing electronic health record-related burnout in providers through a personalized efficiency improvement program. *Journal of the American Medical Informatics Association*, 28 (5). doi:10.1093/jamia/ocaa248.

Louwens, Y. V., Roest-Schalken, M.E., Kleefstra, N., et al. (2014) Excess mortality in mothers of patients with polycystic ovary syndrome. *Human Reproduction*, 29 (8). doi:10.1093/humrep/deu107.

Ludwig, E. (1977) Classification of the types of androgenetic alopecia (common baldness) occurring in the female sex. *British Journal of Dermatology*, 97 (3): 247–254. doi:10.1111/j.1365-2133.1977.tb15179.x.

Lujan, M.E., Chizen, D.R. and Pierson, R.A. (2008) Diagnostic Criteria for Polycystic Ovary Syndrome: Pitfalls and Controversies. *Journal of Obstetrics and Gynaecology Canada*, 30 (8). doi:10.1016/S1701-2163(16)32915-2.

Lunde, O., Magnus, P., Sandvik, L., et al. (1989) Familial clustering in the polycystic ovarian syndrome. *Gynecologic and obstetric investigation*, 28 (1): 23–30. doi:10.1159/000293493.

Luo, P., Su, Z., Li, P., et al. (2023) Effects of Sleeve Gastrectomy on Patients with Obesity and Polycystic Ovary Syndrome: a Meta-analysis. *Obesity surgery*. doi:10.1007/S11695-023-06617-Z.

Luque-Ramírez, M., Alpañés, M., Sanchón, R., et al. (2015) Referral bias in female functional hyperandrogenism and polycystic ovary syndrome. *European Journal of Endocrinology*, 173 (5). doi:10.1530/EJE-15-0646.

de Lusignan, S., Liaw, S.-T., Dedman, D., et al. (2015) An algorithm to improve diagnostic accuracy in diabetes in computerised problem orientated medical records (POMR) compared with an established algorithm developed in episode orientated records (EOMR). *Journal of*

innovation in health informatics, 22 (2): 255–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26245239> (Accessed: 11 July 2016).

Luzzi, V., Mazur, M., Guaragna, M., et al. (2022) Correlations of Obstructive Sleep Apnea Syndrome and Daytime Sleepiness with the Risk of Car Accidents in Adult Working Population: A Systematic Review and Meta-Analysis with a Gender-Based Approach. *Journal of clinical medicine*, 11 (14). doi:10.3390/JCM11143971.

Lyu, X., Lyu, T., Wang, X., et al. (2021) The Antiobesity Effect of GLP-1 Receptor Agonists Alone or in Combination with Metformin in Overweight /Obese Women with Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis. *International journal of endocrinology*, 2021. doi:10.1155/2021/6616693.

Madjar, K., Zucknick, M., Ickstadt, K., et al. (2021) Combining heterogeneous subgroups with graph-structured variable selection priors for Cox regression. *BMC Bioinformatics*, 22 (1). doi:10.1186/s12859-021-04483-z.

Maguire, A., Blak, B.T. and Thompson, M. (2009) The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiology and drug safety*, 18 (1): 76–83. doi:10.1002/pds.1688.

Mamede, S., Van Gog, T., Van Den Berge, K., et al. (2010) Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA*, 304 (11). doi:10.1001/jama.2010.1276.

Mantovani, A., Petracca, G., Beatrice, G., et al. (2021) Non-alcoholic fatty liver disease and risk of incident diabetes mellitus: an updated meta-analysis of 501 022 adult individuals. *Gut*, 70 (5): 962–969. doi:10.1136/GUTJNL-2020-322572.

Maria-Sophie Barthel, F. and Royston, P. (2006) *Graphical representation of interactions*. Marquina, C., Makarounas-Kirchmann, K., Holden, K., et al. (2023) The preventable

productivity burden of sleep apnea in Australia: a lifetime modelling study. *Journal of sleep research*, 32 (2). doi:10.1111/JSR.13748.

Martelli, V., Carelli, E., Tomlinson, G.A., et al. (2022) Prevalence of elevated hemoglobin and hematocrit levels in patients with obstructive sleep apnea and the impact of treatment with continuous positive airway pressure: a meta-analysis. *Hematology (Amsterdam, Netherlands)*, 27 (1): 889–901. doi:10.1080/16078454.2022.2109346.

Martin, J., Girling, A., Nirantharakumar, K., et al. (2016) Intra-cluster and inter-period correlation coefficients for cross-sectional cluster randomised controlled trials for type-2 diabetes in UK primary care. *Trials*, 17 (1): 1–12. doi:10.1186/S13063-016-1532-9/TABLES/7.

Masson, W., Barbagelata, L., Lobo, M., et al. (2023) Plasma Lipoprotein(a) Levels in Polycystic Ovary Syndrome: A Systematic Review and Meta-analysis. *High blood pressure & cardiovascular prevention: the official journal of the Italian Society of Hypertension*. doi:10.1007/S40292-023-00585-2.

McCartney, C.R., Campbell, R.E., Marshall, J.C., et al. (2022) The role of gonadotropin-releasing hormone neurons in polycystic ovary syndrome. *Journal of neuroendocrinology*, 34 (5). doi:10.1111/JNE.13093.

McNamee, R. (2005) Regression modelling and other methods to control confounding. *Occupational and Environmental Medicine*. 62 (7). doi:10.1136/oem.2002.001115.

De Medeiros, S.F., De Medeiros, M.A.S., Ormond, C.M., et al. (2018) Subclinical Hypothyroidism Impact on the Characteristics of Patients with Polycystic Ovary Syndrome.

A Meta-Analysis of Observational Studies. *Gynecologic and Obstetric Investigation*. 83 (2). doi:10.1159/000485619.

de Medeiros, S.F., Yamamoto, M.M.W., de Medeiros, M.A.S., et al. (2020) Changes in clinical and biochemical characteristics of polycystic ovary syndrome with advancing age. *Endocrine Connections*, 9 (2). doi:10.1530/EC-19-0496.

MedSci, K.K.F., Schnecke, V., Haase, C.L., et al. (2023) Weight change and risk of obesity-related complications: A retrospective population-based cohort study of a UK primary care database. *Diabetes, obesity & metabolism*. doi:10.1111/DOM.15154.

Melin, J., Forslund, M., Alesi, S., Piltonen, T., Romualdi, D., Spritzer, P. M., Tay, C. T., Pena, A., Witchel, S. F., Mousa, A., & Teede, H. (2024). Metformin and Combined Oral Contraceptive Pills in the Management of Polycystic Ovary Syndrome: A Systematic Review and Meta-analysis. *The Journal of Clinical Endocrinology & Metabolism*, 109(2), e817–e836. <https://doi.org/10.1210/clinem/dgad465>

Mensch, D. (2023) Optimizing electronic health record efficiency. *Journal of the American Pharmacists Association : JAPhA*, 63 (3): 702–703. doi:10.1016/J.JAPH.2023.04.014.

Merkin, S.S., Azziz, R., Seeman, T., et al. (2011) Socioeconomic status and polycystic ovary syndrome. *Journal of Women's Health*, 20 (3). doi:10.1089/jwh.2010.2303.

Miazgowski, T., Martopullo, I., Widecka, J., et al. (2021) National and regional trends in the prevalence of polycystic ovary syndrome since 1990 within Europe: The modeled estimates from the Global Burden of Disease Study 2016. *Archives of Medical Science*. 17 (2). doi:10.5114/aoms.2019.87112.

Miller, D.P., Gomberg-Maitland, M. and Humbert, M. (2012) Survivor bias and risk assessment. *European Respiratory Journal*. 40 (3). doi:10.1183/09031936.00094112.

Mitchell, M.N. (2012) *A Visual Guide to Stata Graphics, Third Edition*.

Moggetti, P. and Tosi, F. (2021) Insulin resistance and PCOS: chicken or egg? *Journal of*

endocrinological investigation, 44 (2): 233–244. doi:10.1007/S40618-020-01351-0.

Mokhlesi, B., Sam, S. and Ehrmann, D.A. (2017) Obstructive sleep apnea and polycystic ovary syndrome: cause or association? *Sleep Medicine*. 36 pp. 170–171. doi:10.1016/j.sleep.2017.01.001.

Molina, A., Huck, O., Herrera, D., et al. (2023) The association between respiratory diseases and periodontitis: A systematic review and meta-analysis. *Journal of clinical periodontology*, 50 (6). doi:10.1111/JCPE.13767.

Morgan, C.L., Jenkins-Jones, S., Currie, C.J., et al. (2012) Evaluation of adverse outcome in young women with polycystic ovary syndrome versus matched, reference controls: a retrospective, observational study. *The Journal of clinical endocrinology and metabolism*, 97 (9): 3251–60. doi:10.1210/jc.2012-1690.

Mowbray, F.I., Manlongat, D. and Shukla, M. (2022) Sensitivity Analysis: A Method to Promote Certainty and Transparency in Nursing and Health Research. *The Canadian journal of nursing research = Revue canadienne de recherche en sciences infirmieres*, 54 (4). doi:10.1177/08445621221107108.

Nadir, A., Temizkan, S., Ozderya, A., et al. (2019) Dehydroepiandrosterone sulfate/free androgen index ratio predicts a favorable metabolic profile in patients with polycystic ovary syndrome. *Gynecological endocrinology: the official journal of the International Society of Gynecological Endocrinology*, 35 (2): 128–132. doi:10.1080/09513590.2018.1505843.

Nakasian, S.S., Rassen, J.A. and Franklin, J.M. (2017) Effects of expanding the look-back period to all available data in the assessment of covariates. *Pharmacoepidemiology and Drug Safety*, 26 (8). doi:10.1002/pds.4210.

Natsky, A.N., Chai-Coetzer, C.L., Vakulin, A., et al. (2022) Economic evaluation of

diagnostic sleep studies for obstructive sleep apnoea in the adult population: a systematic review. *Sleep Medicine Reviews*. 62. doi:10.1016/j.smrv.2022.101608.

NHS (2014) *Personalised Health and Care 2020 Using Data and Technology to Transform Outcomes for Patients and Citizens A Framework for Action*.

NHS Digital (2023) *General practice workforce 2023 May*. Available at: <https://digital.nhs.uk/data-and-information/publications/statistical/general-and-personal-medical-services/31-may-2023> (Accessed: 20 July 2023).

NICE (2022) *NICE real-world evidence framework Corporate document*. Available at: www.nice.org.uk/corporate/ecd9.

Nicholls, S.G., Langan, S.M. and Benchimol, E.I. (2017) Routinely collected data: The importance of high-quality diagnostic coding to research. *CMAJ*. 189 (33). doi:10.1503/cmaj.170807.

Nicholls, S.G., Quach, P., Von Elm, E., et al. (2015) The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement: Methods for arriving at consensus and developing reporting guidelines. *PLoS ONE*, 10 (5). doi:10.1371/journal.pone.0125620.

Nicholson, A., Ford, E., Davies, K.A., et al. (2013) Optimising use of electronic health records to describe the presentation of rheumatoid arthritis in primary care: a strategy for developing code lists. *PloS one*, 8 (2): e54878. doi:10.1371/journal.pone.0054878.

Nicholson, A., Tate, A.R., Koeling, R., et al. (2011) What does validation of cases in electronic record databases mean? The potential contribution of free text. *Pharmacoepidemiology and Drug Safety*, 20 (3). doi:10.1002/pds.2086.

Nwankwo, M., Okamkpa, C.J. and Danborn, B. (2022) Comparison of diagnostic criteria

and prevalence of metabolic syndrome using WHO, NCEP-ATP III, IDF and harmonized criteria: A case study from urban southeast Nigeria. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 16 (12). doi:10.1016/j.dsx.2022.102665.

Nylander, M., Frøssing, S., Bjerre, A.H., et al. (2017) Ovarian morphology in polycystic ovary syndrome: Estimates from 2D and 3D ultrasound and magnetic resonance imaging and their correlation to anti-Müllerian hormone. *Acta Radiologica*, 58 (8): 997–1004. doi:10.1177/0284185116676656.

Office for National Statistics (2022) *Trends in patient-to-staff numbers at GP practices in England 2022*.

Ohno-Machado, L. (2014) Electronic health record systems: risks and benefits. *Journal of the American Medical Informatics Association: JAMIA*. 21 (e1). doi:10.1136/amiajnl-2014-002635.

Okoth, K., Smith, W.P., Thomas, G.N., et al. (2023) The association between menstrual cycle characteristics and cardiometabolic outcomes in later life: a retrospective matched cohort study of 704,743 women from the UK. *BMC medicine*, 21 (1). doi:10.1186/s12916-023-02794-x.

O'Reilly, M.W., Glisic, M., Kumarendran, B., et al. (2019) Serum testosterone, sex hormone-binding globulin and sex-specific risk of incident type 2 diabetes in a retrospective primary care cohort. *Clinical Endocrinology*, 90 (1): 145–154. doi:10.1111/cen.13862.

O'Reilly, M.W., Taylor, A.E., Crabtree, N.J., et al. (2014) Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *The Journal of clinical endocrinology and metabolism*, 99 (3): 1027–36. doi:10.1210/jc.2013-3399.

Oyinlola, J.O., Campbell, J. and Kousoulis, A.A. (2016) Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Services Research*. 16 (1). doi:10.1186/s12913-016-1562-8.

Ozonze, O., Scott, P.J. and Hopgood, A.A. (2023) Automating Electronic Health Record Data Quality Assessment. *Journal of Medical Systems*. 47 (1). doi:10.1007/s10916-022-01892-2.

Painter, J.L. (2010) “Toward automating an inference model on unstructured terminologies: Oxmis case study.” *In Advances in Experimental Medicine and Biology*. 2010. doi:10.1007/978-1-4419-5913-3_71.

Pardridge, W.M. (1986) Serum bioavailability of sex steroid hormones. *Clinics in Endocrinology and Metabolism*, 15 (2): 259–278. doi:10.1016/S0300-595X(86)80024-X.

Pasquali, R. and Gambineri, A. (2015) A comprehensive approach in diagnosing the polycystic ovary syndrome. *Women's Health*. 11 (4) pp. 501–512. doi:10.2217/whe.15.13.

Peddemul, A., Tejovath, S., Hassan, D., et al. (2022) Influence of Subclinical Hypothyroidism on Women With Polycystic Ovary Syndrome: A Literature Review. *Cureus*. doi:10.7759/cureus.28468.

Pei, C.Z., Jin, L. and Baek, K.H. (2021) Pathogenetic analysis of polycystic ovary syndrome from the perspective of omics. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 142. doi:10.1016/J.BIOPHA.2021.112031.

Petersen, I., Welch, C.A., Nazareth, I., et al. (2019) Health indicator recording in UK primary care electronic health records: Key implications for handling missing data. *Clinical Epidemiology*, 11. doi:10.2147/CLEP.S191437.

Petrides, A.K., Bixho, I., Goonan, E.M., et al. (2017) The benefits and challenges of an

interfaced electronic health record and laboratory information system: Effects on laboratory processes. *Archives of Pathology and Laboratory Medicine*, 141 (3). doi:10.5858/arpa.2016-0146-OA.

Pham, A., Cummings, M., Lindeman, C., et al. (2019) Recognizing misclassification bias in research and medical practice. *Family Practice*, 36 (6). doi:10.1093/fampra/cmy130.

Pines, J.M. (2006) Profiles in patient safety: Confirmation bias in emergency medicine. *Academic Emergency Medicine*, 13 (1). doi:10.1197/j.aem.2005.07.028.

Poorolajal, J., Cheraghi, Z., Irani, A.D., et al. (2011) Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. *Epidemiology and Health*, 33. doi:10.4178/epih/e2011005.

Qiu, S., Cai, X., Sun, Z., et al. (2017) Association between physical activity and risk of nonalcoholic fatty liver disease: a meta-analysis. *Therapeutic advances in gastroenterology*, 10 (9): 701–713. doi:10.1177/1756283X17725977.

Quek, J., Chan, K.E., Wong, Z.Y., et al. (2023) Global prevalence of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in the overweight and obese population: a systematic review and meta-analysis. *The Lancet Gastroenterology and Hepatology*, 8 (1). doi:10.1016/S2468-1253(22)00317-X.

Raciborski, R. (2008) Kountry: A Stata utility for merging cross-country data from multiple sources. *Stata Journal*, 8 (3). doi:10.1177/1536867x0800800305.

Raeisi, T., Rezaie, H., Darand, M., et al. (2021) Circulating resistin and follistatin levels in obese and non-obese women with polycystic ovary syndrome: A systematic review and meta-analysis. *PLoS ONE*. 16 (3 March). doi:10.1371/journal.pone.0246200.

Rajbhandari, P., Auron, M., Worley, S., et al. (2021) Improving documentation of inpatient

problem list in electronic health record: A quality improvement project. *Journal of Patient Safety*, 17 (8). doi:10.1097/PTS.0000000000000490.

Ramezani Tehrani, F., Amiri, M., Behboudi-Gandevani, S., et al. (2020) Cardiovascular events among reproductive and menopausal age women with polycystic ovary syndrome: a systematic review and meta-analysis. *Gynecological Endocrinology*, 36 (1): 12–23. doi:10.1080/09513590.2019.1650337.

Ranganathan, P., Pramesh, C. and Buyse, M. (2016) Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research*, 7 (2). doi:10.4103/2229-3485.179436.

Rao, V.S., Cowan, S., Armour, M., et al. (2022) A Global Survey of Ethnic Indian Women Living with Polycystic Ovary Syndrome: Co-Morbidities, Concerns, Diagnosis Experiences, Quality of Life, and Use of Treatment Methods. *International journal of environmental research and public health*, 19 (23). doi:10.3390/IJERPH192315850.

Ravani, P., Parfrey, P.S., Dicks, E., et al. (2007) Clinical research of kidney diseases II: Problems of study design. *Nephrology Dialysis Transplantation*, 22 (10). doi:10.1093/ndt/gfm433.

RCGP, BMA and DH (2011) *The Good Practice Guidelines for GP electronic patient records*.

Read, J.D. (2009) “Read codes and medical audit.” In *Medical Audit*. Cambridge University Press. pp. 125–137. doi:10.1017/cbo9780511526718.012.

Rebar, R. (2000) *Evaluation of Amenorrhea, Anovulation, and Abnormal Bleeding*. MDText.com, Inc. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK279144/> (Downloaded: 23 July 2022).

Rector, A.L. (1999) Clinical terminology: Why is it so hard? *Methods of Information in Medicine*. 38 (4–5). doi:10.1055/s-0038-1634418.

Rees, D.A., Jenkins-Jones, S. and Morgan, C.L. (2016) Contemporary Reproductive Outcomes for Patients With Polycystic Ovary Syndrome: A Retrospective Observational Study. *The Journal of clinical endocrinology and metabolism*, 101 (4): 1664–1672. doi:10.1210/JC.2015-2682.

Reeves, A., Elliott, M.R., Lewis, T.T., et al. (2022) Study Selection Bias and Racial or Ethnic Disparities in Estimated Age at Onset of Cardiometabolic Disease Among Midlife Women in the US. *JAMA network open*, 5 (11). doi:10.1001/jamanetworkopen.2022.40665.

Reis, Z.S.N., Maia, T.A., Marcolino, M.S., et al. (2017) Is there evidence of cost benefits of electronic medical records, standards, or interoperability in hospital information systems? overview of systematic reviews. *JMIR Medical Informatics*. 5 (3). doi:10.2196/medinform.7400.

Rempert, A.N., Sarria, I., Standeven, L.R., et al. (2023) A Systematic Review of the Psychosocial Impact of Polycystic Ovarian Syndrome Before and After Treatment. *Reproductive sciences (Thousand Oaks, Calif.)*. doi:10.1007/S43032-023-01285-X.

Riazi, K., Azhari, H., Charette, J.H., et al. (2022) The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *The lancet. Gastroenterology & hepatology*, 7 (9): 851–861. doi:10.1016/S2468-1253(22)00165-0.

Riestenberg, C., Jagasia, A., Markovic, D., et al. (2022) Health Care-Related Economic Burden of Polycystic Ovary Syndrome in the United States: Pregnancy-Related and Long-Term Health Consequences. *The Journal of clinical endocrinology and metabolism*, 107 (2): 575–585. doi:10.1210/CLINEM/DGAB613.

Riley, J., Antza, C., Kempegowda, P., et al. (2021) Social deprivation and incident diabetes-related foot disease in patients with type 2 diabetes: A population-based cohort study. *Diabetes Care*, 44 (3): 731–739. doi:10.2337/dc20-1027.

Rivas-Ruiz, F., Pérez-Vicente, S. and González-Ramírez, A.R. (2013) Bias in clinical epidemiological study designs. *Allergologia et Immunopathologia*, 41 (1). doi:10.1016/j.aller.2012.04.005.

Robinson, D., Schulz, E., Brown, P., et al. (1997) Updating the Read Codes: User-interactive Maintenance of a Dynamic Clinical Vocabulary. *Journal of the American Medical Informatics Association*, 4 (6). doi:10.1136/jamia.1997.0040465.

La Rochelle, P. (2014) Basic learning concepts in EBM: The bidimensional hierarchy of evidence. *Evidence-Based Medicine*. 19 (3). doi:10.1136/eb-2013-101610.

Rodgers, L.R., Weedon, M.N., Henley, W.E., et al. (2017) Cohort profile for the MASTERMIND study: Using the Clinical Practice Research Datalink (CPRD) to investigate stratification of response to treatment in patients with type 2 diabetes. *BMJ Open*, 7 (10). doi:10.1136/bmjopen-2017-017989.

Romitti, M., Fabris, V.C., Ziegelmann, P.K., et al. (2018) Association between PCOS and autoimmune thyroid disease: a systematic review and meta-analysis. *Endocrine connections*, 7 (11): 1158–1167. doi:10.1530/EC-18-0309.

Rooney, M.R., Fang, M., Ogurtsova, K., et al. (2023) Global Prevalence of Prediabetes. *Diabetes care*. doi:10.2337/DC22-2376.

Rosenfield, R.L. (2013) Clinical review: Adolescent anovulation: maturational mechanisms and implications. *The Journal of clinical endocrinology and metabolism*, 98 (9): 3572–3583. doi:10.1210/JC.2013-1770.

Rothman, K.J., Wise, L.A. and Hatch, E.E. (2011) Should graphs of risk or rate ratios be plotted on a log scale? *American Journal of Epidemiology*. 174 (3). doi:10.1093/aje/kwr156.

Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group (2004) Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Human reproduction (Oxford, England)*, 19 (1): 41–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14688154> (Accessed: 14 May 2017).

Rubin, K.H., Andersen, M.S., Abrahamsen, B., et al. (2019) Socioeconomic status in Danish women with polycystic ovary syndrome: A register-based cohort study. *Acta Obstetricia et Gynecologica Scandinavica*, 98 (4). doi:10.1111/aogs.13514.

Rudnicka, E., Duszewska, A.M., Kucharski, M., et al. (2022) Oxidative stress and reproductive function: Oxidative stress in polycystic ovary syndrome. *Reproduction*, 164: F145–F144. doi:10.1530/REP.

Rudnicka, E., Suchta, K., Grymowicz, M., et al. (2021) Chronic low grade inflammation in pathogenesis of pcos. *International Journal of Molecular Sciences*, 22 (7). doi:10.3390/ijms22073789.

Rudrapatna, V.A. and Butte, A.J. (2020) Opportunities and challenges in using real-world data for health care. *Journal of Clinical Investigation*. 130 (2). doi:10.1172/JCI129197.

Ruigómez, A., Brobert, G., Vora, P., et al. (2021) Validation of venous thromboembolism diagnoses in patients receiving rivaroxaban or warfarin in The Health Improvement Network. *Pharmacoepidemiology and Drug Safety*, 30 (2). doi:10.1002/pds.5146.

Saeedi, P., Petersohn, I., Salpea, P., et al. (2019) Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Research and Clinical Practice*, 157. doi:10.1016/j.diabres.2019.107843.

Salisbury, C., Johnson, L., Purdy, S., et al. (2011) Epidemiology and impact of multimorbidity in primary care: A retrospective cohort study. *British Journal of General Practice*, 61 (582). doi:10.3399/bjgp11X548929.

Sammon, C.J. and Petersen, I. (2016) Backdating of events in electronic primary health care data: should one censor at the date of last data collection. *Pharmacoepidemiology and drug safety*, 25 (4): 378–84. doi:10.1002/pds.3981.

Sattar, N. and Preiss, D. (2017) Reverse Causality in Cardiovascular Epidemiological Research: More Common Than Imagined? *Circulation*. 135 (24). doi:10.1161/CIRCULATIONAHA.117.028307.

Schneeweiss, S. (2018) Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clinical Epidemiology*. 10. doi:10.2147/CLEP.S166545.

Schneeweiss, S., Rassen, J.A., Brown, J.S., et al. (2019) Graphical depiction of longitudinal study designs in health care databases. *Annals of Internal Medicine*, 170 (6). doi:10.7326/M18-3079.

Schuemie, M.J., Ryan, P.B., Man, K.K.C., et al. (2019) A plea to stop using the case-control design in retrospective database studies. *Statistics in Medicine*, 38 (22). doi:10.1002/sim.8215.

Schulz, E.B., Barrett, J.W. and Price, C. (1998) Read code quality assurance: From simple syntax to semantic stability. *Journal of the American Medical Informatics Association*. 5 (4). doi:10.1136/jamia.1998.0050337.

Schulz, E.B., Price, C. and Brown, P.J.B. (1997) Symbolic anatomic knowledge representation in the Read Codes version 3: structure and application. *Journal of the American Medical Informatics Association : JAMIA*, 4 (1): 38–48.

doi:10.1136/JAMIA.1997.0040038.

Schumm, L.P. and Thisted, R.A. (2005) Reproducible Research Using Stata. *North American Stata Users' Group Meetings*. (16).

Seaman, H.E., De Vries, C.S. and Farmer, R.D.T. (2003a) Differences in the use of combined oral contraceptives amongst women with and without acne. *Human reproduction (Oxford, England)*, 18 (3): 515–521. doi:10.1093/HUMREP/DEG090.

Seaman, H.E., de Vries, C.S. and Farmer, R.D.T. (2003b) The risk of liver disorders in women prescribed cyproterone acetate in combination with ethinyloestradiol (Dianette): a nested case-control study using the GPRD. *Pharmacoepidemiology and drug safety*, 12 (7): 541–550. doi:10.1002/PDS.857.

Seaman, H.E., de Vries, C.S. and Farmer, R.D.T. (2004) Venous thromboembolism associated with cyproterone acetate in combination with ethinyloestradiol (Dianette): observational studies using the UK General Practice Research Database. *Pharmacoepidemiology and drug safety*, 13 (7): 427–436. doi:10.1002/PDS.896.

Senathirajah, Y., Kaufman, D. and Bakken, S. (2014) The clinician in the driver's seat: Part 2 - Intelligent uses of space in a drag/drop user-composable electronic health record. *Journal of Biomedical Informatics*, 52. doi:10.1016/j.jbi.2014.09.008.

Sendur, S.N. and Yildiz, B.O. (2021) Influence of ethnicity on different aspects of polycystic ovary syndrome: a systematic review. *Reproductive biomedicine online*, 42 (4): 799–818. doi:10.1016/J.RBMO.2020.12.006.

Shaheen, A.A., Kaplan, G.G., Sharkey, K.A., et al. (2021) Impact of major depression and antidepressant use on alcoholic and non-alcoholic fatty liver disease: A population-based study. *Liver International*, 41 (10). doi:10.1111/liv.14973.

Shayeb, M. El, Topfer, L.A., Stafinski, T., et al. (2014) Diagnostic accuracy of level 3 portable sleep tests versus level 1 polysomnography for sleep-disordered breathing: A systematic review and meta-analysis. *CMAJ. Canadian Medical Association Journal*, 186 (1). doi:10.1503/cmaj.130952.

Sheikh, J., Khalil, H., Shaikh, S., et al. (2023) Emotional and psychosexual well-being is influenced by ethnicity and birthplace in women and individuals with polycystic ovary syndrome in the UK and India. *BJOG: an international journal of obstetrics and gynaecology*, 130 (8). doi:10.1111/1471-0528.17428.

Shlipak, M., MD, M.P.H., Angeja, B., et al. (2001) Hormone Therapy and In-Hospital Survival After Myocardial Infarction in Postmenopausal Women. *Circulation*, 104 (19).

Silva, M.S.B. and Campbell, R.E. (2022) Polycystic Ovary Syndrome and the Neuroendocrine Consequences of Androgen Excess. *Comprehensive Physiology*, 12 (2): 3347–3369. doi:10.1002/CPHY.C210025.

Smith, H.C., Saxena, S. and Petersen, I. (2022) Maternal Postnatal Depression and Completion of Infant Immunizations: A UK Cohort Study of 196,329 Mother-Infant Pairs, 2006–2015. *Journal of Clinical Psychiatry*, 83 (4). doi:10.4088/JCP.20M13575.

Smith, N., Wilson, A. and Weekes, T. (1995) Use of Read codes in development of a standard data set. *BMJ*, 311 (7000). doi:10.1136/bmj.311.7000.313.

Sola-Leyva, A., Pérez-Prieto, I., Molina, N.M., et al. (2023) Microbial composition across body sites in polycystic ovary syndrome: a systematic review and meta-analysis. *Reproductive biomedicine online*. doi:10.1016/J.RBMO.2023.03.016.

Sommers, S.C. and Wadman, P.J. (1956) Pathogenesis of polycystic ovaries. *American journal of obstetrics and gynecology*, 72 (1): 160–169. doi:10.1016/S0002-9378(16)37526-3.

Springate, D.A., Kontopantelis, E., Ashcroft, D.M., et al. (2014) ClinicalCodes: An online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PLoS ONE*, 9 (6). doi:10.1371/journal.pone.0099825.

Springate, D.A., Parisi, R., Olier, I., et al. (2017) rEHR: An R package for manipulating and analysing electronic health record data. *PLoS ONE*, 12 (2). doi:10.1371/journal.pone.0171784.

van Staa, T.P. and Sprafka, J.M. (2009) Study of adverse outcomes in women using testosterone therapy. *Maturitas*, 62 (1). doi:10.1016/j.maturitas.2008.11.001.

Stein, I.F. and Leventhal, M.L. (1935) Amenorrhea associated with bilateral polycystic ovaries. *American Journal of Obstetrics and Gynecology*, 29 (2): 181–191. doi:10.1016/S0002-9378(15)30642-6.

Stien, Irving.F. (1952) Bilateral polycystic ovaries. *Journal of the South Carolina Medical Association* (1975), 48 (6): 154–155. doi:10.1016/0002-9378(45)90140-2.

Stram, M., Gigliotti, T., Hartman, D., et al. (2020) Logical observation identifiers names and codes for laboratorians potential solutions and challenges for interoperability. *Archives of Pathology and Laboratory Medicine*. 144 (2). doi:10.5858/arpa.2018-0477-RA.

Subramanian, A., Adderley, N.J., Tracy, A., et al. (2019) Risk of incident obstructive sleep apnea among patients with type 2 diabetes. *Diabetes Care*, 42 (5). doi:10.2337/dc18-2004.

Subramanian, A., Anand, A., Adderley, N.J., et al. (2021) Increased COVID-19 infections in women with polycystic ovary syndrome: a population-based study. *European journal of endocrinology*, 184 (5): 637–645. doi:10.1530/EJE-20-1163.

Subramanian, A., Han, D., Braithwaite, T., et al. (2022a) Angiotensin-converting enzyme inhibitors and risk of age-related macular degeneration in individuals with hypertension.

British Journal of Clinical Pharmacology, 88 (9). doi:10.1111/bcp.15366.

Subramanian, A., Lee, S.I., Phillips, K., et al. (2022b) Polycystic ovary syndrome and risk of adverse obstetric outcomes: a retrospective population-based matched cohort study in England. *BMC Medicine*, 20 (1). doi:10.1186/s12916-022-02473-3.

Suissa, S. (2008) Immortal time bias in pharmaco-epidemiology. *American journal of epidemiology*.

Suissa, S. and Ernst, P. (2020) Avoiding immortal time bias in observational studies. *European Respiratory Journal*. 55 (3). doi:10.1183/13993003.02494-2019.

Sun, J., Wang, Y., Zhang, X., et al. (2020) Prevalence of peripheral neuropathy in patients with diabetes: A systematic review and meta-analysis. *Primary Care Diabetes*, 14 (5): 435–444. doi:10.1016/J.PCD.2019.12.005.

Sun, X., Ioannidis, J.P.A., Agoritsas, T., et al. (2014) How to use a subgroup analysis users' guides to the medical literature. *JAMA*. 311 (4). doi:10.1001/jama.2013.285063.

Sun, X., Yu, J., Luo, J., et al. (2022) Meta-analysis of the association between obstructive sleep apnea and postoperative complications. *Sleep medicine*, 91: 1–11. doi:10.1016/J.SLEEP.2021.11.019.

Sutton-Tyrrell, K., Wildman, R.P., Matthews, K.A., et al. (2005) "Sex hormone-binding globulin and the free androgen index are related to cardiovascular risk factors in multiethnic premenopausal and perimenopausal women enrolled in the study of women across the nation (SWAN)." *In Circulation*. 2005. doi:10.1161/01.CIR.0000157697.54255.CE.

Swift, B., Jain, L., White, C., et al. (2018) Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care. *Clinical and Translational Science*. 11 (5). doi:10.1111/cts.12559.

Taggar, J.S., Coleman, T., Lewis, S., et al. (2012) The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. *BMC public health*, 12: 329. doi:10.1186/1471-2458-12-329.

Talari, K. and Goyal, M. (2020) Retrospective studies - Utility and caveats. *Journal of the Royal College of Physicians of Edinburgh*, 50 (4). doi:10.4997/JRCPE.2020.409.

Tang, A.S.P., Chan, K.E., Quek, J., et al. (2022) Non-alcoholic fatty liver disease increases risk of carotid atherosclerosis and ischemic stroke: An updated meta-analysis with 135,602 individuals. *Clinical and molecular hepatology*, 28 (3): 483–496. doi:10.3350/CMH.2021.0406.

Teede, H.J., Misso, M.L., Costello, M.F., et al. (2018a) Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertility and Sterility*, 110 (3): 364–379. doi:10.1016/j.fertnstert.2018.05.004.

Teede, H.J., Misso, M.L., Costello, M.F., et al. (2018b) Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Human reproduction (Oxford, England)*, 33 (9): 1602–1618. doi:10.1093/humrep/dey256.

Teo, Y.H., Tan, B.K.J., Tan, N.K.W., et al. (2022) Obstructive sleep apnea and the incidence and mortality of gastrointestinal cancers: a systematic review and meta-analysis of 5,120,837 participants. *Journal of Gastrointestinal Oncology*, 13 (6): 2789–2798. doi:10.21037/JGO-22-153/COIF).

Teo, Z.L., Tham, Y.C., Yu, M., et al. (2021) Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis. *Ophthalmology*,

128 (11): 1580–1591. doi:10.1016/j.ophtha.2021.04.027.

Tian, D., Chen, W., Xu, Q., et al. (2022) Liraglutide monotherapy and add on therapy on obese women with polycystic ovarian syndromes: a systematic review and meta-analysis. *Minerva medica*, 113 (3): 542–550. doi:10.23736/S0026-4806.21.07085-3.

Tian, Z., Zhang, Y.C., Wang, Y., et al. (2021) Effects of bariatric surgery on patients with obesity and polycystic ovary syndrome: a meta-analysis. *Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery*, 17 (8): 1399–1408. doi:10.1016/J.SOARD.2021.04.009.

Tosi, F., Villani, M., Garofalo, S., Faccin, G., Bonora, E., Fiers, T., Kaufman, J. M., & Moghetti, P. (2022). Clinical Value of Serum Levels of 11-Oxygenated Metabolites of Testosterone in Women With Polycystic Ovary Syndrome. *Journal of Clinical Endocrinology and Metabolism*, 107(5), E2047–E2055. <https://doi.org/10.1210/clinem/dgab920>

Toulis, K.A., Willis, B.H., Marshall, T., et al. (2017) All-cause mortality in patients with diabetes under treatment with dapagliflozin: a population-based, open-cohort study in THIN database. *The Journal of Clinical Endocrinology & Metabolism*. doi:10.1210/jc.2016-3446.

Treasure, W. (2011) The primary care electronic health record: Who’s righting the software? *British Journal of General Practice*, 61 (583). doi:10.3399/bjgp11X556443.

Tremellen, K. and Pearce, K. (2012) Dysbiosis of Gut Microbiota (DOGMA)--a novel theory for the development of Polycystic Ovarian Syndrome. *Medical hypotheses*, 79 (1): 104–112. doi:10.1016/J.MEHY.2012.04.016.

Turcu, A.F. and Auchus, R.J. (2017) Clinical significance of 11-oxygenated androgens. *Current opinion in endocrinology, diabetes, and obesity*, 24 (3): 252–259. doi:10.1097/MED.0000000000000334.

Tyrer, F., Bhaskaran, K. and Rutherford, M.J. (2022) Immortal time bias for life-long conditions in retrospective observational studies using electronic health records. *BMC Medical Research Methodology*, 22 (1). doi:10.1186/s12874-022-01581-1.

Unwin, N., Shaw, J., Zimmet, P., et al. (2002) “Impaired glucose tolerance and impaired fasting glycaemia: The current status on definition and intervention.” *In Diabetic Medicine*. 2002. doi:10.1046/j.1464-5491.2002.00835.x.

Vandenbroucke, J. and Pearce, N. (2015) Point: Incident Exposures, Prevalent Exposures, and Causal Inference: Does Limiting Studies to Persons Who Are Followed from First Exposure Onward Damage Epidemiology? *American Journal of Epidemiology*, 182 (10). doi:10.1093/aje/kwv225.

VanHise, K., Wang, E.T., Norris, K., et al. (2023) Racial and ethnic disparities in polycystic ovary syndrome. *Fertility and Sterility*. 119 (3) pp. 348–354. doi:10.1016/j.fertnstert.2023.01.031.

Vatcheva, K., Lee, M., McCormick, J., et al. (2016) The Effect of Ignoring Statistical Interactions in Regression Analyses Conducted in Epidemiologic Studies: An Example with Survival Analysis Using Cox Proportional Hazards Regression Model. *Epidemiology: Open Access*, 06 (01). doi:10.4172/2161-1165.1000216.

Venn, J. (1880) On the diagrammatic and mechanical representation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10 (59): 1–18. doi:10.1080/14786448008626877.

Vezyridis, P. and Timmons, S. (2016) Evolution of primary care databases in UK: A scientometric analysis of research output. *BMJ Open*, 6 (10). doi:10.1136/bmjopen-2016-012785.

Vounzoulaki, E., Khunti, K., Miksza, J.K., et al. (2022) Screening for type 2 diabetes after a

diagnosis of gestational diabetes by ethnicity: A retrospective cohort study. *Primary Care Diabetes*, 16 (3). doi:10.1016/j.pcd.2022.03.008.

Vuokko, R., Vakkuri, A. and Palojoki, S. (2023) Systematized Nomenclature of Medicine—Clinical Terminology (SNOMED CT) Clinical Use Cases in the Context of Electronic Health Record Systems: Systematic Literature Review. *JMIR Medical Informatics*. 11. doi:10.2196/43750.

Vusirikala, A., Thomas, T., Bhala, N., et al. (2020) Impact of obesity and metabolic health status in the development of non-alcoholic fatty liver disease (NAFLD): A United Kingdom population-based cohort study using the health improvement network (THIN). *BMC Endocrine Disorders*, 20 (1). doi:10.1186/s12902-020-00582-9.

Wall, H., Smith, C. and Hubbard, R. (2012) Body mass index and obstructive sleep apnoea in the UK: a cross-sectional study of the over-50s. *Primary care respiratory journal : journal of the General Practice Airways Group*, 21 (4): 371–6. doi:10.4104/pcrj.2012.00053.

Wang, A., Mo, T., Li, Q., et al. (2019) The effectiveness of metformin, oral contraceptives, and lifestyle modification in improving the metabolism of overweight women with polycystic ovary syndrome: a network meta-analysis. *Endocrine*, 64 (2): 220–232. doi:10.1007/S12020-019-01860-W.

Wang, C., Xu, F., Chen, M., et al. (2022) Association of Obstructive Sleep Apnea-Hypopnea Syndrome with hearing loss: A systematic review and meta-analysis. *Frontiers in neurology*, 13. doi:10.3389/FNEUR.2022.1017982.

Wang, E.T., Cirillo, P.M., Vittinghoff, E., et al. (2011) Menstrual irregularity and cardiovascular mortality. *Journal of Clinical Endocrinology and Metabolism*, 96 (1). doi:10.1210/jc.2010-1709.

Wang, E.T., Kao, C.N., Shinkai, K., et al. (2013) Phenotypic comparison of Caucasian and

Asian women with polycystic ovary syndrome: a cross-sectional study. *Fertility and Sterility*, 100 (1): 214–218. doi:10.1016/J.FERTNSTERT.2013.03.010.

Wang, R., Lagakos, S.W., Ware, J.H., et al. (2007) Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *New England Journal of Medicine*, 357 (21). doi:10.1056/nejmsr077003.

Wang, R. and Mol, B.W.J. (2017) The Rotterdam criteria for polycystic ovary syndrome: Evidence-based criteria? *Human Reproduction*, 32 (2): 261–264. doi:10.1093/humrep/dew287.

Wang, R. and Ware, J.H. (2013) Detecting Moderator Effects Using Subgroup Analyses. *Prevention Science*, 14 (2): 111–120. doi:10.1007/s11121-011-0221-x.

Wang, S. V., Schneeweiss, S., Berger, M.L., et al. (2017) Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Value in Health*, 20 (8). doi:10.1016/j.jval.2017.08.3018.

Wang, X., Piantadosi, S., Le-Rademacher, J., et al. (2021) Statistical Considerations for Subgroup Analyses. *Journal of Thoracic Oncology*, 16 (3). doi:10.1016/j.jtho.2020.12.008.

Wardle, M. and Spencer, A. (2017) Implementation of SNOMED CT in an online clinical database. *Future Hospital Journal*, 4 (2). doi:10.7861/futurehosp.4-2-126.

Warlick, R.D. and Tardiff, B.E. (1998) What in the Word? Development of a Structured Clinical Vocabulary Database. *Proceedings of the AMIA Symposium*, p. 1094. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232136/> (Accessed: 23 July 2023).

Warner, J.L., Smith, J. and Wright, A. (2019) It's time to Wikify clinical documentation: How collaborative authorship can reduce the burden and improve the quality of the electronic health record. *Academic Medicine*. 94 (5). doi:10.1097/ACM.0000000000002613.

Weiskopf, N.G., Dorr, D.A., Jackson, C., et al. (2023) Healthcare utilization is a collider: an introduction to collider bias in EHR data reuse. *Journal of the American Medical Informatics Association*. 30 (5). doi:10.1093/jamia/ocad013.

Wiencek, J.R., McCartney, C.R., Chang, A.Y., et al. (2019) Challenges in the assessment and diagnosis of polycystic ovary syndrome. *Clinical Chemistry*. 65 (3). doi:10.1373/clinchem.2017.284331.

Wierman, M.E. (2021) Hyperandrogenic Anovulation: Differential Diagnosis and Evaluation. *Endocrinology and Metabolism Clinics of North America*. 50 (1) pp. 1–10. doi:10.1016/j.ecl.2020.12.003.

Wijeyaratne, C.N., Balen, A.H., Barth, J.H., et al. (2002) Clinical manifestations and insulin resistance (IR) in polycystic ovary syndrome (PCOS) among South Asians and Caucasians: Is there a difference? *Clinical Endocrinology*, 57 (3): 343–350. doi:10.1046/J.1365-2265.2002.01603.X.

Williams, R., Kontopantelis, E., Buchan, I., et al. (2017) Clinical code set engineering for reusing EHR data for research: A review. *Journal of Biomedical Informatics*. 70. doi:10.1016/j.jbi.2017.04.010.

Witchel, S.F., Oberfield, S.E. and Peña, A.S. (2019) Polycystic Ovary Syndrome: Pathophysiology, Presentation, and Treatment with Emphasis on Adolescent Girls. *Journal of the Endocrine Society*. 3 (8) pp. 1545–1573. doi:10.1210/js.2019-00078.

Wolf, A., Dedman, D., Campbell, J., et al. (2019) Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International Journal of Epidemiology*, 48 (6). doi:10.1093/ije/dyz034.

Wolf, W.M., Wattick, R.A., Kinkade, O.N., et al. (2018) Geographical Prevalence of Polycystic Ovary Syndrome as Determined by Region and Race/Ethnicity. *International*

journal of environmental research and public health, 15 (11). doi:10.3390/IJERPH15112589.

World Health Organisation (2011) *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus Abbreviated Report of a WHO Consultation*.

World Health Organization and International Diabetes Federation (2006) *Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation*. 2006th ed.

Xiao, T., Whitmore, G.A., He, X., et al. (2012) Threshold regression for time-to-event analysis: The stthreg package. *Stata Journal*, 12 (2). doi:10.1177/1536867x1201200206.

Xing, Y., Chen, J., Liu, J., et al. (2021) The Impact of Subclinical Hypothyroidism on Patients with Polycystic Ovary Syndrome: A Meta-Analysis. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme*, 53 (6): 382–390. doi:10.1055/A-1463-3198.

Yang, R., Li, Q., Zhou, Z., et al. (2022) Changes in the Prevalence of Polycystic Ovary Syndrome in China Over the Past Decade. *SSRN Electronic Journal*. doi:10.2139/ssrn.4048835.

Yildiz, B.O., Bolour, S., Woods, K., et al. (2010) Visually scoring hirsutism. *Human reproduction update*, 16 (1): 51–64. doi:10.1093/humupd/dmp024.

Yilmaz, B., Vellanki, P., Ata, B., et al. (2018) Metabolic syndrome, hypertension, and hyperlipidemia in mothers, fathers, sisters, and brothers of women with polycystic ovary syndrome: a systematic review and meta-analysis. *Fertility and sterility*, 109 (2): 356-364.e32. doi:10.1016/J.FERTNSTERT.2017.10.018.

Yu, O., Christ, J.P., Schulze-Rath, R., et al. (2023) Incidence, prevalence, and trends in polycystic ovary syndrome diagnosis: a United States population-based study from 2006 to

2019. *American Journal of Obstetrics and Gynecology*. doi:10.1016/j.ajog.2023.04.010.

Yu, Y.H., Filion, K.B., Reynier, P., et al. (2021) Use of levothyroxine among pregnant women with subclinical hypothyroidism in the United Kingdom: A population-based assessment. *Pharmacology Research and Perspectives*, 9 (5). doi:10.1002/prp2.848.

Zeng, X., Ren, Y., Wu, K., et al. (2023) Association Between Smoking Behavior and Obstructive Sleep Apnea: A Systematic Review and Meta-Analysis. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*, 25 (3): 364–371. doi:10.1093/NTR/NTAC126.

Zhang, D., Ma, Y., Xu, J., et al. (2022a) Association between obstructive sleep apnea (OSA) and atrial fibrillation (AF): A dose-response meta-analysis. *Medicine (United States)*, 101 (30). doi:10.1097/MD.00000000000029443.

Zhang, F., Mamtani, R., Scott, F.I., et al. (2016) Increasing use of prescription drugs in the United Kingdom. *Pharmacoepidemiology and Drug Safety*, 25 (6). doi:10.1002/pds.3947.

Zhang, F. peng and Zhou, Y. (2013) Analyzing left-truncated and right-censored data under Cox models with long-term survivors. *Acta Mathematicae Applicatae Sinica*, 29 (2). doi:10.1007/s10255-013-0215-5.

Zhang, J., Symons, J., Agapow, P., et al. (2022b) Best practices in the real-world data life cycle. *PLOS Digital Health*, 1 (1). doi:10.1371/journal.pdig.0000003.

Zhang, Y., Peloquin, C.E., Dubreuil, M., et al. (2015) Sleep Apnea and the Risk of Incident Gout: A Population-Based, Body Mass Index-Matched Cohort Study. *Arthritis & rheumatology (Hoboken, N.J.)*, 67 (12): 3298–302. doi:10.1002/art.39330.

Zhang, Z., Reinikainen, J., Adeleke, K.A., et al. (2018) Time-varying covariates and coefficients in Cox regression models. *Annals of Translational Medicine*, 6 (7).

doi:10.21037/atm.2018.02.12.

Zhao, X. and Guan, J. (2018) Autonomic nervous system might be related with circadian rhythms and have the intricate effects in obstructive sleep apnea with metabolic syndrome. *Journal of clinical hypertension (Greenwich, Conn.)*, 20 (10): 1553. doi:10.1111/JCH.13378.

Zhou, Y., Wang, X., Jiang, Y., et al. (2017) Association between polycystic ovary syndrome and the risk of stroke and all-cause mortality: insights from a meta-analysis. *Gynecological endocrinology: the official journal of the International Society of Gynecological Endocrinology*, 33 (12): 904–910. doi:10.1080/09513590.2017.1347779.

Zhu, S., Zhang, B., Jiang, X., et al. (2019) Metabolic disturbances in non-obese women with polycystic ovary syndrome: a systematic review and meta-analysis. *Fertility and sterility*, 111 (1): 168–177. doi:10.1016/J.FERTNSTERT.2018.09.013.

Chapter 2: General methods

2.1. Study design

We used retrospective open cohort (Talari and Goyal, 2020) study design to determine the risk of NAFLD, OSA and Dysglycaemia (T2DM and prediabetes) among women with PCOS. In the hierarchy of evidence in health research (La Rochelle, 2014), I chose retrospective cohort study as 1) this study focus on three outcomes of a single exposure, 2) an interventional study is not required to achieve the outcome, 3) a prospective study will not be feasible considering the longer latency period required to develop the outcomes and the resource needs, 4) documented evidence of exposure from the time of follow up period, 5) outcomes have developed at the start of the study, 6) availability of The Health Improvement Network - THIN data (Blak et al., 2011) in electronic data format removing the need for data collection or data entry, 7) University of Birmingham having the access to THIN data, and 8) availability of DExtER (Gokhale et al., 2021), a novel data extraction platform developed by our research team.

The cohort in the THIN database is an open cohort, also known as a dynamic cohort. This differs from a closed cohort, which has a fixed start date and end date, and all participants have a fixed period of follow-up unless an outcome occurs beforehand. In contrast, an open cohort allows participants to join or leave at any time during the follow-up period (Ravani et al., 2007).

2.2. Study setting

The data for this retrospective cohort study was derived from The Health Improvement Network (THIN) which collaborated with selected general practices in the UK in gathering the primary care data. Present general practice system in the NHS was established along with

the establishment of NHS in the aftermath of World War II (1948) as a state's responsibility of providing care to everyone without a direct payment but through the tax (Greengross et al., 1999). There are around 6,800 general practice surgeries (GP surgeries) in the UK. The workforce of general practices in the UK in May 2023 includes 36,194 full time equivalent (FTE) of general practitioners, 16,831 FTE of nurses, 16,314 FTE direct patient care staff and 74,638 admin or non-clinical staff (NHS Digital, 2023). Patient per doctor in the general practices was 1700 in the October 2022 (Office for National Statistics, 2022). GP surgeries have approximately five annual consultations rates per person with each consultation lasting nearly nine minutes and used predominantly by females especially in the reproductive age group (Hobbs et al., 2016). Patients with multimorbidity had relatively higher number of consultations when compared with patients without multimorbidity although longitudinal continuity of care level was lower for vast majority of the patients (Salisbury et al., 2011). Patients have 24-hour access to care from their general practitioner, who will coordinate referrals to specialists and communicate the outcome to the patient. Results of tests and investigations are sent directly to the referring GP practice, and the general practitioner is informed of medical treatment provided elsewhere.

2.3. Study data source

2.3.1. Real-world data or routine health data

Real world data are “data relating to patient health or experience, or care delivery collected outside the context of a highly controlled clinical trial” (NICE, 2022). Varied data contributes to real world data such as clinical, medications, health finance, omics data, family history, mobile health, environmental, patient reported data, social media, and literature (Swift et al., 2018) and a classified based on the source, type, temporality, geographic scope, population, and analytical potential. Best practices guidelines are available to improve the real-world data

cycle throughout its acquisition, aggregation or enrichment, maintenance, and usage (Zhang et al., 2022b). Larger and more diverse datasets, longer follow up periods, real world setting as opposed to the context of conventional clinical trials, external validity, relatively lesser ethical considerations, opportunity to detect rare events such as adverse events, rapid availability of data, low cost and potential to generate hypothesis favours the use of real-world data. When compared with primary data studies (either observational studies with or without invasive tests for research purposes or interventional studies), real world data would have relatively lesser ethical issues due to its 1) purpose of collection of data for routine management and, 2) likelihood of more strict data security and privacy. Despite all these benefits, real world data or real-world evidence is susceptible for data quality or limited clinical information, sizable potential for bias such as selection, misclassification, temporal and publications biases, confounding factors especially from lack of randomisation, data integration challenges and issues related to confidentiality, data ownership and access issues (Rudrapatna and Butte, 2020). Diverse uses of real-world data include, but not limited, determination of effectiveness of routine treatment or management, assessing the effectiveness of implementation of healthcare innovation, estimates required to design randomized control trials and generation of synthetic control (Basch and Schrag, 2019). Clinical guidelines increasingly use evidence from real world data (Oyinlola et al., 2016).

2.3.2. Electronic health data

Electronic health records (EHR) are useful with regards to patient record management, efficient documentation and reporting, billing and coding, care coordination and quality of care (Reis et al., 2017), quality improvement (Carr et al., 2023) and compliance, clinical decision support, medical education, patient engagement, access and reporting, data analytics and research, public health surveillance and population health management (Ohno-Machado,

2014; King et al., 2014; Chao et al., 2013). Interfacing with systems of other units such as laboratory information systems with EHR benefits both clinical unit and the laboratory (Petrides et al., 2017). Inclusion of cloud technology widens the scope of EHR (Ahmadi and Aslani, 2018). Integration of smartphones with electronic patient records enables remote monitoring using patient generated data (Austin et al., 2020) and user composable approaches (Senathirajah et al., 2014). Use of EHR expands beyond the healthcare settings such as to occupational settings (Isakari et al., 2023).

Quality of entry in electronic health record may be hampered by workaround or short-cut methods adopted by users such as ignoring the pop-ups, pre-starting a patient's visit, copy-pasting, using separate text fields, leaving data fields empty, sharing log-in details, entering incorrect data, using paper, using shadow system, giving verbal consent for dispensing medication, separating scanners from its computer on wheels (Boonstra et al., 2021). Quality of EHR could be improved and maintained during the development and implementation such as user centred design, usability testing, user training and education (Lourie et al., 2021) and patient engagement (Warner et al., 2019), during data entry and processing such as improvement of clinical documentation (Cusick et al., 2020; Rajbhandari et al., 2021), standardisation of data entry and data validation and verification, during linkage such as interoperability, incorporating clinical decision support and facilitating continuity of care via single source, data governance such as privacy, data back-up and recovery, security and access control, and during quality assurance such as setting up quality assurance and quality enhancement programs (Ozonze et al., 2023), regular updates and maintenance, user feedback, support and regular auditing and task and error analysis (Carstens et al., 2014; Mensch, 2023).

Although the level of usability of information technology in the emergency department

electronic health records in the UK needs more effort to reach internationally validated standard (Bloom et al., 2021), NHS UK continues to uphold improving the electronic health records in the NHS as a priority (House of Parliament, 2016). Interruptions during the implementation of electronic health record is transient (Jacobs et al., 2019).

2.3.3. General practice electronic health data in the UK

The general practice system in the UK observed a rapid progress in computerisation evident from a sharp rise from 10% to 96% of the practices to use computer for clinical practice by 1996 (Benson, 2002). Still, many problems were encountered such as lack of education, training, and support, not having a common data model to enable interoperability, not having a common coding scheme, lack of information exchange, lack of data validation or verification procedures, system support and underlying research according to a viewpoint (Ellis and Kidd, 2000). Efforts to overcome these problems include use of well tested electronic patient records namely, EMIS, SystemOne, vision or Evolution, incentives to improve data quality such as Quality Outcomes Framework (QOF), anonymised or pseudonymised data sharing, legislation cover such as General Data Protection Regulation (GDPR), IT system based prescription of data, processing of data through primary care clinical systems providers with options to link datasets, development of research databases, system to access primary care data for research and overseeing support by NHS Digital (Bradley et al., 2018). A systematic review finds that majority of the diagnostic codes were well documented (Khan et al., 2010).

2.3.4. General practice databases in the UK

Decades of development of general practice electronic health records have paved way for generation of many primary care databases such as The Health Improvement Network – THIN (Blak et al., 2011), Clinical Practice Research Datalink – CPRD Gold (Herrett et al.,

2015), CPRD Aurum (Wolf et al., 2019), CPRD Gold (Gulliford et al., 2020) and QResearch (Hippisley-Cox et al., 2004). THIN and CPRD Gold have major overlap (Cai et al., 2012; Carbonari et al., 2015). More than 1600 peer reviewed journal articles have used at least one of these databases.

CPRD Aurum database has been used to study various chronic conditions or women health conditions or drug related studies including the association between weight change and 13 obesity related complications including PCOS (MedSci et al., 2023), likelihood of pregnancy (Haas et al., 2023), risk of adverse pregnancy outcome (Subramanian et al., 2022b), screening of T2DM among women with past GDM (Vounzoulaki et al., 2022), PCOS as an associated factor with prescription of levothyroxine among pregnant women with subclinical hypothyroidism (Yu et al., 2021), risk of cardiovascular events (Berni et al., 2021), association with autism (Cherskov et al., 2018) and adverse mental Health and neurodevelopmental outcomes (Berni et al., 2018),

Similarly, CPRD contributed to study burden of hidradenitis suppurativa with PCOS as one of the associated factors (Ingram et al., 2018), response to treatment among T2DM with exclusion of PCOS (Rodgers et al., 2017), reproductive outcomes (Rees et al., 2016), adverse outcome in young women with PCOS (Morgan et al., 2012), risk of ovarian cancer with PCOS as one of the covariates (Bodmer et al., 2011), venous thromboembolism associated with cyproterone acetate in combination with ethinyloestradiol (Dianette) (Seaman et al., 2004), risk of liver disorder in prescribing of cyproterone acetate in combination with ethinyloestradiol (Dianette) (Seaman et al., 2003b) and use of combined oral contraceptives amongst women with and without acne (Seaman et al., 2003a).

QResearch database has been used to study PCOS such as in development and validation of risk prediction algorithm to estimate future risk of T2DM with PCOS as one of the risk

factors (Hippisley-Cox and Coupland, 2017).

Overall, primary care databases in the UK make immense contributions to research outputs. These contributions include an increasing number of research outputs, diverse document types and disciplines, wider co-authorship from many institutions across many countries, and publication in top journals with many citations. (Vezyridis and Timmons, 2016).

2.3.5. The Health Improvement Network data (THIN Data)

2.3.5.1. Introduction to THIN Data

The Health Improvement Network (THIN) was established in 2002 as an anonymised longitudinal open cohort data based on representatively selected UK general practices (Kontopantelis et al., 2018). Data for the THIN data is gathered by the provider of Vision GP records i.e. software In Practice Systems (INPS) from the general practices in the UK which has agreed to join The Health Improvement Network (THIN). This data is provided to an organisation named IMS to facilitate researchers using this data. The nationally accredited committee that approved THIN also provides ethics approval for research using THIN data (IMS Health, 2015).

Although this is a general practice-based data, 98% coverage of registration with a general practitioner in the UK makes this database more valid (Lis and Mann, 1995). THIN data included data of more than 14 million patient records from above 760 practices (Kumarendran et al., 2019).

2.3.5.2. The development of THIN data

When a practice joins THIN, a complete retrospective dataset is collected in a single Full Data Collection. Following that, daily Incremental Data Collections are performed with minimal interruption to daily practice. IMS processes the data three times a year, adding the

updates to the existing data to create the research dataset. Majority of the practices that contribute to THIN have amassed more than 25 years of data in their systems. THIN offers free Vision software training to contributing practices to improve data recording in general practice. This training can be delivered on-site or at training seminars, where experienced Vision trainers can provide advice and expertise on how to use the system to input high-quality data that meets data standards. Initial audits of THIN Data practices were conducted to verify that their data recording met national standards, enabling them to join the scheme. THIN sends feedback reports to practices on a regular basis to highlight areas where completeness of recording could be improved. The aggregated results of the feedback reports can be shared with researchers to help them select subsets of practices with strengths in specific areas of research (IMS Health, 2015).

2.3.5.3. Generalisability and validity of THIN data

THIN data is proven to be generalisable for morbidity and mortality of important health conditions with opportunities for adjustment for demographic factors and social deprivation (Blak et al., 2011). This database uses Read code system which is a hierarchical coding system developed for general practice data (Booth, 1994).

THIN database has been validated for the use many health conditions such as venous thromboembolism (Ruigómez et al., 2021), small cell lung cancer (Cea Soriano et al., 2019b), intracranial bleeding (Cea Soriano et al., 2019a), haemorrhagic stroke (Cea Soriano et al., 2019a), chronic kidney disease (Denburg et al., 2011) and pharmacoepidemiologic research (Lewis et al., 2007).

THIN data base was utilised to study various health conditions such as NAFLD (Shaheen et al., 2021; Vusirikala et al., 2020; Alexander et al., 2018), OSA (Subramanian et al., 2019; Alshehri et al., 2022; Zhang et al., 2015; Guest et al., 2014; Wall et al., 2012),

prediabetes/Diabetes (Gardner et al., 2023; Boursi et al., 2022; Cahn et al., 2020) and testosterone therapy (van Staa and Sprafka, 2009).

THIN data base has been used in PCOS studies including diagnosis and management of PCOS (Ding et al., 2016), incidence and prevalence of diabetes among women with PCOS (Ding et al., 2018a), risk of Covid19 infection among women with PCOS (Subramanian et al., 2021), and prescribing of cyproterone acetate /ethinylestradiol (Cea Soriano et al., 2017a).

2.3.5.4. Data structure of THIN data

The data structure of THIN data includes mainly six files namely, practice files, demography files, patient file, therapy file, medical file and Additional Health Data (AHD) file. Practice file is linked by practice ID with the other files. Files holding patient related information are linked by patient ID. Postcode Variable Indicators (PVI) are included in patient file. Therapy file includes information on prescriptions such as pack size (via pack size code), drug codes and generic medicine prescription and dosages. Medical file captures the medical episodes using medical codes that links with symptoms, sign and procedures and anonymised comments via text ID. AHD file contains data on investigations, lifestyle and preventions and is linked with medical codes, anonymised comments and AHD codes. The consult file can be used to track a patient's therapy, ADHD, and medical history and includes details about where, when, and how long each consultation lasted. Practice file holds dates of Acceptable Mortality Reporting (AMR), computerisation, last collection, Vision date, etc. (IMS Health, 2015)

2.3.5.5. Strength and characteristics of THIN data

THIN data has many strengths arising from its special characteristics. The THIN data collection scheme is non-interventional and reflects real-world data as it is collected from GP practices, which are the first point of contact for all sections of healthcare (Cea-Soriano et al.,

2018). The use of standardised coding systems and national guidelines enables researchers to develop research methodologies using pre-collected data and to confirm population selection and trial length for pre-clinical trials (Martin et al., 2016; Danaei et al., 2018). It is estimated to be approximately 6% of the UK population actively registered, although the size of the THIN data source population is not straightforward to determine (Blak et al., 2011). The THIN dataset serves as a comprehensive record of patient care in primary care, though it may not include all events that are relevant to research (Petersen et al., 2019). The THIN data enables research to get patient counts and linkage within families sharing the same GP practice as it collects anonymized demographic data, including registration and departure dates, age, sex, and household identifiers without identifiable information (Smith et al., 2022). The THIN database collects all medical conditions and symptoms recorded electronically during consultations, referrals to secondary care, hospital admissions, and discharge medication and diagnosis (Eberhardt et al., 2022). Prescribing data in the THIN database is well-recorded despite having a possibility of under-recording for prescriptions not issued electronically (Zhang et al., 2016). The THIN Data includes information on lifestyle and preventative healthcare, in addition to GP consultations and other health care interventions (Breeze et al., 2020). IMS Health offers postcode linked indicators that maintain anonymity for studies utilizing THIN Data, thereby augmenting its value for epidemiological, public health, and health services research purposes (Riley et al., 2021). IMS Health has developed a secure linkage methodology to connect patient records in THIN with their record in Hospital Episodes Statistics – HES (Cea Soriano et al., 2017b), which allows researchers to access secondary care information for research purposes (IMS Health, 2015).

2.4. Codes and code list used in the study

2.4.1. Read codes: The structured clinical vocabulary used in THIN

Structured clinical vocabulary or coded clinical data facilitates documentation of clinical data according to a standard coding system which enables pooling of data documented by more than one practitioner (Warlick and Tardiff, 1998). Structured clinical vocabulary largely replaced ‘free text’ data to improve consistency, standardisation for reporting and surveillance, clinical decision support, patient safety, interoperability and semantic interoperability, quality improvement measures, data aggregation, data analysis and research, health information exchange, efficiency of workflow and billing and reimbursement (Awaysheh et al., 2018). Yet there are room for improvement in coding systems such as incomplete coverage, ambiguity and overlapping concepts, limited flexibility, handling new or rare technologies, lack of standardisation, limitation in expressiveness, complexity of certain cases, time consuming data entry, evolutionary challenges, semantic differences, maintenance and update, cost of implementation and learning curve (Rector, 1999). Common structure clinical vocabulary or clinical coding systems include Oxford Medical Information System - OXMIS codes (Painter, 2010), Read code (Booth, 1994), Systematized Nomenclature of Medicine – Clinical Terms - SNOMED CT (Vuokko et al., 2023), Logical Observation Identifiers Names and Codes – LOINC (Stram et al., 2020), Current Procedural Terminology – CPT (Burns et al., 2020) and Pathology Bounded Code List - PBCL.

James Read originated the Read code system in 1982 with just 25 codes to document clinical data from his own practice (Benson, 2012). Exponential increase in number to more than 250,000 (Chisholm, 1990) and complexity of Read codes raised concern to develop strategies for quality assurance (Schulz et al., 1998). Semantic definition of disorders (Brown et al., 1998) and symbolic anatomic knowledge representation (Schulz et al., 1997) were added to

facilitate understanding of the codes by the users correctly. Read codes were originally developed for general practice and was recognised as standard classification of general practice data for the United Kingdom in the 1988 (Chisholm, 1990). Read code was chosen as the logical choice for the development of a standard data set for the general practice setting (Smith et al., 1995). Later, Read codes were expanded for hospital settings (Read, 2009).

Table 2.1 Summary of coding systems for electronic databases

Full name	Abbreviation
Free text	Free text
Structured Clinical Vocabulary	SCV
Oxford Medical Information System	OXMIS codes
Read code	Read code
Systematized Nomenclature of Medicine – Clinical Terms	SNOMED CT
Logical Observation Identifiers Names and Codes	LOINC
Current Procedural Terminology	CPT
Pathology Bounded Code List	PBCL
Common Structure Clinical Vocabulary	CSCV

Read codes consisted of four-digit alphanumeric codes with each of the four levels to represent more detailed description of the condition. Automatic encoding enables a drop down to list the possible codes as someone starts typing the condition of interest (Benson, 2012). The Read code sets are updated quarterly to reflect the needs of users, but this can lead to tensions between different users and uses of coded clinical data of different versions (Robinson et al., 1997). When compared with Read code system, SNOMED CT has many advantages, namely, consistent sharing of vital information, comprehensive coverage across

clinical specialties, support for clinical decision making, and facilitation of extensive clinical audit and research (NHS, 2014). Considering many advantages of using SNOMED CT over to Read codes (Wardle and Spencer, 2017), NHS initiated a transition to move from Read code to SNOMED CT from 2016 (NHS, 2014) and was expected to complete it by 2022 (Jones, 2022). We used the version of THIN data for this retrospective cohort study which used Read codes.

2.4.2. Preparation of code list

Code list is the set of codes used to define all variables such as exposures, outcomes and covariates in a study using secondary data (NICE, 2022). As opposed to a prospective cohort study which generates data custom made for the study concerned, a retrospective cohort study depends entirely on past data and requires a code list to develop standardised definitions and research methodology to reduce bias such as misclassification bias and to improve time efficiency of data abstraction, consistency with previous studies, transparency, and reproducibility (Nicholls et al., 2017). Understanding this importance of applying a systematic approach in preparing code list (Davé and Petersen, 2009; Nicholson et al., 2013), we developed, adhered to and published an iterative eight-step approach (Kumarendran et al., 2019).

In line with the recommendation to cover the codes broadly than confining only the diagnostic codes, we included Comma Separated Values (CSV) documents containing medical Read codes, drug codes and Additional Health Data (AHD) codes and converted the CSV into Microsoft Excel (MS Excel) workbook format to ease categorisation of codes. Firstly, I sorted the codes based on the systems using the sort column option in the MS Excel document. Secondly, I assigned ranks based on the level of hierarchy based on relevance in the descending order while 1 indicating a confirmed category, 2 indicating codes requiring

second opinion but high likelihood of inclusion, 3 indicating codes requiring second opinion but high likelihood of exclusion and 4 indicating definite exclusion. Similar approach has been recommended previously (Davé and Petersen, 2009). Thirdly, I gathered the code list published in other studies such as validation studies (Nicholson et al., 2011) and looked for the missing but relevant codes. Fourthly, I searched an online repository of clinical codes (<https://clinicalcodes.rss.mhs.man.ac.uk/>) for code list. Use of Online clinical codes repository has been recommended to improve the validity and reproducibility of research using electronic medical records (Springate et al., 2014). Fifthly, Read codes corresponding to Quality and Outcome Framework (QoF) were searched for as these codes were better documented (Taggar et al., 2012). Sixthly, I made a list of key words based on previous articles and textbooks and used find option (Ctrl + F) to look up for codes in the MS Excel document. While using this search option, I used only the relevant letters of the keyword to minimise missing of any codes, for instance, using ‘ovar’ to cover ovary, ovaries and ovarian. Although there was an option to use STATA to get an automated extraction (Davé and Petersen, 2009), I opted to use manual search as the latter would help identifying further relevant codes due to clustering during the eyeballing despite it being a time-consuming process. Seventhly, I conducted discussion with our team which comprised of multidisciplinary experts from relevant to the study such as androgen experts, epidemiologists, endocrinologist, gynaecologist and sleep medicine. Eighthly, I discussed with general practitioners in further improving the code list. The entire process was iterated many times until saturation was reached.

My contribution as co-author in other research (Caleyachetty et al., 2017; Toulis et al., 2017; Dafoulas et al., 2017; O’Reilly et al., 2019) provided me opportunity in improving my skills in developing code list. As per the guidance (NICE, 2022; Williams et al., 2017), we have published code list as supplement material for all three studies added in this thesis.

2.5. Study period

Study period creation needs to consider several time points in this retrospective open cohort study using primary care electronic health database to derive participants. A primary care open cohort would allow patients to join or leave the cohort at point in the follow up or observation period. Recruitment of participants using electronic database requires enrolment after assurance of satisfactory quality of documentation of data in the database. Therefore, derivation of study period will include, but not limited to, cohort start date, acceptable computerisation date, Vision date, acceptable mortality rate date, registration date, index date (or cohort entry date), transfer date, death rate, outcome date, date of last data collection, exit date and study end date (Horsfall et al., 2013; Wang et al., 2017).

The cohort start date, or start date, is a fixed date set by the researcher from which to start looking for data in an existing longitudinal data set. Despite the availability of data for an earlier period, the cohort start date may be chosen on a later date, considering the availability of quality data, changes in clinical practice or administrative guidelines, or patients' behaviour. Considering the start of The Health Improvement Network in 2002 (IMS Health, 2015) and to cover documentation of immediate past records, we chose 1st of January 2000 as the cohort start date. The acceptable computerisation date is the date when the set standard of computerisation of data for the determined items and frequency per patient is attained (Horsfall et al., 2013). Although acceptable computerisation is reached, adaptation to a new electronic health record platform may require further time to attain good quality data entry (de Lusignan et al., 2015; Treasure, 2011). The THIN database derives data from the practices using EMR platform provided by a commercial provider In Practice Systems and named Vison® (Bradley et al., 2018). The Vison date indicates the date of implementation of Vision system in the practice (Subramanian et al., 2022a). The practice registration date

indicates the date in which participant registered with the general practice (registration date). From the date of registration of a participant in the general practice, studies utilising electronic health database allow a lag period for recruitment to enable complete documentation of past records (Schneeweiss et al., 2019).

The index date (or cohort entry date) is set to the date in which a participant completes this lag period provided other eligibility requirements are met such as age. We set the index date one year after the latest of the registration date, Vison date, Acceptable mortality rate to ensure good quality data and to allow sufficient time to record important covariates (Schneeweiss, 2018). The latest among these dates (washout window for exposure) is chosen to improve the sensitivity of incident exposures. A participant becomes eligible for the study from the index date if the age requirement and the matching requirement for the non-exposed group are met. Use of baseline covariates which developed prior to the first documentation of the exposure is essential in establishing the temporal association (Schuemie et al., 2019) and this period is called covariate assessment window (Nakasian et al., 2017). Covariate assessment window allows adjustment for confounding, capture time-dependent effects, and establish a robust temporal association between exposure and outcome (Wagner et al., 2021).

The definition of the index date would vary depending on whether the participant is exposed or not exposed to the exposure of interest. For those who are exposed, the index date would be defined differently for incident and prevalent exposures (Vandenbroucke and Pearce, 2015). The eligible participants are categorized as exposed if they have evidence of exposure documented in the database at any time before or after the cohort entry date, but on or before the study end date. For participants whose exposure is first documented on or after the cohort entry date (i.e. incident exposure), the index date is the date on which their exposure is first

documented. For participants whose exposure is first documented prior to the cohort entry date (i.e. prevalent exposure), the index date is their cohort entry date. For participants who never had evidence of exposure at any time on or prior to the study end date, the index date is set to the same index date as their corresponding exposed patient to avoid immortal time bias (Tyrer et al., 2022).

Likewise, in addition to the timelines related to cohort entry, there are also a set of dates related to exit from the cohort, such as transfer date, death date, outcome date, date of last data collection, exit date and study end date. In the context of a general practice database, the transfer date marks the date on which a patient's registration is struck off with the index practice and they move to another general practice (CPRD, 2022). The Good Practice Guidelines for GP electronic patient records recommends documentation of transfer out of patients from the practice (RCGP et al., 2011). Death date denotes the date on which a patient is dead. Documentation of death is crucial to calculating the person time of follow up and more importantly when one of the outcomes is death. The average lag between the actual date of death and the date of documentation in the primary care database is one month (Gallagher et al., 2019). However, this lag is unlikely to have a considerable adverse impact on the estimates, as studies with a large person-time follow-up typically have a low rate of misclassification. Data are collected as part of routine practice and uploaded to secure servers at a frequency of usually once per month. The date of last data upload from each practice is referred to as the date of last data collection. Censoring the follow-up period ahead of the date of last data collection is recommended for studies using outcomes diagnosed outside of general practice (Sammon and Petersen, 2016). This helps to minimize the adverse impact of the lag period in getting data updated in the general practice data. For those who develop an outcome after the index date (which will be discussed in the next paragraph), follow-up is

censored at the date of documentation of the outcome (Schneeweiss et al., 2019) and we denoted this as outcome date in our study. The cohort end date, or end date, is a fixed date set by the researcher to stop looking for data beyond that date from an existing longitudinal data set.

The cutoff dates discussed so far collectively form two composite endpoints in a retrospective cohort study based on a primary care database: the index date marks the start of the follow-up period, and the exit date marks the end (Langan et al., 2018). Follow-up end date (exit date) is determined from the earliest of outcome date, transfer date, death date or date of last data collection. The earliest among these dates (washout window for outcome) is chosen to improve the sensitivity of the definition of incident outcome. We provided visual representation of study timeline and election process for each study in compliance with the recommendation (Schneeweiss et al., 2019).

2.6. Study population

We generated two types of cohorts with PCOS diagnosis based primary cohort and secondary cohorts for biochemical androgen excess. We could not depend on the primary cohorts based on the PCOS diagnosis for the determination of the association between androgen excess and outcomes (NAFLD, OSA, and dysglycaemia) due to the possible under-documentation of androgen excess. This is because 1) objective assessment of hirsutism and acne during brief encounters in general practice requires longer duration, privacy, conducive room temperature for exposing the body during assessments, especially in winter, and a chaperone, 2) clinical androgen features might be perceived as non-specific symptoms by the patient as well as practitioner, 3) lack of awareness, 4) stigma or embarrassment, 5) subclinical cases, 6) limited access to specialised care, 7) diagnostic challenges, variability in coding practices, 8) lack of

routine screening 9) and missed diagnosis or delayed diagnosis

and 10) challenges in diagnosis of clinical androgen excess among adolescents, and 11) not being included in the quality outcome framework to get incentives (Joham et al., 2022; Cussen et al., 2022; Wiencek et al., 2019; Agapova et al., 2014).

2.6.1. PCOS diagnosis based retrospective cohorts

Women aged 18–50 years at the index date who did not have evidence of documentation of outcome by the index date were eligible for the study. We excluded women with any documentation of excess alcohol use in the OSA outcome study, to minimise the confounding due to alcohol increasing the risk of accumulation of fat in the liver. The nonexposed group comprised eligible women who did not have a documented diagnosis of PCOS and were matched within the same practice for age (± 1 year) and BMI (± 2 kg/m²) of each woman with PCOS. Age and BMI were used as matching variables because of their strong association with both PCOS and outcomes in each study (NAFLD, OSA, and dysglycemia). Practice was matched to minimise the confounding effect of systematic difference between the practice in diagnosis, management, follow-up, and referral. When there were several patients who could be matched, two were randomly matched for each exposed patient. We generated three independent PCOS cohorts for each of the three outcomes namely, NAFLD, OSA and dysglycaemia (Table 2.1). Code lists used to define the exposure, outcomes and covariates are provided under the appendices of the respective studies.

Table 2.1. Distribution of participants among cohorts

	Primary analysis (Prevalent and incident cases)		Sensitivity analysis (Incident cases)		Sensitivity analysis (PCOS specific Read codes only)	
	PCOS	Control	PCOS	Control	PCOS	Control
NAFLD	63,210	121,064	20,162	37,544	19,425	35,967
OSA	76,978	143,077	23,349	43,106	24,603	44,991
Dysglycaemia	64,051	123,545	20,333	38,989	18,780	35,938

* Some women with exposure did not have two controls due to 1) non-availability of eligible women for matching, and 2) women with excessive drinking were excluded from the cohorts for NAFLD outcome, despite the intention to match two women without exposure for each woman with exposure.

The number of women in each exposed group and their corresponding control groups vary for each outcome due to update of databases by THIN and exclusion criteria, if any.

2.6.2. Androgen excess cohorts

A cohort of women with serum testosterone measurements (Testosterone cohort) and another cohort of women with serum SHBG measurements (SHBG cohort) were derived using the same eligibility, entry, and exit criteria as the primary cohort based on PCOS diagnosis status, for the outcome of NAFLD and Dysglycaemia separately. This approach was crucial because the androgen excess features (such as acne, hirsutism and hair loss) and biomarkers related to androgen excess were inadequately documented within the PCOS cohort. Consequently, PCOS cohort lacked sufficient statistical power to detect associations related to these factors. The determination of diabetes risk among women with serum sex hormone measurements was published as part of a larger study (O'Reilly et al., 2019).

Follicle Stimulating Hormone (FSH) and Luteinising Hormone (LH) were not included in the analysis due to 1) the need to consider the timing of sample collection (follicular phase or luteal phase) and mismatch between the levels of these hormones and clinical androgen excess features (ASRM, 2021).

2.7. Data extraction

Data extraction in a retrospective cohort study using electronic database needs to apply timelines (anchors) in a systematic manner considering the calendar time, patient event time and index date (Schneeweiss et al., 2019). Base anchors include date of last collection and study period. Cohort entry date is considered as first order anchor. Second order anchors include washout window for exposure (to define incident exposure), washout window for outcome (to define incident outcome), exclusion assessment window, exclusion assessment window, covariate assessment window, exposure assessment window and follow-up assessment window. Accomplishing all these constraints and then matching for three variables in a large dataset manually is not feasible. Creating Structured Query Language and Natural Language Processing Algorithm specific to each study was a resource intensive process. Therefore, automated extraction software was developed to ease primary care routine data studies such as DExtER (Gokhale et al., 2021), ACE (Callahan et al., 2021) and rEHR (Springate et al., 2017). We used DExtER (Data extraction for epidemiological research) software developed by two members of this research team (Krishna Gokhale and Krish Nirantharakumar) for automated data extraction. This software uses extract, transform and load (ETL) method. DExtER has been used in just above 80 peer reviewed research publications.

Although we could not share the data due to our agreement with IMS Health, any researcher

can request the data through their data request process.

2.8. Data analysis

2.8.1. STATA software for data analysis

We used STATA software for data analysis as STATA would provide robust data management tools to clean (Dupont, 2021), manipulate and merge datasets (Raciborski, 2008), handling large datasets, options for rich statistical methods such as survival analysis (Cleves et al., 2016), reproducibility using do-files (Schumm and Thisted, 2005), graphic visualisation during data cleaning and in generating graphs such as forest plot (Mitchell, 2012), support in using third party user written packages (Xiao et al., 2012), support from active user community and compatibility with other software (Hanson, 2005).

2.8.2. Study population characteristics

As per the NICE guidelines on real-world evidence framework (NICE, 2022), we provided participants flow diagrams to present number of practices eligible for the study, number of patients available in the dataset, number of patients eligible for the study period, number with exposure, number of matched comparison group with no documented evidence of exposure, matching criteria, number excluded (if any) and median years of follow up and number of patients with outcome for exposed and non-exposed. We choose median to present the follow-up (Betensky, 2015) as opposed to mean follow up time, as median is a robust measure to deal with outliers due to participants with unusually short or long term follow-up, variability of follow-up duration in an open cohort, right skewed distribution of follow-up duration, easy to interpret, sharing non-parametric nature, avoids misinterpretation due to missing follow-up times or censoring and clinical relevance due to the estimate for half of the study participants.

The participants' basic characteristics for exposed and non-exposed was tabulated for mean ($\pm$ standard deviation) age in years, quintiles of Townsend deprivation score (and missing data), BMI (<25 , $25-30$, >30 kg/m² and missing or implausible data), smoking status (smokers, non-smoker and missing or implausible data), ethnicity, and chronic conditions such as diabetes, hypertension, hypothyroidism and impaired glucose regulation. Our definition of Impaired Glucose Regulation (IGR) included 1) Impaired Fasting Glucose (IFG) demarcated by fasting plasma glucose of 6.1 to 6.9 mmol/L, and 2) Impaired Glucose Tolerance (IGT) demarcated by plasma glucose of 7.8 to 11.1 mmol/L measured 120 minutes after ingestion of 75g glucose in the oral glucose tolerance test (Unwin et al., 2002).

We provided another table for PCOS diagnostic features (anovulation, hirsutism, acene, alopecia, and polycystic ovaries) and use of medication such as antiandrogens (cyproterone, Drospirenone and other anti-androgen), metformin and lipid modifying drugs.

No statistical tests were performed to compare the patients with PCOS and the control patients at baseline as per the recommendation (Bland and Altman, 1995) to avoid inflation of alpha error (Ranganathan et al., 2016).

2.8.3. Primary analysis

We presented the primary analysis with total number of participants, person-years, number and percentage of outcome, incidence rates per 10,000 person years, for women with PCOS and women without PCOS separately. In an open cohort, incidence rates are calculated as a measure of risk, considering the person-time of follow-up, due to variations in participants' individual follow-up periods (Ravani et al., 2007). Incidence of each outcome, rather than mortality due to each outcome, was considered due to lower baseline risk of mortality due to these outcomes in this cohort with median age of 30 years.

We also presented unadjusted hazard ratios with 95% confidence intervals, and hazard ratios

adjusted for age, Townsend score, BMI category, and diabetes or impaired glucose regulation (the latter was added as baseline and follow-up separately). We opted for BMI categories instead of numerical BMI because categorical BMI allows straight forward interpretation, captures non-linear effects, improves clinical relevance, and reduces the risk of overfitting. Additionally, we considered that the impact of cut-points would be less pronounced due to the lower proportion of Asians. We did not include hormone replacement therapy (HRT) as one of the confounders because the proportion of women using HRT would be very low. This is due to the median age of menopause in the UK being 51 years, while the median age of the cohort was 30 years.

We chose Cox-regression analysis (Koletsis and Pandis, 2017) considering its suitability for time to event analysis (Abd Elhafeez et al., 2021), handling left truncated and right censored data (Zhang and Zhou, 2013), adjustment for time varying covariates (Zhang et al., 2018), control for confounding variables, flexibility in data analysis as Cox regression does not require assumptions about distribution of survival times (Deo et al., 2021), ease of interpretation of hazard ratio as it represents weighted relative risk (George et al., 2020), applicability to large datasets (Bertrand and Maumy-Bertrand, 2021), and easy implementation in STATA statistical software (Lambert, 2017).

2.8.4. Sensitivity analysis

Sensitivity analysis is recommended to ensure robustness of the study (NICE, 2022; Mowbray et al., 2022; von Elm et al., 2007). For each outcome, we conducted two additional sets of sensitivity analyses. The first analysis was limited to women with the PCOS Read code only (excluding women with PCO Read codes only) and their matched comparison group. The second analysis included women with incident cases of PCOS (excluding the women with prevalent cases of PCOS) and their matched comparison group of women

without PCOS.

Limiting the analysis to incident exposure would avoid reverse causality (Sattar and Preiss, 2017), enhance temporal sequence, improve the validity for causal inference, minimise recall bias, better represent source population better by avoiding survival bias, minimise prevalence-incidence bias (Delgado-Rodríguez and Llorca, 2004), and reduce confounding. In the meantime, restricting the analysis only to incidence exposure and their control may reduce the statistical power and provide relatively wider confidence intervals.

2.8.5. Determination if the risk of outcome is independent of BMI

We performed subgroup analyses and Cox regression with interaction term to determine if the association between PCOS and outcomes is independent of BMI. First, we examined whether the risk of outcome (e.g. NAFLD) was independent of BMI status by conducting a subgroup analysis within each BMI category. Subgroup analysis is useful in determine if the association is consistent across different levels of a third variable (in this case, BMI), detect effect modification, identify high risk group, improve clinical decision making, identify potential confounding effect, and to generate hypothesis for further research (Sun et al., 2014; Wang et al., 2007; Farrokhyar et al., 2022; Fletcher, 2007; Dijkman et al., 2009; Burke et al., 2015; Wang et al., 2021).

We developed Cox regression models to determine if the association between PCOS status and outcome (e.g. NAFLD) was independent of BMI by adding interaction term between PCOS status and BMI category. Women without PCOS and with a BMI of less than 25 kg/m² were used as the reference category in the regression model to verify the existence of a dose-response pattern with increasing BMI among women and with PCOS. The interaction terms in a Cox regression model helps to test effect modification, sub-group effects, account for heterogeneity, identify high risk group and uncover hidden relationship, and improve

model fit and prediction accuracy (Madjar et al., 2021; Wang and Ware, 2013; Vatcheva et al., 2016; Brankovic et al., 2019). We graphically presented the findings using dose response hazard plot (Maria-Sophie Barthel and Royston, 2006).

2.8.6. Determining the risk of outcomes among women with PCOS when compared with women without PCOS by adding PCOS diagnostic features as predictor

We performed another sensitivity analysis by adding PCOS diagnostic features as predictors in the Cox regression models to determine the association between PCOS and outcome (e.g., NAFLD) and to statistically adjust for the possibility of misclassification bias in the diagnosis of PCOS in the primary care dataset. These models included PCOS status, age, Townsend index category, BMI, diabetes or IGR and clinical features of androgen excess (hirsutism, alopecia and acne).

2.8.7. Determining the role of factors other than BMI in outcomes among women with PCOS and without PCOS

We developed Cox regression models to determine the role of factors other than BMI in outcomes among women with PCOS and without PCOS independently. These models included age, Townsend category, BMI category, diabetes or impaired glucose regulation, anovulation, individual androgen excess features (hirsutism, acne and alopecia), lipid modifying drugs, metformin and antiandrogen drugs. Firstly, we incorporated clinical androgen excess features into the primary model to assess their impact on the effect size. Subsequently, we conducted a focused analysis among exposed patients to identify predictors of NAFLD in women with PCOS. In this analysis, we included phenotypic features and antiandrogenic medications as covariates in addition to the previously specified covariates. It

is important to note that these two analyses were constrained due to the availability of documented androgen excess features in only a third of patients.

2.8.8. Data analysis of biochemical androgen excess cohorts

Serum testosterone level concentration was categorised into 1.5 to 1.99, 2 to 2.49, 2.5 to 2.99, 3 to 3.49 and ≥ 3.5 (nmol/L). Serum SHBG concentration was categorised into <20 , 20 to 29.99, 30 to 39.99, 40 to 49.99, 50 to 59.99 and ≥ 60 (nmol/L). Serum SHBG and testosterone categories were created based on clinically meaningful cutoff values and by consensus among the study group to determine if there is a dose-response relationship (Clark et al., 2019; Healy et al., 2014). We provided a Venn diagram (Venn, 1880) to illustrate the overlap between Women with PCOS, women with serum testosterone measurements and women with serum SHBG measurements and another Venn diagram to illustrate the overlap between Women without PCOS, women with serum testosterone measurements and women with serum SHBG measurements.

For each of the sex hormone cohorts, we tabulated the baseline characteristics, including number of participants, age (mean $\pm$ standard deviation), Townsend index categories, BMI categories, smoking status, medical conditions at baseline (diabetes mellitus, hypertension, hypothyroidism, impaired glucose regulation, and importantly, PCOS).

The association between hormonal levels (testosterone and SHBG) and outcomes (NAFLD and dysglacaemia) was examined using Cox regression models that considered potential confounding factors. For each cohort, we presented number of women included, their median age with interquartile range, number of incident outcomes (NAFLD and dysglycaemia), person years, incidence rate (per 10,000 person years), hazard ratio adjusted for important covariates and 95% confidence interval of hazard ratio and p-value. We performed Cox regression to determine the factors associated with outcomes of interest (NAFLD and

dysglycaemia) among each of the cohort of women with available serum testosterone or serum SHBG measurements (as a categorical variable), adjusted for age, Townsend score, BMI, hypothyroidism, and diabetes or impaired glucose regulation. When choosing the reference level for sex hormones in the Cox regression model, we chose the lowest-level category for serum testosterone and the highest-level category for serum SHBG, considering the increased cardiovascular risk associated with higher levels of serum testosterone (Benn et al., 2015) and lower levels of serum SHBG (Sutton-Tyrrell et al., 2005).

We presented dose-response hazard ratio (HR) plot to illustrate the dose response association of serum testosterone and serum SHBG with individual outcomes (NAFLD and dysglycaemia) separately. We used the Y-axis of the dose-response hazard ratio plot to transform the wide range of values with right skew into a symmetrical representation, which is the standard practice to simplify comparisons and improve statistical interpretation and visual clarity (Rothman et al., 2011; Kim et al., 2018).

As a sensitivity analysis, we performed independent Cox regression for each sex hormone cohort to verify if the association between sex hormone levels and outcomes (NAFLD and dysglycaemia) was retained, by excluding women with PCOS/PCO Read codes. The data may be limited by the fact that some of the women with serum testosterone or serum SHBG measurements may have had androgen excess, which is a possible sign of PCOS, but this was not recorded as PCOS.

The determination of diabetes risk among women with serum sex hormone measurements was published as part of a larger study (O'Reilly et al., 2019).

2.9. Role of bias

This retrospective cohort study using routinely collected real world data from general

practices in the UK is susceptible to several bias arising from clinical epidemiological study design (Rivas-Ruiz et al., 2013). We have taken utmost care in minimising these biases when we design the study and tried to verify the possibility of some bias through sensitivity analysis.

2.9.1. Selection bias

Selection bias results when participants selected in a study systematic differ from the population of interest (Catalog of bias, 2023b). Selection bias in real world data or routine data studies would arise from non-referral (Jarada et al., 2023), systematic difference among subgroup in accessing the general practice services (diagnostic access bias) such as ethnicity (Reeves et al., 2022), adherence bias due to systematic difference in adherence to follow-up of a condition or management (Shlipak et al., 2001), admission rate bias or Berkson's bias or referral bias due to systematic difference in getting GP appointment based on the severity of the condition or systematic difference in referral to secondary care by the GP (Luque-Ramírez et al., 2015), and chronological bias due to systematic difference in diagnostic criteria over the time such as for PCOS (Catalog of bias, 2023a).

We minimised or explored selection bias via use of well-defined code list derived through a structured iterative process, clearly defined inclusion and exclusion criteria, use of well-structured index date and exit date, deriving the participants from THIN data base which has proven to represent the UK population, making baseline comparison of characteristics of participants, and by performing sensitivity analysis.

2.9.2. Misclassification bias

Misclassification bias results from systematic difference in the classification of exposure or outcome (Pham et al., 2019). Misclassification bias may arise in various contexts such as

from systematic difference in time point of documentation of important variables such as BMI or not applying ethnicity specific cut-off for BMI (Haas et al., 2023), ascertainment bias or detection bias due to systematic difference in missing documentation such as laboratory findings (Flory et al., 2017; FitzSimmons and Davidson, 1990), and collider bias due to increased healthcare utilisation by both the exposure and outcome in a general practice setting (Weiskopf et al., 2023). We performed sensitivity analysis based on different types of diagnosis of PCOS such as precision of diagnostic codes for PCOS (PCOS only codes and broader category) and incident timeline of diagnosis (cases and prevalent cases).

We minimised or explored the misclassification bias via applying clear definition for exposure and outcomes using a rigorous code list, applying data quality checks such as baseline comparison and sensitivity analysis, record linkage of medical codes, drug codes and AHD codes, sub-group analysis such as BMI categories, temporal considerations such as incident and prevalent cases and use of DExtER (Gokhale et al., 2021) to automate the extraction.

2.9.3. Survivor bias

Survivor bias results from systematically different survival of participants either prior to study start or after the study end. Restricting the analysis to incident cases is one of the options to minimise survivor bias (Lee et al., 2012; Humbert et al., 2010). We performed sensitivity analysis based on the timeline of diagnosis namely incident cases i.e. diagnosis made after the case meeting the eligibility criteria and prevalent cases i.e. diagnosis made before the case meeting the eligibility criteria (Miller et al., 2012).

2.9.4. Immortal time bias

Immortal time is a window period during observation in which the outcome cannot occur

(Suissa, 2008). Cohort studies (Lévesque et al., 2010) including retrospective cohort studies using electronic health records (Tyrer et al., 2022) and studies recording prescriptions (Duchesneau et al., 2022) are susceptible for immortal time bias. We assigned the same index date of the women with PCOS was assigned to the matched control women without PCOS to minimise the immortal time bias (Suissa and Ernst, 2020; Maguire et al., 2009).

2.9.5. Cognitive bias

Cognitive bias (Khadilkar and Khadilkar, 2020) in a general practice setting can act through availability bias (Mamede et al., 2010), base rate neglect bias (Ashinoff et al., 2022), confirmation bias (Pines, 2006), or diagnostic momentum bias and resulting in systematic difference in the diagnosis and management.

2.10. Role of confounders

Confounders in a retrospective cohort study are managed at the designed stage by restriction and matching and at the analysis stage by stratified analysis and multiple regression analysis (McNamee, 2005). We applied restriction by recruiting women to 18 years to 50 years, setting the eligibility date one year after 1) registration in the practice – registration date, 2) introduction of VISION EMR – vision date and accepted mortality recording – AMR date to ensure satisfactory documentation in the practice. We matched controls for practice, age (± 1 year) and BMI (± 2 kg/m²). We performed stratified analysis based on BMI, type of codes used to define PCOS and timeline of diagnosis (incident and prevalent cases). We performed cox regression adjusting for potential confounders such as age, BMI and Townsend deprivation index.

2.11. Ethics approval

Ethics approval was originally sought from the South-East Multi-center Research Ethics

Committee in 2003 for the use of the THIN data for research with exception from the need for informed consent. Further registration and authorization for this project were obtained from the relevant Scientific Review Committee (17THIN026) as per requirement of the ethical approval.

2.12. Reporting guidelines

The Strengthening the reporting of observational studies in epidemiology (STROBE) is the recommended guidelines to report observational studies (Von Elm et al., 2007) including cohort studies (Poorolajal et al., 2011). The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement (Langan et al., 2018) was developed to fill the gap when using the STROBE guidelines in reporting observational studies arising from routinely collected health data (Nicholls et al., 2015). We reported the adherence to RECORD checklist for each of the three papers as supplementary information.

Although all three papers added in this thesis were published prior to NICE real-world evidence framework was published in 2022 by National Institute for Health and Care Excellence UK, most of the recommendations were already incorporated in these three paper such as priori registration of protocol, use of reporting guidelines, open publishing of code list, detailed description of data extraction, discussion on role of bias, assessing robustness of results via sensitivity analysis, stratification, matching, multivariable regression adjustment and use of appropriate reporting guidelines (NICE, 2022).

2.13. References of Chapter 2 : General Methods

Abd Elhafeez, S., D'Arrigo, G., Leonardis, D., et al. (2021) Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. *Oxidative Medicine and Cellular Longevity*, 2021. doi:10.1155/2021/1302811.

Abdel-Maboud, M., Menshawy, A., Hasabo, E.A., et al. (2021) The comparative effectiveness of 55 interventions in obese patients with polycystic ovary syndrome: A network meta-analysis of 101 randomized trials. *PloS one*, 16 (7). doi:10.1371/JOURNAL.PONE.0254412.

Agapova, S.E., Cameo, T., Sopher, A.B., et al. (2014) Diagnosis and challenges of polycystic ovary syndrome in adolescence. *Seminars in Reproductive Medicine*, 32 (3). doi:10.1055/s-0034-1371091.

Ahmadi, M. and Aslani, N. (2018) Capabilities and advantages of cloud computing in the implementation of Electronic Health Record. *Acta Informatica Medica*, 26 (1). doi:10.5455/aim.2018.26.24-28.

Aini, N., Chu, H., Banda, K.J., et al. (2023) Prevalence of sleep-related breathing disorders and associated risk factors among people with dementia: A meta-analysis. *Sleep medicine*, 103: 51–61. doi:10.1016/J.SLEEP.2023.01.020.

Alexander, M., Loomis, A.K., Fairburn-Beech, J., et al. (2018) Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Medicine*, 16 (1). doi:10.1186/s12916-018-1103-x.

Al-Sadawi, M., Theodoropoulos, K., Saeidifard, F., et al. (2022) Sleep Apnea as a Risk Factor for Diastolic Dysfunction: A Systematic Review and Meta-Analysis. *Respiration; international review of thoracic diseases*, 101 (11): 1051–1068. doi:10.1159/000525782.

Alshehri, Z., Subramanian, A., Adderley, N.J., et al. (2022) Risk of incident obstructive sleep apnoea in patients with type 1 diabetes: a population-based retrospective cohort study. *Diabetologia*, 65 (8). doi:10.1007/s00125-022-05714-5.

Alur-Gupta, S., Lee, I., Chemerinski, A., et al. (2021) Racial differences in anxiety, depression, and quality of life in women with polycystic ovary syndrome. *F and S Reports*, 2 (2). doi:10.1016/j.xfre.2021.03.003.

Amiri, M., Ramezani Tehrani, F., Behboudi-Gandevani, S., et al. (2020) Risk of hypertension in women with polycystic ovary syndrome: a systematic review, meta-analysis and meta-regression. *Reproductive biology and endocrinology: RB&E*, 18 (1). doi:10.1186/S12958-020-00576-1.

de Araujo Dantas, A.B., Gonçalves, F.M., Martins, A.A., et al. (2023) Worldwide prevalence and associated risk factors of obstructive sleep apnea: a meta-analysis and meta-regression. *Sleep & breathing = Schlaf & Atmung*. doi:10.1007/S11325-023-02810-7.

Ashinoff, B.K., Buck, J., Woodford, M., et al. (2022) The effects of base rate neglect on sequential belief updating and real-world beliefs. *PLoS Computational Biology*, 18 (12). doi:10.1371/journal.pcbi.1010796.

ASRM. (2021). Diagnosis and treatment of luteal phase deficiency: a committee opinion. *Fertility and Sterility*, 115(6), 1416–1423. <https://doi.org/10.1016/j.fertnstert.2021.02.010>

Austin, L., Sharp, C.A., van der Veer, S.N., et al. (2020) Providing “the bigger picture”: Benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. *Rheumatology (United Kingdom)*, 59 (2). doi:10.1093/rheumatology/kez207.

Awaysheh, A., Wilcke, J., Elvinger, F., et al. (2018) A review of medical terminology standards and structured reporting. *Journal of Veterinary Diagnostic Investigation*. 30 (1).

doi:10.1177/1040638717738276.

Azziz, R. (2021) Adiposity in polycystic ovary syndrome: excess versus dysfunction. *Fertility and Sterility*. 116 (1) pp. 87–88. doi:10.1016/j.fertnstert.2021.04.033.

Azziz, R., Carmina, E., Chen, Z., et al. (2016) Polycystic ovary syndrome. *Nature Reviews Disease Primers*. 2. doi:10.1038/nrdp.2016.57.

Azziz, R., Carmina, E., Dewailly, D., et al. (2006) Position statement: Criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: An androgen excess society guideline. *Journal of Clinical Endocrinology and Metabolism*, 91 (11): 4237–4245. doi:10.1210/jc.2006-0178.

Azziz, R., Kintziger, K., Li, R., et al. (2019) Recommendations for epidemiologic and phenotypic research in polycystic ovary syndrome: an androgen excess and PCOS society resource. *Human reproduction (Oxford, England)*, 34 (11): 2254–2265. doi:10.1093/HUMREP/DEZ185.

Balakrishnan, M. (2021) Women Have a Lower Risk of Non alcoholic Fatty Liver Disease but Higher Risk of Non alcoholic Fatty Liver Disease Fibrosis Than Men: Summary of the Findings of a Systematic Review and Meta-analysis. *Clinical Liver Disease*, 18 (5): 251. doi:10.1002/CLD.1139.

Balen, A., Laven, J.S.E., Tan, S.L., et al. (2003) Ultrasound assessment of the polycystic ovary: International consensus definitions. *Human Reproduction Update*. 9 (6) pp. 505–514. doi:10.1093/humupd/dmg044.

Balen, A. and Michelmore, K. (2002) What is polycystic ovary syndrome? Are national views important? *Human Reproduction*. 17 (9) pp. 2219–2227. doi:10.1093/humrep/17.9.2219.

Ballestri, S., Zona, S., Targher, G., et al. (2016) Nonalcoholic fatty liver disease is associated

with an almost twofold increased risk of incident type 2 diabetes and metabolic syndrome. Evidence from a systematic review and meta-analysis. *Journal of gastroenterology and hepatology*, 31 (5): 936–944. doi:10.1111/JGH.13264.

Bambhroliya, Z., Sandrugu, J., Lowe, M., et al. (2022) Diabetes, Polycystic Ovarian Syndrome, Obstructive Sleep Apnea, and Obesity: A Systematic Review and Important Emerging Themes. *Cureus*, 14 (6). doi:10.7759/cureus.26325.

Baranova, a, Tran, T.P., Birerdinc, a, et al. (2011) Systematic review: association of polycystic ovary syndrome with metabolic syndrome and non-alcoholic fatty liver disease. *Alimentary pharmacology & therapeutics*, 33 (7): 801–814. doi:10.1111/j.1365-2036.2011.04579.x.

Barry, J.A., Azizia, M.M. and Hardiman, P.J. (2014) Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Human reproduction update*, 20 (5): 748–758. doi:10.1093/HUMUPD/DMU012.

Basch, E. and Schrag, D. (2019) The Evolving Uses of “Real-World” Data. *JAMA*. 321 (14). doi:10.1001/jama.2019.4064.

Bazarganipour, F., Taghavi, S.A., Montazeri, A., et al. (2015) The impact of polycystic ovary syndrome on the health-related quality of life: A systematic review and meta-analysis. *Iranian Journal of Reproductive Medicine*, 13 (2).

Behboodi Moghadam, Z., Fereidooni, B., Saffari, M., et al. (2018) Measures of health-related quality of life in pcos women: A systematic review. *International Journal of Women's Health*. 10. doi:10.2147/IJWH.S165794.

Benjafield, A. V., Ayas, N.T., Eastwood, P.R., et al. (2019) Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *The Lancet Respiratory Medicine*, 7 (8). doi:10.1016/S2213-2600(19)30198-5.

Benn, M., Voss, S.S., Holmegard, H.N., et al. (2015) Extreme concentrations of endogenous sex hormones, ischemic heart disease, and death in women. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 35 (2). doi:10.1161/ATVBAHA.114.304821.

Benson, T. (2002) Why general practitioners use computers and hospital doctors do not—Part 1: Incentives. *BMJ*, 325 (7372). doi:10.1136/bmj.325.7372.1086.

Benson, T. (2012) The history of the Read Codes: The inaugural James Read memorial lecture 2011. *Informatics in Primary Care*. 19 (3). doi:10.14236/jhi.v19i3.811.

Berga, S. and Naftolin, F. (2012) Neuroendocrine control of ovulation. *Gynecological endocrinology: the official journal of the International Society of Gynecological Endocrinology*, 28 Suppl 1 (SUPPL.1): 9–13. doi:10.3109/09513590.2012.651929.

Bernhardt, L., Brady, E.M., Freeman, S.C., et al. (2022) Diagnostic accuracy of screening questionnaires for obstructive sleep apnoea in adults in different clinical cohorts: a systematic review and meta-analysis. *Sleep & breathing = Schlaf & Atmung*, 26 (3): 1053–1078. doi:10.1007/S11325-021-02450-9.

Berni, T.R., Morgan, C.L. and Aled Rees, D. (2021) Women with Polycystic Ovary Syndrome Have an Increased Risk of Major Cardiovascular Events: A Population Study. *Journal of Clinical Endocrinology and Metabolism*, 106 (9). doi:10.1210/clinem/dgab392.

Berni, T.R., Morgan, C.L., Berni, E.R., et al. (2018) Polycystic Ovary Syndrome Is Associated With Adverse Mental Health and Neurodevelopmental Outcomes. *The Journal of Clinical Endocrinology & Metabolism*, 103 (6): 2116–2125. doi:10.1210/jc.2017-02667.

Beroukhim, G., Esencan, E. and Seifer, D.B. (2022) Impact of sleep patterns upon female neuroendocrinology and reproductive outcomes: a comprehensive review. *Reproductive biology and endocrinology: RB&E*, 20 (1). doi:10.1186/S12958-022-00889-3.

Bertrand, F. and Maumy-Bertrand, M. (2021) Fitting and Cross-Validating Cox Models to Censored Big Data With Missing Values Using Extensions of Partial Least Squares Regression Models. *Frontiers in Big Data*, 4. doi:10.3389/fdata.2021.684794.

Betensky, R.A. (2015) Measures of Follow-up in Time-to-Event Studies: Why provide them and what should they be? *Clinical trials (London, England)*, 12 (4): 403. doi:10.1177/1740774515586176.

Blak, B.T., Thompson, M., Dattani, H., et al. (2011) Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Informatics in primary care*, 19 (4): 251–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22828580> (Accessed: 1 December 2015).

Bland, J.M. and Altman, D.G. (1995) Multiple significance tests: the Bonferroni method. *BMJ*, 310 (6973). Available at: <http://www.bmj.com/content/310/6973/170.long> (Accessed: 24 May 2017).

Bloom, B.M., Pott, J., Thomas, S., et al. (2021) Usability of electronic health record systems in UK EDs. *Emergency Medicine Journal*, 38 (6). doi:10.1136/emmermed-2020-210401.

Bodmer, M., Becker, C., Meier, C., et al. (2011) Use of metformin and the risk of ovarian cancer: A case-control analysis. *Gynecologic Oncology*, 123 (2): 200–204. doi:10.1016/j.ygyno.2011.06.038.

Boonstra, A., Jonker, T.L., van Offenbeek, M.A.G., et al. (2021) Persisting workarounds in Electronic Health Record System use: types, risks and benefits. *BMC Medical Informatics and Decision Making*, 21 (1). doi:10.1186/s12911-021-01548-0.

Booth, N. (1994) What are the Read Codes? *Health libraries review*. 11 (3) pp. 177–82. doi:10.1046/j.1365-2532.1994.1130177.x.

- Bouhanna, P. (2000) Multifactorial classification of male and female androgenetic alopecia. *Dermatologic Surgery*. 26 (6) pp. 555–561. doi:10.1046/j.1524-4725.2000.00009.x.
- Boursi, B., Finkelman, B., Giantonio, B.J., et al. (2022) A clinical prediction model to assess risk for pancreatic cancer among patients with prediabetes. *European Journal of Gastroenterology and Hepatology*, 34 (1). doi:10.1097/MEG.0000000000002052.
- Bozdag, G., Mumusoglu, S., Zengin, D., et al. (2016) The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction*, 31 (12): 2841–2855. doi:10.1093/humrep/dew218.
- Bradley, S.H., Lawrence, N.R. and Carder, P. (2018) Using primary care data for health research in England – an overview. *Future Healthcare Journal*, 5 (3). doi:10.7861/futurehosp.5-3-207.
- Brankovic, M., Kardys, I., Steyerberg, E.W., et al. (2019) Understanding of interaction (subgroup) analysis in clinical trials. *European Journal of Clinical Investigation*. 49 (8). doi:10.1111/eci.13145.
- Breeze, P., Thomas, C., Thokala, P., et al. (2020) The Impact of Including Costs and Outcomes of Dementia in a Health Economic Model to Evaluate Lifestyle Interventions to Prevent Diabetes and Cardiovascular Disease. *Medical Decision Making*, 40 (7). doi:10.1177/0272989X20946758.
- Brown, P.J.B., O’Neil, M. and Price, C. (1998) Semantic definition of disorders in version 3 of the read codes. *Methods of Information in Medicine*. 37 (4–5). doi:10.1055/s-0038-1634556.
- Bulloch, G., Seth, I., Zhu, Z., et al. (2023) Ocular manifestations of obstructive sleep apnea: a systematic review and meta-analysis. *Graefe’s archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle*

Ophthalmologie. doi:10.1007/S00417-023-06103-3.

Burke, J.F., Sussman, J.B., Kent, D.M., et al. (2015) Three simple rules to ensure reasonably credible subgroup analyses. *BMJ (Online)*, 351. doi:10.1136/bmj.h5651.

Burns, M.L., Mathis, M.R., Vandervest, J., et al. (2020) Classification of current procedural terminology codes from electronic health record data using machine learning. *Anesthesiology*. doi:10.1097/ALN.0000000000003150.

Butler, A.E., Abouseif, A., Dargham, S.R., et al. (2020) Metabolic comparison of polycystic ovarian syndrome and control women in Middle Eastern and UK Caucasian populations. *Scientific Reports*, 10 (1). doi:10.1038/s41598-020-75109-2.

Cahn, A., Shoshan, A., Sagiv, T., et al. (2020) Prediction of progression from pre-diabetes to diabetes: Development and validation of a machine learning model. *Diabetes/Metabolism Research and Reviews*, 36 (2). doi:10.1002/dmrr.3252.

Cai, B., Xu, W., Bortnichak, E., et al. (2012) An algorithm to identify medical practices common to both the General Practice Research Database and The Health Improvement Network database. *Pharmacoepidemiology and drug safety*, 21 (7): 770–774. doi:10.1002/PDS.3277.

Cai, J., Zhang, Y., Wang, Y., et al. (2019) High thyroid stimulating hormone level is associated with hyperandrogenism in euthyroid polycystic ovary syndrome (PCOS) women, independent of age, BMI, and thyroid autoimmunity: A cross-sectional analysis. *Frontiers in Endocrinology*, 10 (APR). doi:10.3389/fendo.2019.00222.

Caleyachetty, R., Thomas, G.N., Toulis, K.A., et al. (2017) Metabolically Healthy Obese and Incident Cardiovascular Disease Events Among 3.5 Million Men and Women. *Journal of the American College of Cardiology*, 70 (12): 1429–1437. doi:10.1016/j.jacc.2017.07.763.

Callahan, A., Polony, V., Posada, J.D., et al. (2021) ACE: The Advanced Cohort Engine for searching longitudinal patient records. *Journal of the American Medical Informatics Association*, 28 (7). doi:10.1093/jamia/ocab027.

Cambras, T., Romero, O., Díez-Noguera, A., et al. (2021) Circadian patterns of patients with type 2 diabetes and obstructive sleep apnea. *Journal of Clinical Medicine*, 10 (2). doi:10.3390/jcm10020244.

Carbonari, D.M., Saine, M.E., Newcomb, C.W., et al. (2015) Use of demographic and pharmacy data to identify patients included within both the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Pharmacoepidemiology and Drug Safety*, 24 (9). doi:10.1002/pds.3844.

Carmina, E., Dreno, B., Lucky, W.A., et al. (2022) Female Adult Acne and Androgen Excess: A Report From the Multidisciplinary Androgen Excess and PCOS Committee. *Journal of the Endocrine Society*, 6 (3). doi:10.1210/jendso/bvac003.

Carr, L.H., Christ, L. and Ferro, D.F. (2023) The Electronic Health Record as a Quality Improvement Tool: Exceptional Potential with Special Considerations. *Clinics in perinatology*, 50 (2): 473–488. doi:10.1016/J.CLP.2023.01.008.

Carstens, D.S., Rodriguez, W. and Wood, M.B. (2014) Task and error analysis balancing benefits over business of electronic medical records. *International Journal of Electronic Healthcare*, 7 (4). doi:10.1504/IJEH.2014.064333.

Catalog of bias (2023a) *Chronological bias - Catalog of Bias*. Available at: <https://catalogofbias.org/biases/chronological-bias/> (Accessed: 25 July 2023).

Catalog of bias (2023b) *Selection bias - Catalog of Bias*. Available at: <https://catalogofbias.org/biases/selection-bias/> (Accessed: 24 July 2023).

Cea Soriano, L., Asiimwe, A. and García Rodríguez, L.A. (2017a) Prescribing of cyproterone acetate/ethinylestradiol in UK general practice: a retrospective descriptive study using The Health Improvement Network. *Contraception*, 95 (3). doi:10.1016/j.contraception.2016.10.003.

Cea Soriano, L., Gaist, D., Soriano-Gabarró, M., et al. (2019a) The importance of validating intracranial bleeding diagnoses in The Health Improvement Network, United Kingdom: Misclassification of onset and its impact on the risk associated with low-dose aspirin therapy. *Pharmacoepidemiology and Drug Safety*, 28 (2). doi:10.1002/pds.4561.

Cea Soriano, L., López-García, E., Schulze-Rath, R., et al. (2017b) Incidence, treatment and recurrence of endometriosis in a UK-based population analysis using data from The Health Improvement Network and the Hospital Episode Statistics database. *European Journal of Contraception and Reproductive Health Care*, 22 (5). doi:10.1080/13625187.2017.1374362.

Cea Soriano, L., Zong, J. and García Rodríguez, L.A. (2019b) Feasibility and validity of The Health Improvement Network database of primary care electronic health records to identify and characterise patients with small cell lung cancer in the United Kingdom. *BMC Cancer*, 19 (1). doi:10.1186/s12885-019-5305-1.

Cea-Soriano, L., García-Rodríguez, L.A., Brodovicz, K.G., et al. (2018) Real world management of pregestational diabetes not achieving glycemic control for many patients in the UK. *Pharmacoepidemiology and Drug Safety*, 27 (8). doi:10.1002/pds.4553.

Chai, T.Y.L., Rajaratnam, R.M., Deng, D., et al. (2021) The prevalence of gestational diabetes mellitus in women diagnosed with non-alcoholic fatty liver disease during pregnancy: A systematic review and meta-analysis. *Journal of diabetes and its complications*, 35 (9). doi:10.1016/J.JDIACOMP.2021.107991.

Chan, J.L., Kar, S., Vanky, E., et al. (2017) Racial and ethnic differences in the prevalence of

metabolic syndrome and its components of metabolic syndrome in women with polycystic ovary syndrome: a regional cross-sectional study. *American Journal of Obstetrics and Gynecology*, 217 (2): 189.e1-189.e8. doi:10.1016/j.ajog.2017.04.007.

Chang, C.W., Chang, C.H., Chuang, H.Y., et al. (2022) What is the association between secondhand smoke (SHS) and possible obstructive sleep apnea: a meta-analysis. *Environmental health : a global access science source*, 21 (1). doi:10.1186/S12940-022-00868-6.

Chao, W.C., Hu, H., Ung, C.O.L., et al. (2013) Benefits and challenges of electronic health record system on stakeholders: A qualitative study of outpatient physicians. *Journal of Medical Systems*, 37 (4). doi:10.1007/s10916-013-9960-5.

Chen, H., Wang, J., Li, Z., et al. (2019) Consumption of Sugar-Sweetened Beverages Has a Dose-Dependent Effect on the Risk of Non-Alcoholic Fatty Liver Disease: An Updated Systematic Review and Dose-Response Meta-Analysis. *International journal of environmental research and public health*, 16 (12). doi:10.3390/IJERPH16122192.

Chen, H., Zheng, Z., Chen, R., et al. (2022a) A meta-analysis of the diagnostic value of NoSAS in patients with sleep apnea syndrome. *Sleep & breathing = Schlaf & Atmung*, 26 (2): 519–531. doi:10.1007/S11325-021-02410-3.

Chen, L., Zou, S. and Wang, J. (2022b) Association of Obstructive Sleep Apnea Syndrome (OSA/OSAHS) with Coronary Atherosclerosis Risk: Systematic Review and Meta-Analysis. *Computational and mathematical methods in medicine*, 2022. doi:10.1155/2022/8905736.

Chen, T., Yu, Y., Jia, F., et al. (2022c) The relationship between polycystic ovary syndrome and insulin resistance from 1983 to 2022: A bibliometric analysis. *Frontiers in public health*, 10. doi:10.3389/FPUBH.2022.960965.

Chen, X., He, S. and Wang, D. (2021) Effects of metformin on body weight in polycystic ovary syndrome patients: model-based meta-analysis. *Expert Review of Clinical Pharmacology*, 14 (1). doi:10.1080/17512433.2021.1863788.

Chen, Y., Bai, W., Mao, D., et al. (2023) The relationship between non-alcoholic fatty liver disease and incidence of chronic kidney disease for diabetic and non-diabetic subjects: A meta-analysis. *Advances in clinical and experimental medicine: official organ Wroclaw Medical University*, 32 (4): 0–0. doi:10.17219/ACEM/155017.

Chen, Z. and Cai, Z. (2022) Effects of oral contraceptives plus orlistat in patients with polycystic ovary syndrome and overweight/obesity: A meta-analysis. *The journal of obstetrics and gynaecology research*, 48 (6): 1399–1408. doi:10.1111/JOG.15211.

Cheong, A.J.Y., Ho, O.T.W., Wang, S.K.X., et al. (2023) Association between obstructive sleep apnea and floppy eyelid syndrome: A systematic review and metaanalysis. *Survey of ophthalmology*, 68 (2): 257–264. doi:10.1016/J.SURVOPHTHAL.2022.11.006.

Cheong, A.J.Y., Tan, B.K.J., Teo, Y.H., et al. (2022) Obstructive Sleep Apnea and Lung Cancer: A Systematic Review and Meta-Analysis. *Annals of the American Thoracic Society*. 19 (3). doi:10.1513/AnnalsATS.202108-960OC.

Cherskov, A., Pohl, A., Allison, C., et al. (2018) Polycystic ovary syndrome and autism: A test of the prenatal sex steroid theory. *Translational Psychiatry*, 8 (1). doi:10.1038/s41398-018-0186-7.

Chiaffarino, F., Cipriani, S., Dalmartello, M., et al. (2022) Prevalence of polycystic ovary syndrome in European countries and USA: A systematic review and meta-analysis. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 279. doi:10.1016/j.ejogrb.2022.10.020.

Chisholm, J. (1990) The Read clinical classification. *British Medical Journal*. 300 (6732).

doi:10.1136/bmj.300.6732.1092.

Chiu, H.Y., Chen, P.Y., Chuang, L.P., et al. (2017) Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: A bivariate meta-analysis. *Sleep Medicine Reviews*. 36 pp. 57–70. doi:10.1016/j.smr.2016.10.004.

Cho, J., Lee, I., Park, D.H., et al. (2021) Relationships between socioeconomic status, handgrip strength, and non-alcoholic fatty liver disease in middle-aged adults. *International Journal of Environmental Research and Public Health*, 18 (4). doi:10.3390/ijerph18041892.

Ciardullo, S., Grassi, G., Mancina, G., et al. (2022) Nonalcoholic fatty liver disease and risk of incident hypertension: a systematic review and meta-analysis. *European journal of gastroenterology & hepatology*, 34 (4): 365–371. doi:10.1097/MEG.0000000000002299.

Clark, R. V., Wald, J.A., Swerdloff, R.S., et al. (2019) Large divergence in testosterone concentrations between men and women: Frame of reference for elite athletes in sex-specific competition in sports, a narrative review. *Clinical Endocrinology*. 90 (1). doi:10.1111/cen.13840.

Cleves, M.A., Gould, W. and Marchenko, Y. V. (2016) *An Introduction to Survival Analysis Using Stata*. STATA.

Colonetti, L., Grande, A.J., Toreti, I.R., et al. (2022) Green tea promotes weight loss in women with polycystic ovary syndrome: Systematic review and meta-analysis. *Nutrition Research*. 104. doi:10.1016/j.nutres.2022.03.009.

Conway, G., Dewailly, D., Diamanti-Kandarakis, E., et al. (2014) The polycystic ovary syndrome: a position statement from the European Society of Endocrinology. *European journal of endocrinology / European Federation of Endocrine Societies*, 171 (4): P1-29. doi:10.1530/EJE-14-0253.

Cooney, L.G., Lee, I., Sammel, M.D., et al. (2017) High prevalence of moderate and severe depressive and anxiety symptoms in polycystic ovary syndrome: a systematic review and meta-analysis. *Human reproduction (Oxford, England)*, 32 (5): 1075–1091. doi:10.1093/HUMREP/DEX044.

CPRD (2022) *CPRD GOLD Data Specification*.

Cronin, F.M., Hurley, S.M., Buckley, T., et al. (2022) Mediators of socioeconomic differences in overweight and obesity among youth in Ireland and the UK (2011–2021): a systematic review. *BMC Public Health*, 22 (1). doi:10.1186/s12889-022-14004-z.

Cusick, M.M., Sholle, E.T., Davila, M.A., et al. (2020) A Method to Improve Availability and Quality of Patient Race Data in an Electronic Health Record System. *Applied Clinical Informatics*, 11 (5). doi:10.1055/s-0040-1718756.

Cussen, L., McDonnell, T., Bennett, G., et al. (2022) Approach to androgen excess in women: Clinical and biochemical insights. *Clinical Endocrinology*. 97 (2). doi:10.1111/cen.14710.

Dafoulas, G.E., Toulis, K.A., Mccorry, D., et al. (2017) Type 1 diabetes mellitus and risk of incident epilepsy: a population-based, open-cohort study. *Diabetologia*, 60 (2): 258–261. doi:10.1007/s00125-016-4142-x.

Danaei, G., García Rodríguez, L.A., Cantero, O.F., et al. (2018) Electronic medical records can be used to emulate target trials of sustained treatment strategies. *Journal of Clinical Epidemiology*, 96. doi:10.1016/j.jclinepi.2017.11.021.

Davé, S. and Petersen, I. (2009) Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiology and Drug Safety*, 18 (8): 704–707. doi:10.1002/pds.1770.

Delgado-Rodríguez, M. and Llorca, J. (2004) Bias. *Journal of epidemiology and community health*, 58 (8): 635–641. doi:10.1136/jech.2003.008466.

Denburg, M.R., Haynes, K., Shults, J., et al. (2011) Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease. *Pharmacoepidemiology and drug safety*, 20 (11): 1138–49. doi:10.1002/pds.2203.

Deo, S.V., Deo, V. and Sundaram, V. (2021) Survival analysis—part 2: Cox proportional hazards model. *Indian Journal of Thoracic and Cardiovascular Surgery*, 37 (2). doi:10.1007/s12055-020-01108-7.

Derks, I.P.M., Koster, A., Schram, M.T., et al. (2017) The association of early life socioeconomic conditions with prediabetes and type 2 diabetes: results from the Maastricht study. *International Journal for Equity in Health*, 16 (1). doi:10.1186/s12939-017-0553-7.

Deshmukh, H., Shah, N., Papageorgiou, M., et al. (2022) Genetic risk for the polycystic ovary syndrome, bone mineral density and fractures in women and men: A UK Biobank Mendelian randomisation study. *Bone*, 155. doi:10.1016/j.bone.2021.116285.

Deswal, R., Yadav, A. and Dang, A.S. (2018) Sex hormone binding globulin - an important biomarker for predicting PCOS risk: A systematic review and meta-analysis. *Systems Biology in Reproductive Medicine*, 64 (1): 12–24. doi:10.1080/19396368.2017.1410591.

Dijkman, B., Kooistra, B., Bhandari, M., et al. (2009) How to work with a subgroup analysis. *Canadian Journal of Surgery*, 52 (6).

Ding, T., Baio, G., Hardiman, P.J., et al. (2016) Diagnosis and management of polycystic ovary syndrome in the UK: a retrospective cohort study. *BMJ Open*, 6: 5–8. doi:10.1136/bmjopen-2016-012461.

Ding, T., Hardiman, P.J., Petersen, I., et al. (2018a) Incidence and prevalence of diabetes and

cost of illness analysis of polycystic ovary syndrome: a Bayesian modelling study. *Human Reproduction*, 33 (7): 1299–1306. doi:10.1093/humrep/dey093.

Ding, X., Yang, L., Wang, J., et al. (2018b) Subclinical Hypothyroidism in Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis. *Frontiers in Endocrinology*. 9. doi:10.3389/fendo.2018.00700.

Duchesneau, E.D., Jackson, B.E., Webster-Clark, M., et al. (2022) The Timing, the Treatment, the Question: Comparison of Epidemiologic Approaches to Minimize Immortal Time Bias in Real-World Data Using a Surgical Oncology Example. *Cancer Epidemiology Biomarkers and Prevention*, 31 (11). doi:10.1158/1055-9965.EPI-22-0495.

Dunn, A.J. (2021) Environmental Influences and Polycystic Ovarian Syndrome. *Clinical obstetrics and gynecology*, 64 (1): 33–38. doi:10.1097/GRF.0000000000000562.

Dupont, W.D. (2021) Review of Michael N. Mitchell's Data Management Using Stata: A Practical Handbook, Second Edition. *Stata Journal*, 21 (3). doi:10.1177/1536867X211045581.

Eberhardt, J., Bright, C.J., Gildea, C., et al. (2022) Impact of the third national 'Be Clear on Cancer' Breast Cancer in Women over 70 Campaign on general practitioner attendance and referral, diagnosis rates and prevalence awareness. *European Journal of Cancer Care*, 31 (3). doi:10.1111/ecc.13583.

Echouffo-Tcheugui, J.B., Perreault, L., Ji, L., et al. (2023) Diagnosis and Management of Prediabetes: A Review. *JAMA*, 329 (14): 1206–1216. doi:10.1001/jama.2023.4063.

Einarson, T.R., Acs, A., Ludwig, C., et al. (2018) Economic Burden of Cardiovascular Disease in Type 2 Diabetes: A Systematic Review. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 21 (7): 881–890. doi:10.1016/J.JVAL.2017.12.019.

Elhassan, Y.S., Idkowiak, J., Smith, K., et al. (2018) Causes, patterns, and severity of androgen excess in 1205 consecutively recruited women. *Journal of Clinical Endocrinology and Metabolism*, 103 (3): 1214–1223. doi:10.1210/jc.2017-02426.

Ellis, N.T. and Kidd, M.R. (2000) General practice computerisation: Lessons from the United Kingdom. *Medical Journal of Australia*. 172 (1). doi:10.5694/j.1326-5377.2000.tb123873.x.

Von Elm, E., Altman, D.G., Egger, M., et al. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *PLOS Medicine*, 4 (10): e296. doi:10.1371/JOURNAL.PMED.0040296.

von Elm, E., Altman, D.G., Egger, M., et al. (2007) The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*, 370 (9596): 1453–1457. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25046131>.

Emanuel, R.H.K., Roberts, J., Docherty, P.D., et al. (2022) A review of the hormones involved in the endocrine dysfunctions of polycystic ovary syndrome and their interactions. *Frontiers in Endocrinology*, 13. doi:10.3389/fendo.2022.1017468.

Etindele Soso, F. and Matos, E. (2021) Socioeconomic disparities in obstructive sleep apnea: a systematic review of empirical research. *Sleep and Breathing*. 25 (4). doi:10.1007/s11325-020-02274-z.

Fadlalmola, H.A., Elhusein, A.M., Al-Sayaghi, K.M., et al. (2023) Efficacy of resveratrol in women with polycystic ovary syndrome: a systematic review and meta-analysis of randomized clinical trials. *Pan African Medical Journal*, 44. doi:10.11604/pamj.2023.44.134.32404.

Farrokhyar, F., Skorzewski, P., Phillips, M.R., et al. (2022) When to believe a subgroup

analysis: revisiting the 11 criteria. *Eye (Basingstoke)*. 36 (11). doi:10.1038/s41433-022-01948-0.

Feltner, C., Wallace, I.F., Aymes, S., et al. (2022) Screening for Obstructive Sleep Apnea in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 328 (19) pp. 1951–1971. doi:10.1001/jama.2022.18357.

FitzSimmons, J. and Davidson, S. (1990) Ascertainment bias in diabetic outcome data. *American Journal of Obstetrics and Gynecology*. 163 (2). doi:10.1016/0002-9378(90)91235-5.

Fletcher, J. (2007) Subgroup analyses: How to avoid being misled. *British Medical Journal*, 335 (7610). doi:10.1136/bmj.39265.596262.AD.

Flory, J.H., Roy, J., Gagne, J.J., et al. (2017) Missing laboratory results data in electronic health databases: Implications for monitoring diabetes risk. *Journal of Comparative Effectiveness Research*, 6 (1). doi:10.2217/cer-2016-0033.

Forslund, M., Landin-Wilhelmsen, K., Brännström, M., et al. (2023) No difference in morbidity between perimenopausal women with PCOS with and without previous wedge resection. *European journal of obstetrics, gynecology, and reproductive biology*, 285: 74–78. doi:10.1016/J.EJOGRB.2023.04.008.

Forslund, M., Schmidt, J., Brännström, M., et al. (2022) Morbidity and mortality in PCOS: A prospective follow-up up to a mean age above 80 years. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 271. doi:10.1016/j.ejogrb.2022.02.020.

Fox, R. and Hull, M. (1993) Ultrasound Diagnosis of Polycystic Ovaries. *Annals of the New York Academy of Sciences*, 687 (1): 217–223. doi:10.1111/j.1749-6632.1993.tb43868.x.

Freitas De-Medeiros, S., Yamamoto, M.M.W., Souto De-Medeiros, M.A., et al. (2017)

Should subclinical hypothyroidism be an exclusion criterion for the diagnosis of polycystic ovary syndrome? *Journal of Reproduction and Infertility*, 18 (2).

Fu, C.E., Ng, C.H., Yong, J.N., et al. (2023a) A Meta-analysis on Associated Risk of Mortality in Nonalcoholic Fatty Liver Disease. *Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists*, 29 (1): 33–39. doi:10.1016/J.EPRAC.2022.10.007.

Fu, L., Xie, N., Qu, F., et al. (2023b) The Association Between Polycystic Ovary Syndrome and Metabolic Syndrome in Adolescents: a Systematic Review and Meta-analysis. *Reproductive Sciences*. 30 (1). doi:10.1007/s43032-022-00864-8.

Gallagher, A.M., Dedman, D., Padmanabhan, S., et al. (2019) The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiology and Drug Safety*, 28 (5). doi:10.1002/pds.4747.

Gao, Y., Liu, H., Qiao, L., et al. (2023) Study of Burden in Polycystic Ovary Syndrome at Global, Regional, and National Levels from 1990 to 2019. *Healthcare (Switzerland)*, 11 (4). doi:10.3390/healthcare11040562.

Garcia-Beltran, C., Bassols, J., Carreras-Badosa, G., et al. (2023) Raised Thyroid Stimulating Hormone in Girls with Polycystic Ovary Syndrome: Effects of Randomized Interventions. *Hormone Research in Paediatrics*. doi:10.1159/000529183.

Gardner, M.P., Wang, J., Hazlehurst, J.M., et al. (2023) Risk of progression from pre-diabetes to type 2 diabetes in a large UK adult cohort. *Diabetic Medicine*, 40 (3). doi:10.1111/dme.14996.

Ge, J.J., Wang, D.J., Song, W., et al. (2022) The effectiveness and safety of liraglutide in treating overweight/obese patients with polycystic ovary syndrome: a meta-analysis. *Journal*

of endocrinological investigation, 45 (2): 261–273. doi:10.1007/S40618-021-01666-6.

George, A., Stead, T.S. and Ganti, L. (2020) What's the Risk: Differentiating Risk Ratios, Odds Ratios, and Hazard Ratios? *Cureus*. doi:10.7759/cureus.10047.

Glintborg, D., Mumm, H., Hougaard, D., et al. (2010) Ethnic differences in Rotterdam criteria and metabolic risk factors in a multiethnic group of women with PCOS studied in Denmark. *Clinical endocrinology*, 73 (6): 732–8. doi:10.1111/j.1365-2265.2010.03873.x.

Glintborg, D., Rubin, K.H., Nybo, M., et al. (2019) Increased risk of thyroid disease in danish women with polycystic ovary syndrome: A cohort study. *Endocrine Connections*, 8 (10). doi:10.1530/EC-19-0377.

Gokhale, K.M., Chandan, J.S., Toulis, K., et al. (2021) Data extraction for epidemiological research (DExtER): a novel tool for automated clinical epidemiology studies. *European Journal of Epidemiology*, 36 (2): 165–178. doi:10.1007/S10654-020-00677-6/TABLES/6.

Goldstein, S.R. (1990) Incorporating endovaginal ultrasonography into the overall gynecologic examination. *American journal of obstetrics and gynecology*, 162 (3): 625–632. doi:10.1016/0002-9378(90)90971-9.

Greengross, P., Grant, K. and Collini, E. (1999) *The history and development of the UK National Health Service 1948 - 1999*. Available at: www.healthsystemsrc.org.

Guay-Gagnon, M., Vat, S., Forget, M.F., et al. (2022) Sleep apnea and the risk of dementia: A systematic review and meta-analysis. *Journal of sleep research*, 31 (5). doi:10.1111/JSR.13589.

Guest, J.F., Panca, M., Sladkevicius, E., et al. (2014) Clinical outcomes and cost-effectiveness of continuous positive airway pressure to manage obstructive sleep apnea in patients with type 2 diabetes in the U.K. *Diabetes Care*, 37 (5). doi:10.2337/dc13-2539.

Gulliford, M.C., Sun, X., Anjuman, T., et al. (2020) Comparison of antibiotic prescribing records in two UK primary care electronic health record systems: Cohort study using CPRD GOLD and CPRD Aurum databases. *BMJ Open*, 10 (6). doi:10.1136/bmjopen-2020-038767.

Guo, L., Gordon, N.P., Chandra, M., et al. (2021) The risks of polycystic ovary syndrome and diabetes vary by ethnic subgroup among young asian women. *Diabetes Care*. 44 (6). doi:10.2337/dc21-0373.

Gutiérrez-Tobal, G.C., Álvarez, D., Kheirandish-Gozal, L., et al. (2022) Reliability of machine learning to diagnose pediatric obstructive sleep apnea: Systematic review and meta-analysis. *Pediatric Pulmonology*, 57 (8): 1931–1943. doi:10.1002/ppul.25423.

Haas, C.L., Varbo, A., Laursen, P.N., et al. (2023) Association between body mass index, weight loss and the chance of pregnancy in women with polycystic ovary syndrome and overweight or obesity: a retrospective cohort study in the UK. *Human reproduction (Oxford, England)*, 38 (3). doi:10.1093/humrep/deac267.

Hacivelioglu, S., Gungor, A.N.C., Gencer, M., et al. (2013) Acne severity and the Global Acne Grading System in polycystic ovary syndrome. *International Journal of Gynecology & Obstetrics*, 123 (1): 33–36. doi:10.1016/j.ijgo.2013.05.005.

Hadjiconstantinou, M., Mani, H., Patel, N., et al. (2017) Understanding and supporting women with polycystic ovary syndrome: A qualitative study in an ethnically diverse UK sample. *Endocrine Connections*, 6 (5). doi:10.1530/EC-17-0053.

Hanson, M.S. (2005) Stata tip 26: Maximizing compatibility between macintosh and windows. *Stata Journal*. 5 (4). doi:10.1177/1536867x0500500411.

Heald, A.H., Livingston, M., Holland, D., et al. (2018) Polycystic ovarian syndrome: Assessment of approaches to diagnosis and cardiometabolic monitoring in UK primary care. *International Journal of Clinical Practice*, 72 (1). doi:10.1111/ijcp.13046.

Healy, M.L., Gibney, J., Pentecost, C., et al. (2014) Endocrine profiles in 693 elite athletes in the postcompetition setting. *Clinical Endocrinology*, 81 (2). doi:10.1111/cen.12445.

Helvaci, N., Karabulut, E., Demir, A.U., et al. (2017) Polycystic ovary syndrome and the risk of obstructive sleep apnea: A meta-analysis and review of the literature. *Endocrine Connections*, 6 (7): 437–445. doi:10.1530/EC-17-0129.

Herrett, E., Gallagher, A.M., Bhaskaran, K., et al. (2015) Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International journal of epidemiology*, 44 (3): 827–836. doi:10.1093/IJE/DYV098.

Hippisley-Cox, J. and Coupland, C. (2017) Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. *BMJ (Clinical research ed.)*, 359. doi:10.1136/bmj.j5019.

Hippisley-Cox, J., Stables, D. and Pringle, M. (2004) QRESEARCH: A new general practice database for research. *Informatics in Primary Care*. 12 (1). doi:10.14236/jhi.v12i1.108.

Hobbs, F.D.R., Bankhead, C., Mukhtar, T., et al. (2016) Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *The Lancet*, 387 (10035). doi:10.1016/S0140-6736(16)00620-6.

Horsfall, L., Walters, K. and Petersen, I. (2013) Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiology and drug safety*, 22 (1): 64–9. doi:10.1002/pds.3368.

Hou, H., Zhao, Y., Yu, W., et al. (2018) Association of obstructive sleep apnea with hypertension: A systematic review and meta-analysis. *Journal of Global Health*, 8 (1). doi:10.7189/jogh.08.010405.

House of Parliament (2016) *Government Proposals for Electronic Health Records*. Available

at: www.parliament.uk/post.

Humbert, M., Sitbon, O., Yaïci, A., et al. (2010) Survival in incident and prevalent cohorts of patients with pulmonary arterial hypertension. *European Respiratory Journal*, 36 (3). doi:10.1183/09031936.00057010.

Ichikawa, M., Akiyama, T., Tsujimoto, Y., et al. (2022) Diagnostic accuracy of home sleep apnea testing using peripheral arterial tonometry for sleep apnea: A systematic review and meta-analysis. *Journal of sleep research*, 31 (6). doi:10.1111/JSR.13682.

IMS Health (2015) *THIN Data Guide for Researchers*.

Ingram, J.R., Jenkins-Jones, S., Knipe, D.W., et al. (2018) Population-based Clinical Practice Research Datalink study using algorithm modelling to identify the true burden of hidradenitis suppurativa. *British Journal of Dermatology*, 178 (4). doi:10.1111/bjd.16101.

Isakari, M., Sanchez, A., Conic, R., et al. (2023) Benefits and Challenges of Transitioning Occupational Health to an Enterprise Electronic Health Record. *Journal of Occupational & Environmental Medicine*, 65 (7): 615–620. doi:10.1097/JOM.0000000000002864.

Ismail, S.U., Asamane, E.A., Osei-Kwasi, H.A., et al. (2022) Socioeconomic Determinants of Cardiovascular Diseases, Obesity, and Diabetes among Migrants in the United Kingdom: A Systematic Review. *International journal of environmental research and public health*, 19 (5). doi:10.3390/IJERPH19053070.

Jacobs, M., Boersma, L.J., Swart, R., et al. (2019) Electronic Health Record implementation in a large academic radiotherapy department: Temporarily disruptions but long-term benefits. *International Journal of Medical Informatics*, 129. doi:10.1016/j.ijmedinf.2019.07.008.

Jarada, T.N., O’Sullivan, D.E., Brenner, D.R., et al. (2023) Selection Bias in Real-World Data Studies Used to Support Health Technology Assessments: A Case Study in Metastatic

Cancer. *Current Oncology*, 30 (2). doi:10.3390/curroncol30020151.

Joham, A.E., Piltonen, T., Lujan, M.E., et al. (2022) Challenges in diagnosis and understanding of natural history of polycystic ovary syndrome. *Clinical Endocrinology*. 97 (2). doi:10.1111/cen.14757.

Jones, M. (2022) Benefits of using SNOMED CT in the UK National Health Service (NHS). *Roadmap to Successful Digital Health Ecosystems: A Global Perspective*, pp. 489–497. doi:10.1016/B978-0-12-823413-6.00023-9.

Kaczmarek, C., Haller, D.M. and Yaron, M. (2016) Health-Related Quality of Life in Adolescents and Young Adults with Polycystic Ovary Syndrome: A Systematic Review. *Journal of Pediatric and Adolescent Gynecology*. 29 (6). doi:10.1016/j.jpog.2016.05.006.

Kahal, H., Kyrou, I., Uthman, O.A., et al. (2020) The prevalence of obstructive sleep apnoea in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Sleep and Breathing*, 24 (1): 339–350. doi:10.1007/s11325-019-01835-1.

Kazemi, M., Kim, J.Y., Parry, S.A., et al. (2021) Disparities in cardio metabolic risk between Black and White women with polycystic ovary syndrome: a systematic review and meta-analysis. *American Journal of Obstetrics and Gynecology*. 224 (5). doi:10.1016/j.ajog.2020.12.019.

van Keizerswaard, J., Dietz de Loos, A.L.P., Louwers, Y. V., et al. (2022) Changes in individual polycystic ovary syndrome phenotypical characteristics over time: a long-term follow-up study. *Fertility and sterility*, 117 (5): 1059–1066. doi:10.1016/J.FERTNSTERT.2022.01.014.

Kelly dos Santos, I., Ashe, M.C., Cobucci, R.N., et al. (2020) The effect of exercise as an intervention for women with polycystic ovary syndrome: A systematic review and meta-analysis. *Medicine*, 99 (16): E19644. doi:10.1097/MD.00000000000019644.

Kempegowda, P., Melson, E., Manolopoulos, K.N., et al. (2020) Implicating androgen excess in propagating metabolic disease in polycystic ovary syndrome. *Therapeutic advances in endocrinology and metabolism*, 11. doi:10.1177/2042018820934319.

Keramati, M., Musazadeh, V., Malekhamdi, M., et al. (2022) Cinnamon, an effective anti-obesity agent: Evidence from an umbrella meta-analysis. *Journal of Food Biochemistry*. 46 (8). doi:10.1111/jfbc.14166.

Khadilkar, S. V. and Khadilkar, S.S. (2020) Bias in Clinical Practice. *Journal of Obstetrics and Gynecology of India*. 70 (1). doi:10.1007/s13224-019-01304-5.

Khalil, C., Zarabi, S., Kirkham, K., et al. (2023) Validity of non-contact methods for diagnosis of Obstructive Sleep Apnea: a systematic review and meta-analysis. *Journal of clinical anesthesia*, 87. doi:10.1016/J.JCLINANE.2023.111087.

Khan, M.J., Ullah, A. and Basit, S. (2019) Genetic Basis of Polycystic Ovary Syndrome (PCOS): Current Perspectives. *The application of clinical genetics*, 12: 249–260. doi:10.2147/TACG.S200341.

Khan, N.F., Harrison, S.E. and Rose, P.W. (2010) Validity of diagnostic coding within the General Practice Research Database: A systematic review. *British Journal of General Practice*. 60 (572). doi:10.3399/bjgp10X483562.

Kim, A.S. and Wild, R.A. (2022) The natural history of polycystic ovary syndrome: commentary on a longitudinal study evaluating changes in phenotype with age. *Fertility and Sterility*, 117 (5): 1067–1068. doi:10.1016/J.FERTNSTERT.2022.03.003.

Kim, D.H., Kim, S.W. and Hwang, S.H. (2022) Diagnostic value of smartphone in obstructive sleep apnea syndrome: A systematic review and meta-analysis. *PloS one*, 17 (5). doi:10.1371/JOURNAL.PONE.0268585.

Kim, H.J., Adams, J.M., Gudmundsson, J.A., et al. (2017) Polycystic ovary morphology: age-based ultrasound criteria. *Fertility and Sterility*, 108 (3): 548–553. doi:10.1016/j.fertnstert.2017.07.005.

Kim, J., Kaufman, J.S. and Bang, H. (2018) Graphing Ratio Measures on Forest Plot. *Journal of the American College of Cardiology*. 71 (5). doi:10.1016/j.jacc.2017.10.098.

King, J., Patel, V., Jamoom, E.W., et al. (2014) Clinical benefits of electronic health record use: National findings. *Health Services Research*, 49 (1 PART 2). doi:10.1111/1475-6773.12135.

Koletsis, D. and Pandis, N. (2017) Survival analysis, part 3: Cox regression. *American Journal of Orthodontics and Dentofacial Orthopedics*. 152 (5). doi:10.1016/j.ajodo.2017.07.009.

Kontopantelis, E., Stevens, R.J., Helms, P.J., et al. (2018) Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: A cross-sectional population study. *BMJ Open*, 8 (2). doi:10.1136/bmjopen-2017-020738.

Kumarendran, B., Sumilo, D., O'Reilly, M.W., et al. (2019) Increased risk of obstructive sleep apnoea in women with polycystic ovary syndrome: A population-based cohort study. *European Journal of Endocrinology*, 180 (4): 265–272. doi:10.1530/EJE-18-0693.

Lambert, P.C. (2017) The estimation and modeling of cause-specific cumulative incidence functions using time-dependent weights. *Stata Journal*, 17 (1). doi:10.1177/1536867x1701700110.

Langan, S.M., Schmidt, S.A., Wing, K., et al. (2018) The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *The BMJ*. 363. doi:10.1136/bmj.k3532.

Lavrentaki, A., Thomas, T., Subramanian, A., et al. (2019) Increased risk of non-alcoholic

fatty liver disease in women with gestational diabetes mellitus: A population-based cohort study, systematic review and meta-analysis. *Journal of diabetes and its complications*, 33 (10). doi:10.1016/J.JDIACOMP.2019.06.006.

Le, M.H., Le, D.M., Baez, T.C., et al. (2023) Global incidence of non-alcoholic fatty liver disease: a systematic review and meta-analysis of 63 studies and 1,201,807 persons. *Journal of hepatology*. doi:10.1016/J.JHEP.2023.03.040.

Lee, B., Jung, E.A., Yoo, J.J., et al. (2020) Prevalence, incidence and risk factors of tamoxifen-related non-alcoholic fatty liver disease: A systematic review and meta-analysis. *Liver international: official journal of the International Association for the Study of the Liver*, 40 (6): 1344–1355. doi:10.1111/LIV.14434.

Lee, W.T.N., Ling, Y., Sheares, K.K., et al. (2012) Predicting survival in pulmonary arterial hypertension in the UK. *European Respiratory Journal*, 40 (3). doi:10.1183/09031936.00196611.

Legro, R.S. (2003) Diagnostic Criteria in Polycystic Ovary Syndrome. *Seminars in Reproductive Medicine*, 21 (3): 267–275. doi:10.1055/s-2003-43304.

Legro, R.S., Arslanian, S.A., Ehrmann, D.A., et al. (2013) Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *The Journal of clinical endocrinology and metabolism*, 98 (12): 4565–92. doi:10.1210/jc.2013-2350.

Legro, R.S., Schlaff, W.D., Diamond, M.P., et al. (2010) Total testosterone assays in women with polycystic ovary syndrome: Precision and correlation with hirsutism. *Journal of Clinical Endocrinology and Metabolism*, 95 (12): 5305–5313. doi:10.1210/jc.2010-1123.

Lévesque, L.E., Hanley, J.A., Kezouh, A., et al. (2010) Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes. *BMJ (Online)*,

340 (7752): 907–911. doi:10.1136/bmj.b5087.

Lewis, J.D., Schinnar, R., Bilker, W.B., et al. (2007) Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiology and drug safety*, 16 (4): 393–401. doi:10.1002/pds.1335.

Li, W., Wen, W., Xie, D., et al. (2022) Association between non-alcoholic fatty liver disease and risk of incident heart failure: a meta-analysis of observational studies. *Therapeutic advances in chronic disease*, 13. doi:10.1177/20406223221119626.

Li, Y., Li, Y., Yu Ng, E.H., et al. (2011) Polycystic ovary syndrome is associated with negatively variable impacts on domains of health-related quality of life: Evidence from a meta-analysis. *Fertility and Sterility*, 96 (2). doi:10.1016/j.fertnstert.2011.05.072.

Liao, D., Zhong, C., Li, C., et al. (2018) Meta-analysis of the effects of probiotic supplementation on glycemia, lipidic profiles, weight loss and C-reactive protein in women with polycystic ovarian syndrome. *Minerva Medica*. 109 (6). doi:10.23736/S0026-4806.18.05728-2.

Lim, S.S., Kakoly, N.S., Tan, J.W.J., et al. (2019) Metabolic syndrome in polycystic ovary syndrome: a systematic review, meta-analysis and meta-regression. *Obesity Reviews*. 20 (2). doi:10.1111/obr.12762.

Lim, S.S., Norman, R.J., Davies, M.J., et al. (2013) The effect of obesity on polycystic ovary syndrome: a systematic review and meta-analysis. *Obesity reviews : an official journal of the International Association for the Study of Obesity*, 14 (2): 95–109. doi:10.1111/j.1467-789X.2012.01053.x.

Lis, Y. and Mann, R.D. (1995) The VAMP research multi-purpose database in the U.K. *Journal of Clinical Epidemiology*, 48 (3). doi:10.1016/0895-4356(94)00137-F.

Liu, J., Wu, Q., Hao, Y., et al. (2021) Measuring the global disease burden of polycystic ovary syndrome in 194 countries: Global Burden of Disease Study 2017. *Human Reproduction*, 36 (4): 1108–1119. doi:10.1093/humrep/deaa371.

Livadas, S., Macut, D., Bothou, C., et al. (2020) Insulin resistance, androgens, and lipids are gradually improved in an age-dependent manner in lean women with polycystic ovary syndrome: insights from a large Caucasian cohort. *Hormones*, 19 (4). doi:10.1007/s42000-020-00211-z.

Lizneva, D., Gavrilova-Jordan, L., Walker, W., et al. (2016a) Androgen excess: Investigations and management. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 37: 98–118. doi:10.1016/j.bpobgyn.2016.05.003.

Lizneva, D., Suturina, L., Walker, W., et al. (2016b) Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertility and Sterility*, 106 (1): 6–15. doi:10.1016/j.fertnstert.2016.05.003.

Lo, J.C., Feigenbaum, S.L., Yang, J., et al. (2006) Epidemiology and Adverse Cardiovascular Risk Profile of Diagnosed Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology & Metabolism*, 91 (4): 1357–1363. doi:10.1210/jc.2005-2430.

Di Lorenzo, B., Pau, M.C., Zinellu, E., et al. (2023) Association between Red Blood Cell Distribution Width and Obstructive Sleep Apnea Syndrome: A Systematic Review and Meta-Analysis. *Journal of clinical medicine*, 12 (9). doi:10.3390/JCM12093302.

Lourie, E.M., Utidjian, L.H., Ricci, M.F., et al. (2021) Reducing electronic health record-related burnout in providers through a personalized efficiency improvement program. *Journal of the American Medical Informatics Association*, 28 (5). doi:10.1093/jamia/ocaa248.

Louwens, Y. V., Roest-Schalken, M.E., Kleefstra, N., et al. (2014) Excess mortality in

mothers of patients with polycystic ovary syndrome. *Human Reproduction*, 29 (8). doi:10.1093/humrep/deu107.

Ludwig, E. (1977) Classification of the types of androgenetic alopecia (common baldness) occurring in the female sex. *British Journal of Dermatology*, 97 (3): 247–254. doi:10.1111/j.1365-2133.1977.tb15179.x.

Lujan, M.E., Chizen, D.R. and Pierson, R.A. (2008) Diagnostic Criteria for Polycystic Ovary Syndrome: Pitfalls and Controversies. *Journal of Obstetrics and Gynaecology Canada*, 30 (8). doi:10.1016/S1701-2163(16)32915-2.

Lunde, O., Magnus, P., Sandvik, L., et al. (1989) Familial clustering in the polycystic ovarian syndrome. *Gynecologic and obstetric investigation*, 28 (1): 23–30. doi:10.1159/000293493.

Luo, P., Su, Z., Li, P., et al. (2023) Effects of Sleeve Gastrectomy on Patients with Obesity and Polycystic Ovary Syndrome: a Meta-analysis. *Obesity surgery*. doi:10.1007/S11695-023-06617-Z.

Luque-Ramírez, M., Alpañés, M., Sanchón, R., et al. (2015) Referral bias in female functional hyperandrogenism and polycystic ovary syndrome. *European Journal of Endocrinology*, 173 (5). doi:10.1530/EJE-15-0646.

de Lusignan, S., Liaw, S.-T., Dedman, D., et al. (2015) An algorithm to improve diagnostic accuracy in diabetes in computerised problem orientated medical records (POMR) compared with an established algorithm developed in episode orientated records (EOMR). *Journal of innovation in health informatics*, 22 (2): 255–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26245239> (Accessed: 11 July 2016).

Luzzi, V., Mazur, M., Guaragna, M., et al. (2022) Correlations of Obstructive Sleep Apnea Syndrome and Daytime Sleepiness with the Risk of Car Accidents in Adult Working Population: A Systematic Review and Meta-Analysis with a Gender-Based Approach.

Journal of clinical medicine, 11 (14). doi:10.3390/JCM11143971.

Lyu, X., Lyu, T., Wang, X., et al. (2021) The Antiobesity Effect of GLP-1 Receptor Agonists Alone or in Combination with Metformin in Overweight /Obese Women with Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis. *International journal of endocrinology*, 2021. doi:10.1155/2021/6616693.

Madjar, K., Zucknick, M., Ickstadt, K., et al. (2021) Combining heterogeneous subgroups with graph-structured variable selection priors for Cox regression. *BMC Bioinformatics*, 22 (1). doi:10.1186/s12859-021-04483-z.

Maguire, A., Blak, B.T. and Thompson, M. (2009) The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiology and drug safety*, 18 (1): 76–83. doi:10.1002/pds.1688.

Mamede, S., Van Gog, T., Van Den Berge, K., et al. (2010) Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA*, 304 (11). doi:10.1001/jama.2010.1276.

Mantovani, A., Petracca, G., Beatrice, G., et al. (2021) Non-alcoholic fatty liver disease and risk of incident diabetes mellitus: an updated meta-analysis of 501 022 adult individuals. *Gut*, 70 (5): 962–969. doi:10.1136/GUTJNL-2020-322572.

Maria-Sophie Barthel, F. and Royston, P. (2006) *Graphical representation of interactions*.

Marquina, C., Makarounas-Kirchmann, K., Holden, K., et al. (2023) The preventable productivity burden of sleep apnea in Australia: a lifetime modelling study. *Journal of sleep research*, 32 (2). doi:10.1111/JSR.13748.

Martelli, V., Carelli, E., Tomlinson, G.A., et al. (2022) Prevalence of elevated hemoglobin and hematocrit levels in patients with obstructive sleep apnea and the impact of treatment

with continuous positive airway pressure: a meta-analysis. *Hematology (Amsterdam, Netherlands)*, 27 (1): 889–901. doi:10.1080/16078454.2022.2109346.

Martin, J., Girling, A., Nirantharakumar, K., et al. (2016) Intra-cluster and inter-period correlation coefficients for cross-sectional cluster randomised controlled trials for type-2 diabetes in UK primary care. *Trials*, 17 (1): 1–12. doi:10.1186/S13063-016-1532-9/TABLES/7.

Masson, W., Barbagelata, L., Lobo, M., et al. (2023) Plasma Lipoprotein(a) Levels in Polycystic Ovary Syndrome: A Systematic Review and Meta-analysis. *High blood pressure & cardiovascular prevention: the official journal of the Italian Society of Hypertension*. doi:10.1007/S40292-023-00585-2.

McCartney, C.R., Campbell, R.E., Marshall, J.C., et al. (2022) The role of gonadotropin-releasing hormone neurons in polycystic ovary syndrome. *Journal of neuroendocrinology*, 34 (5). doi:10.1111/JNE.13093.

McNamee, R. (2005) Regression modelling and other methods to control confounding. *Occupational and Environmental Medicine*. 62 (7). doi:10.1136/oem.2002.001115.

De Medeiros, S.F., De Medeiros, M.A.S., Ormond, C.M., et al. (2018) Subclinical Hypothyroidism Impact on the Characteristics of Patients with Polycystic Ovary Syndrome. A Meta-Analysis of Observational Studies. *Gynecologic and Obstetric Investigation*. 83 (2). doi:10.1159/000485619.

de Medeiros, S.F., Yamamoto, M.M.W., de Medeiros, M.A.S., et al. (2020) Changes in clinical and biochemical characteristics of polycystic ovary syndrome with advancing age. *Endocrine Connections*, 9 (2). doi:10.1530/EC-19-0496.

MedSci, K.K.F., Schnecke, V., Haase, C.L., et al. (2023) Weight change and risk of obesity-

related complications: A retrospective population-based cohort study of a UK primary care database. *Diabetes, obesity & metabolism*. doi:10.1111/DOM.15154.

Mensch, D. (2023) Optimizing electronic health record efficiency. *Journal of the American Pharmacists Association : JAPhA*, 63 (3): 702–703. doi:10.1016/J.JAPH.2023.04.014.

Merkin, S.S., Azziz, R., Seeman, T., et al. (2011) Socioeconomic status and polycystic ovary syndrome. *Journal of Women's Health*, 20 (3). doi:10.1089/jwh.2010.2303.

Miazgowski, T., Martopullo, I., Widecka, J., et al. (2021) National and regional trends in the prevalence of polycystic ovary syndrome since 1990 within Europe: The modeled estimates from the Global Burden of Disease Study 2016. *Archives of Medical Science*. 17 (2). doi:10.5114/aoms.2019.87112.

Miller, D.P., Gomberg-Maitland, M. and Humbert, M. (2012) Survivor bias and risk assessment. *European Respiratory Journal*. 40 (3). doi:10.1183/09031936.00094112.

Mitchell, M.N. (2012) *A Visual Guide to Stata Graphics, Third Edition*.

Moggetti, P. and Tosi, F. (2021) Insulin resistance and PCOS: chicken or egg? *Journal of endocrinological investigation*, 44 (2): 233–244. doi:10.1007/S40618-020-01351-0.

Mokhlesi, B., Sam, S. and Ehrmann, D.A. (2017) Obstructive sleep apnea and polycystic ovary syndrome: cause or association? *Sleep Medicine*. 36 pp. 170–171. doi:10.1016/j.sleep.2017.01.001.

Molina, A., Huck, O., Herrera, D., et al. (2023) The association between respiratory diseases and periodontitis: A systematic review and meta-analysis. *Journal of clinical periodontology*, 50 (6). doi:10.1111/JCPE.13767.

Morgan, C.L., Jenkins-Jones, S., Currie, C.J., et al. (2012) Evaluation of adverse outcome in young women with polycystic ovary syndrome versus matched, reference controls: a

retrospective, observational study. *The Journal of clinical endocrinology and metabolism*, 97 (9): 3251–60. doi:10.1210/jc.2012-1690.

Mowbray, F.I., Manlongat, D. and Shukla, M. (2022) Sensitivity Analysis: A Method to Promote Certainty and Transparency in Nursing and Health Research. *The Canadian journal of nursing research = Revue canadienne de recherche en sciences infirmieres*, 54 (4). doi:10.1177/08445621221107108.

Nadir, A., Temizkan, S., Ozderya, A., et al. (2019) Dehydroepiandrosterone sulfate/free androgen index ratio predicts a favorable metabolic profile in patients with polycystic ovary syndrome. *Gynecological endocrinology : the official journal of the International Society of Gynecological Endocrinology*, 35 (2): 128–132. doi:10.1080/09513590.2018.1505843.

Nakasian, S.S., Rassen, J.A. and Franklin, J.M. (2017) Effects of expanding the look-back period to all available data in the assessment of covariates. *Pharmacoepidemiology and Drug Safety*, 26 (8). doi:10.1002/pds.4210.

Natsky, A.N., Chai-Coetzer, C.L., Vakulin, A., et al. (2022) Economic evaluation of diagnostic sleep studies for obstructive sleep apnoea in the adult population: a systematic review. *Sleep Medicine Reviews*. 62. doi:10.1016/j.smrv.2022.101608.

NHS (2014) *Personalised Health and Care 2020 Using Data and Technology to Transform Outcomes for Patients and Citizens A Framework for Action*.

NHS Digital (2023) *General practice workforce 2023 May*. Available at: <https://digital.nhs.uk/data-and-information/publications/statistical/general-and-personal-medical-services/31-may-2023> (Accessed: 20 July 2023).

NICE (2022) *NICE real-world evidence framework Corporate document*. Available at: www.nice.org.uk/corporate/ecd9.

Nicholls, S.G., Langan, S.M. and Benchimol, E.I. (2017) Routinely collected data: The importance of high-quality diagnostic coding to research. *CMAJ*. 189 (33). doi:10.1503/cmaj.170807.

Nicholls, S.G., Quach, P., Von Elm, E., et al. (2015) The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement: Methods for arriving at consensus and developing reporting guidelines. *PLoS ONE*, 10 (5). doi:10.1371/journal.pone.0125620.

Nicholson, A., Ford, E., Davies, K.A., et al. (2013) Optimising use of electronic health records to describe the presentation of rheumatoid arthritis in primary care: a strategy for developing code lists. *PloS one*, 8 (2): e54878. doi:10.1371/journal.pone.0054878.

Nicholson, A., Tate, A.R., Koeling, R., et al. (2011) What does validation of cases in electronic record databases mean? The potential contribution of free text. *Pharmacoepidemiology and Drug Safety*, 20 (3). doi:10.1002/pds.2086.

Nwankwo, M., Okamkpa, C.J. and Danborno, B. (2022) Comparison of diagnostic criteria and prevalence of metabolic syndrome using WHO, NCEP-ATP III, IDF and harmonized criteria: A case study from urban southeast Nigeria. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 16 (12). doi:10.1016/j.dsx.2022.102665.

Nylander, M., Frøssing, S., Bjerre, A.H., et al. (2017) Ovarian morphology in polycystic ovary syndrome: Estimates from 2D and 3D ultrasound and magnetic resonance imaging and their correlation to anti-Müllerian hormone. *Acta Radiologica*, 58 (8): 997–1004. doi:10.1177/0284185116676656.

Office for National Statistics (2022) *Trends in patient-to-staff numbers at GP practices in England 2022*.

Ohno-Machado, L. (2014) Electronic health record systems: risks and benefits. *Journal of the*

American Medical Informatics Association: JAMIA. 21 (e1). doi:10.1136/amiajnl-2014-002635.

Okoth, K., Smith, W.P., Thomas, G.N., et al. (2023) The association between menstrual cycle characteristics and cardiometabolic outcomes in later life: a retrospective matched cohort study of 704,743 women from the UK. *BMC medicine*, 21 (1). doi:10.1186/s12916-023-02794-x.

O'Reilly, M.W., Glisic, M., Kumarendran, B., et al. (2019) Serum testosterone, sex hormone-binding globulin and sex-specific risk of incident type 2 diabetes in a retrospective primary care cohort. *Clinical Endocrinology*, 90 (1): 145–154. doi:10.1111/cen.13862.

O'Reilly, M.W., Taylor, A.E., Crabtree, N.J., et al. (2014) Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *The Journal of clinical endocrinology and metabolism*, 99 (3): 1027–36. doi:10.1210/jc.2013-3399.

Oyinlola, J.O., Campbell, J. and Kousoulis, A.A. (2016) Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Services Research*. 16 (1). doi:10.1186/s12913-016-1562-8.

Ozonze, O., Scott, P.J. and Hopgood, A.A. (2023) Automating Electronic Health Record Data Quality Assessment. *Journal of Medical Systems*. 47 (1). doi:10.1007/s10916-022-01892-2.

Painter, J.L. (2010) “Toward automating an inference model on unstructured terminologies: Oxmis case study.” *In Advances in Experimental Medicine and Biology*. 2010. doi:10.1007/978-1-4419-5913-3_71.

Pardridge, W.M. (1986) Serum bioavailability of sex steroid hormones. *Clinics in Endocrinology and Metabolism*, 15 (2): 259–278. doi:10.1016/S0300-595X(86)80024-X.

Pasquali, R. and Gambineri, A. (2015) A comprehensive approach in diagnosing the polycystic ovary syndrome. *Women's Health*. 11 (4) pp. 501–512. doi:10.2217/whe.15.13.

Peddemul, A., Tejavath, S., Hassan, D., et al. (2022) Influence of Subclinical Hypothyroidism on Women With Polycystic Ovary Syndrome: A Literature Review. *Cureus*. doi:10.7759/cureus.28468.

Pei, C.Z., Jin, L. and Baek, K.H. (2021) Pathogenetic analysis of polycystic ovary syndrome from the perspective of omics. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 142. doi:10.1016/J.BIOPHA.2021.112031.

Petersen, I., Welch, C.A., Nazareth, I., et al. (2019) Health indicator recording in UK primary care electronic health records: Key implications for handling missing data. *Clinical Epidemiology*, 11. doi:10.2147/CLEP.S191437.

Petrides, A.K., Bixho, I., Goonan, E.M., et al. (2017) The benefits and challenges of an interfaced electronic health record and laboratory information system: Effects on laboratory processes. *Archives of Pathology and Laboratory Medicine*, 141 (3). doi:10.5858/arpa.2016-0146-OA.

Pham, A., Cummings, M., Lindeman, C., et al. (2019) Recognizing misclassification bias in research and medical practice. *Family Practice*, 36 (6). doi:10.1093/fampra/cmy130.

Pines, J.M. (2006) Profiles in patient safety: Confirmation bias in emergency medicine. *Academic Emergency Medicine*, 13 (1). doi:10.1197/j.aem.2005.07.028.

Poorolajal, J., Cheraghi, Z., Irani, A.D., et al. (2011) Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. *Epidemiology and Health*, 33. doi:10.4178/epih/e2011005.

Qiu, S., Cai, X., Sun, Z., et al. (2017) Association between physical activity and risk of

nonalcoholic fatty liver disease: a meta-analysis. *Therapeutic advances in gastroenterology*, 10 (9): 701–713. doi:10.1177/1756283X17725977.

Quek, J., Chan, K.E., Wong, Z.Y., et al. (2023) Global prevalence of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in the overweight and obese population: a systematic review and meta-analysis. *The Lancet Gastroenterology and Hepatology*, 8 (1). doi:10.1016/S2468-1253(22)00317-X.

Raciborski, R. (2008) Kountry: A Stata utility for merging cross-country data from multiple sources. *Stata Journal*, 8 (3). doi:10.1177/1536867x0800800305.

Raeisi, T., Rezaie, H., Darand, M., et al. (2021) Circulating resistin and follistatin levels in obese and non-obese women with polycystic ovary syndrome: A systematic review and meta-analysis. *PLoS ONE*. 16 (3 March). doi:10.1371/journal.pone.0246200.

Rajbhandari, P., Auron, M., Worley, S., et al. (2021) Improving documentation of inpatient problem list in electronic health record: A quality improvement project. *Journal of Patient Safety*, 17 (8). doi:10.1097/PTS.0000000000000490.

Ramezani Tehrani, F., Amiri, M., Behboudi-Gandevani, S., et al. (2020) Cardiovascular events among reproductive and menopausal age women with polycystic ovary syndrome: a systematic review and meta-analysis. *Gynecological Endocrinology*, 36 (1): 12–23. doi:10.1080/09513590.2019.1650337.

Ranganathan, P., Pramesh, C. and Buyse, M. (2016) Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research*, 7 (2). doi:10.4103/2229-3485.179436.

Rao, V.S., Cowan, S., Armour, M., et al. (2022) A Global Survey of Ethnic Indian Women Living with Polycystic Ovary Syndrome: Co-Morbidities, Concerns, Diagnosis Experiences,

Quality of Life, and Use of Treatment Methods. *International journal of environmental research and public health*, 19 (23). doi:10.3390/IJERPH192315850.

Ravani, P., Parfrey, P.S., Dicks, E., et al. (2007) Clinical research of kidney diseases II: Problems of study design. *Nephrology Dialysis Transplantation*, 22 (10). doi:10.1093/ndt/gfm433.

RCGP, BMA and DH (2011) *The Good Practice Guidelines for GP electronic patient records*.

Read, J.D. (2009) “Read codes and medical audit.” In *Medical Audit*. Cambridge University Press. pp. 125–137. doi:10.1017/cbo9780511526718.012.

Rebar, R. (2000) *Evaluation of Amenorrhea, Anovulation, and Abnormal Bleeding*. MDText.com, Inc. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK279144/> (Downloaded: 23 July 2022).

Rector, A.L. (1999) Clinical terminology: Why is it so hard? *Methods of Information in Medicine*. 38 (4–5). doi:10.1055/s-0038-1634418.

Rees, D.A., Jenkins-Jones, S. and Morgan, C.L. (2016) Contemporary Reproductive Outcomes for Patients With Polycystic Ovary Syndrome: A Retrospective Observational Study. *The Journal of clinical endocrinology and metabolism*, 101 (4): 1664–1672. doi:10.1210/JC.2015-2682.

Reeves, A., Elliott, M.R., Lewis, T.T., et al. (2022) Study Selection Bias and Racial or Ethnic Disparities in Estimated Age at Onset of Cardiometabolic Disease Among Midlife Women in the US. *JAMA network open*, 5 (11). doi:10.1001/jamanetworkopen.2022.40665.

Reis, Z.S.N., Maia, T.A., Marcolino, M.S., et al. (2017) Is there evidence of cost benefits of electronic medical records, standards, or interoperability in hospital information systems?

overview of systematic reviews. *JMIR Medical Informatics*. 5 (3). doi:10.2196/medinform.7400.

Rempert, A.N., Sarria, I., Standeven, L.R., et al. (2023) A Systematic Review of the Psychosocial Impact of Polycystic Ovarian Syndrome Before and After Treatment. *Reproductive sciences (Thousand Oaks, Calif.)*. doi:10.1007/S43032-023-01285-X.

Riazi, K., Azhari, H., Charette, J.H., et al. (2022) The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *The lancet. Gastroenterology & hepatology*, 7 (9): 851–861. doi:10.1016/S2468-1253(22)00165-0.

Riestenberg, C., Jagasia, A., Markovic, D., et al. (2022) Health Care-Related Economic Burden of Polycystic Ovary Syndrome in the United States: Pregnancy-Related and Long-Term Health Consequences. *The Journal of clinical endocrinology and metabolism*, 107 (2): 575–585. doi:10.1210/CLINEM/DGAB613.

Riley, J., Antza, C., Kempegowda, P., et al. (2021) Social deprivation and incident diabetes-related foot disease in patients with type 2 diabetes: A population-based cohort study. *Diabetes Care*, 44 (3): 731–739. doi:10.2337/dc20-1027.

Rivas-Ruiz, F., Pérez-Vicente, S. and González-Ramírez, A.R. (2013) Bias in clinical epidemiological study designs. *Allergologia et Immunopathologia*, 41 (1). doi:10.1016/j.aller.2012.04.005.

Robinson, D., Schulz, E., Brown, P., et al. (1997) Updating the Read Codes: User-interactive Maintenance of a Dynamic Clinical Vocabulary. *Journal of the American Medical Informatics Association*, 4 (6). doi:10.1136/jamia.1997.0040465.

La Rochelle, P. (2014) Basic learning concepts in EBM: The bidimensional hierarchy of evidence. *Evidence-Based Medicine*. 19 (3). doi:10.1136/eb-2013-101610.

Rodgers, L.R., Weedon, M.N., Henley, W.E., et al. (2017) Cohort profile for the MASTERMIND study: Using the Clinical Practice Research Datalink (CPRD) to investigate stratification of response to treatment in patients with type 2 diabetes. *BMJ Open*, 7 (10). doi:10.1136/bmjopen-2017-017989.

Romitti, M., Fabris, V.C., Ziegelmann, P.K., et al. (2018) Association between PCOS and autoimmune thyroid disease: a systematic review and meta-analysis. *Endocrine connections*, 7 (11): 1158–1167. doi:10.1530/EC-18-0309.

Rooney, M.R., Fang, M., Ogurtsova, K., et al. (2023) Global Prevalence of Prediabetes. *Diabetes care*. doi:10.2337/DC22-2376.

Rosenfield, R.L. (2013) Clinical review: Adolescent anovulation: maturational mechanisms and implications. *The Journal of clinical endocrinology and metabolism*, 98 (9): 3572–3583. doi:10.1210/JC.2013-1770.

Rothman, K.J., Wise, L.A. and Hatch, E.E. (2011) Should graphs of risk or rate ratios be plotted on a log scale? *American Journal of Epidemiology*. 174 (3). doi:10.1093/aje/kwr156.

Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group (2004) Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Human reproduction (Oxford, England)*, 19 (1): 41–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14688154> (Accessed: 14 May 2017).

Rubin, K.H., Andersen, M.S., Abrahamsen, B., et al. (2019) Socioeconomic status in Danish women with polycystic ovary syndrome: A register-based cohort study. *Acta Obstetrica et Gynecologica Scandinavica*, 98 (4). doi:10.1111/aogs.13514.

Rudnicka, E., Duszewska, A.M., Kucharski, M., et al. (2022) Oxidative stress and reproductive function: Oxidative stress in polycystic ovary syndrome. *Reproduction*, 164:

F145–F144. doi:10.1530/REP.

Rudnicka, E., Suchta, K., Grymowicz, M., et al. (2021) Chronic low grade inflammation in pathogenesis of pcos. *International Journal of Molecular Sciences*, 22 (7). doi:10.3390/ijms22073789.

Rudrapatna, V.A. and Butte, A.J. (2020) Opportunities and challenges in using real-world data for health care. *Journal of Clinical Investigation*. 130 (2). doi:10.1172/JCI129197.

Ruigómez, A., Brobert, G., Vora, P., et al. (2021) Validation of venous thromboembolism diagnoses in patients receiving rivaroxaban or warfarin in The Health Improvement Network. *Pharmacoepidemiology and Drug Safety*, 30 (2). doi:10.1002/pds.5146.

Saeedi, P., Petersohn, I., Salpea, P., et al. (2019) Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Research and Clinical Practice*, 157. doi:10.1016/j.diabres.2019.107843.

Salisbury, C., Johnson, L., Purdy, S., et al. (2011) Epidemiology and impact of multimorbidity in primary care: A retrospective cohort study. *British Journal of General Practice*, 61 (582). doi:10.3399/bjgp11X548929.

Sammon, C.J. and Petersen, I. (2016) Backdating of events in electronic primary health care data: should one censor at the date of last data collection. *Pharmacoepidemiology and drug safety*, 25 (4): 378–84. doi:10.1002/pds.3981.

Sattar, N. and Preiss, D. (2017) Reverse Causality in Cardiovascular Epidemiological Research: More Common Than Imagined? *Circulation*. 135 (24). doi:10.1161/CIRCULATIONAHA.117.028307.

Schneeweiss, S. (2018) Automated data-adaptive analytics for electronic healthcare data to

study causal treatment effects. *Clinical Epidemiology*. 10. doi:10.2147/CLEP.S166545.

Schneeweiss, S., Rassen, J.A., Brown, J.S., et al. (2019) Graphical depiction of longitudinal study designs in health care databases. *Annals of Internal Medicine*, 170 (6). doi:10.7326/M18-3079.

Schuemie, M.J., Ryan, P.B., Man, K.K.C., et al. (2019) A plea to stop using the case-control design in retrospective database studies. *Statistics in Medicine*, 38 (22). doi:10.1002/sim.8215.

Schulz, E.B., Barrett, J.W. and Price, C. (1998) Read code quality assurance: From simple syntax to semantic stability. *Journal of the American Medical Informatics Association*. 5 (4). doi:10.1136/jamia.1998.0050337.

Schulz, E.B., Price, C. and Brown, P.J.B. (1997) Symbolic anatomic knowledge representation in the Read Codes version 3: structure and application. *Journal of the American Medical Informatics Association : JAMIA*, 4 (1): 38–48. doi:10.1136/JAMIA.1997.0040038.

Schumm, L.P. and Thisted, R.A. (2005) Reproducible Research Using Stata. *North American Stata Users' Group Meetings*. (16).

Seaman, H.E., De Vries, C.S. and Farmer, R.D.T. (2003a) Differences in the use of combined oral contraceptives amongst women with and without acne. *Human reproduction (Oxford, England)*, 18 (3): 515–521. doi:10.1093/HUMREP/DEG090.

Seaman, H.E., de Vries, C.S. and Farmer, R.D.T. (2003b) The risk of liver disorders in women prescribed cyproterone acetate in combination with ethinyloestradiol (Dianette): a nested case-control study using the GPRD. *Pharmacoepidemiology and drug safety*, 12 (7): 541–550. doi:10.1002/PDS.857.

Seaman, H.E., de Vries, C.S. and Farmer, R.D.T. (2004) Venous thromboembolism associated with cyproterone acetate in combination with ethinyloestradiol (Dianette): observational studies using the UK General Practice Research Database. *Pharmacoepidemiology and drug safety*, 13 (7): 427–436. doi:10.1002/PDS.896.

Senathirajah, Y., Kaufman, D. and Bakken, S. (2014) The clinician in the driver's seat: Part 2 - Intelligent uses of space in a drag/drop user-composable electronic health record. *Journal of Biomedical Informatics*, 52. doi:10.1016/j.jbi.2014.09.008.

Sendur, S.N. and Yildiz, B.O. (2021) Influence of ethnicity on different aspects of polycystic ovary syndrome: a systematic review. *Reproductive biomedicine online*, 42 (4): 799–818. doi:10.1016/J.RBMO.2020.12.006.

Shaheen, A.A., Kaplan, G.G., Sharkey, K.A., et al. (2021) Impact of major depression and antidepressant use on alcoholic and non-alcoholic fatty liver disease: A population-based study. *Liver International*, 41 (10). doi:10.1111/liv.14973.

Shayeb, M. El, Topfer, L.A., Stafinski, T., et al. (2014) Diagnostic accuracy of level 3 portable sleep tests versus level 1 polysomnography for sleep-disordered breathing: A systematic review and meta-analysis. *CMAJ. Canadian Medical Association Journal*, 186 (1). doi:10.1503/cmaj.130952.

Sheikh, J., Khalil, H., Shaikh, S., et al. (2023) Emotional and psychosexual well-being is influenced by ethnicity and birthplace in women and individuals with polycystic ovary syndrome in the UK and India. *BJOG: an international journal of obstetrics and gynaecology*, 130 (8). doi:10.1111/1471-0528.17428.

Shlipak, M., MD, M.P.H., Angeja, B., et al. (2001) Hormone Therapy and In-Hospital Survival After Myocardial Infarction in Postmenopausal Women. *Circulation*, 104 (19).

Silva, M.S.B. and Campbell, R.E. (2022) Polycystic Ovary Syndrome and the

Neuroendocrine Consequences of Androgen Excess. *Comprehensive Physiology*, 12 (2): 3347–3369. doi:10.1002/CPHY.C210025.

Smith, H.C., Saxena, S. and Petersen, I. (2022) Maternal Postnatal Depression and Completion of Infant Immunizations: A UK Cohort Study of 196,329 Mother-Infant Pairs, 2006–2015. *Journal of Clinical Psychiatry*, 83 (4). doi:10.4088/JCP.20M13575.

Smith, N., Wilson, A. and Weekes, T. (1995) Use of Read codes in development of a standard data set. *BMJ*, 311 (7000). doi:10.1136/bmj.311.7000.313.

Sola-Leyva, A., Pérez-Prieto, I., Molina, N.M., et al. (2023) Microbial composition across body sites in polycystic ovary syndrome: a systematic review and meta-analysis. *Reproductive biomedicine online*. doi:10.1016/J.RBMO.2023.03.016.

Sommers, S.C. and Wadman, P.J. (1956) Pathogenesis of polycystic ovaries. *American journal of obstetrics and gynecology*, 72 (1): 160–169. doi:10.1016/S0002-9378(16)37526-3.

Springate, D.A., Kontopantelis, E., Ashcroft, D.M., et al. (2014) ClinicalCodes: An online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PLoS ONE*, 9 (6). doi:10.1371/journal.pone.0099825.

Springate, D.A., Parisi, R., Olier, I., et al. (2017) rEHR: An R package for manipulating and analysing electronic health record data. *PLoS ONE*, 12 (2). doi:10.1371/journal.pone.0171784.

van Staa, T.P. and Sprafka, J.M. (2009) Study of adverse outcomes in women using testosterone therapy. *Maturitas*, 62 (1). doi:10.1016/j.maturitas.2008.11.001.

Stein, I.F. and Leventhal, M.L. (1935) Amenorrhea associated with bilateral polycystic ovaries. *American Journal of Obstetrics and Gynecology*, 29 (2): 181–191. doi:10.1016/S0002-9378(15)30642-6.

Stien, Irving.F. (1952) Bilateral polycystic ovaries. *Journal of the South Carolina Medical Association* (1975), 48 (6): 154–155. doi:10.1016/0002-9378(45)90140-2.

Stram, M., Gigliotti, T., Hartman, D., et al. (2020) Logical observation identifiers names and codes for laboratorians potential solutions and challenges for interoperability. *Archives of Pathology and Laboratory Medicine*. 144 (2). doi:10.5858/arpa.2018-0477-RA.

Subramanian, A., Adderley, N.J., Tracy, A., et al. (2019) Risk of incident obstructive sleep apnea among patients with type 2 diabetes. *Diabetes Care*, 42 (5). doi:10.2337/dc18-2004.

Subramanian, A., Anand, A., Adderley, N.J., et al. (2021) Increased COVID-19 infections in women with polycystic ovary syndrome: a population-based study. *European journal of endocrinology*, 184 (5): 637–645. doi:10.1530/EJE-20-1163.

Subramanian, A., Han, D., Braithwaite, T., et al. (2022a) Angiotensin-converting enzyme inhibitors and risk of age-related macular degeneration in individuals with hypertension. *British Journal of Clinical Pharmacology*, 88 (9). doi:10.1111/bcp.15366.

Subramanian, A., Lee, S.I., Phillips, K., et al. (2022b) Polycystic ovary syndrome and risk of adverse obstetric outcomes: a retrospective population-based matched cohort study in England. *BMC Medicine*, 20 (1). doi:10.1186/s12916-022-02473-3.

Suissa, S. (2008) Immortal time bias in pharmaco-epidemiology. *American journal of epidemiology*.

Suissa, S. and Ernst, P. (2020) Avoiding immortal time bias in observational studies. *European Respiratory Journal*. 55 (3). doi:10.1183/13993003.02494-2019.

Sun, J., Wang, Y., Zhang, X., et al. (2020) Prevalence of peripheral neuropathy in patients with diabetes: A systematic review and meta-analysis. *Primary Care Diabetes*, 14 (5): 435–444. doi:10.1016/J.PCD.2019.12.005.

Sun, X., Ioannidis, J.P.A., Agoritsas, T., et al. (2014) How to use a subgroup analysis users' guides to the medical literature. *JAMA*. 311 (4). doi:10.1001/jama.2013.285063.

Sun, X., Yu, J., Luo, J., et al. (2022) Meta-analysis of the association between obstructive sleep apnea and postoperative complications. *Sleep medicine*, 91: 1–11. doi:10.1016/J.SLEEP.2021.11.019.

Sutton-Tyrrell, K., Wildman, R.P., Matthews, K.A., et al. (2005) “Sex hormone-binding globulin and the free androgen index are related to cardiovascular risk factors in multiethnic premenopausal and perimenopausal women enrolled in the study of women across the nation (SWAN).” *In Circulation*. 2005. doi:10.1161/01.CIR.0000157697.54255.CE.

Swift, B., Jain, L., White, C., et al. (2018) Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care. *Clinical and Translational Science*. 11 (5). doi:10.1111/cts.12559.

Taggar, J.S., Coleman, T., Lewis, S., et al. (2012) The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. *BMC public health*, 12: 329. doi:10.1186/1471-2458-12-329.

Talari, K. and Goyal, M. (2020) Retrospective studies - Utility and caveats. *Journal of the Royal College of Physicians of Edinburgh*, 50 (4). doi:10.4997/JRCPE.2020.409.

Tang, A.S.P., Chan, K.E., Quek, J., et al. (2022) Non-alcoholic fatty liver disease increases risk of carotid atherosclerosis and ischemic stroke: An updated meta-analysis with 135,602 individuals. *Clinical and molecular hepatology*, 28 (3): 483–496. doi:10.3350/CMH.2021.0406.

Teede, H.J., Misso, M.L., Costello, M.F., et al. (2018a) Recommendations from the

international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertility and Sterility*, 110 (3): 364–379. doi:10.1016/j.fertnstert.2018.05.004.

Teede, H.J., Misso, M.L., Costello, M.F., et al. (2018b) Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Human reproduction (Oxford, England)*, 33 (9): 1602–1618. doi:10.1093/humrep/dey256.

Teo, Y.H., Tan, B.K.J., Tan, N.K.W., et al. (2022) Obstructive sleep apnea and the incidence and mortality of gastrointestinal cancers: a systematic review and meta-analysis of 5,120,837 participants. *Journal of Gastrointestinal Oncology*, 13 (6): 2789–2798. doi:10.21037/JGO-22-153/COIF).

Teo, Z.L., Tham, Y.C., Yu, M., et al. (2021) Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis. *Ophthalmology*, 128 (11): 1580–1591. doi:10.1016/j.ophtha.2021.04.027.

Tian, D., Chen, W., Xu, Q., et al. (2022) Liraglutide monotherapy and add on therapy on obese women with polycystic ovarian syndromes: a systematic review and meta-analysis. *Minerva medica*, 113 (3): 542–550. doi:10.23736/S0026-4806.21.07085-3.

Tian, Z., Zhang, Y.C., Wang, Y., et al. (2021) Effects of bariatric surgery on patients with obesity and polycystic ovary syndrome: a meta-analysis. *Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery*, 17 (8): 1399–1408. doi:10.1016/J.SOARD.2021.04.009.

Toulis, K.A., Willis, B.H., Marshall, T., et al. (2017) All-cause mortality in patients with diabetes under treatment with dapagliflozin: a population-based, open-cohort study in THIN database. *The Journal of Clinical Endocrinology & Metabolism*. doi:10.1210/jc.2016-3446.

Treasure, W. (2011) The primary care electronic health record: Who's righting the software? *British Journal of General Practice*, 61 (583). doi:10.3399/bjgp11X556443.

Tremellen, K. and Pearce, K. (2012) Dysbiosis of Gut Microbiota (DOGMA)--a novel theory for the development of Polycystic Ovarian Syndrome. *Medical hypotheses*, 79 (1): 104–112. doi:10.1016/J.MEHY.2012.04.016.

Turcu, A.F. and Auchus, R.J. (2017) Clinical significance of 11-oxygenated androgens. *Current opinion in endocrinology, diabetes, and obesity*, 24 (3): 252–259. doi:10.1097/MED.0000000000000334.

Tyrer, F., Bhaskaran, K. and Rutherford, M.J. (2022) Immortal time bias for life-long conditions in retrospective observational studies using electronic health records. *BMC Medical Research Methodology*, 22 (1). doi:10.1186/s12874-022-01581-1.

Unwin, N., Shaw, J., Zimmet, P., et al. (2002) “Impaired glucose tolerance and impaired fasting glycaemia: The current status on definition and intervention.” *In Diabetic Medicine*. 2002. doi:10.1046/j.1464-5491.2002.00835.x.

Vandenbroucke, J. and Pearce, N. (2015) Point: Incident Exposures, Prevalent Exposures, and Causal Inference: Does Limiting Studies to Persons Who Are Followed from First Exposure Onward Damage Epidemiology? *American Journal of Epidemiology*, 182 (10). doi:10.1093/aje/kwv225.

VanHise, K., Wang, E.T., Norris, K., et al. (2023) Racial and ethnic disparities in polycystic ovary syndrome. *Fertility and Sterility*. 119 (3) pp. 348–354. doi:10.1016/j.fertnstert.2023.01.031.

Vatcheva, K., Lee, M., McCormick, J., et al. (2016) The Effect of Ignoring Statistical Interactions in Regression Analyses Conducted in Epidemiologic Studies: An Example with Survival Analysis Using Cox Proportional Hazards Regression Model. *Epidemiology: Open*

Access, 06 (01). doi:10.4172/2161-1165.1000216.

Venn, J. (1880) On the diagrammatic and mechanical representation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10 (59): 1–18. doi:10.1080/14786448008626877.

Vezyridis, P. and Timmons, S. (2016) Evolution of primary care databases in UK: A scientometric analysis of research output. *BMJ Open*, 6 (10). doi:10.1136/bmjopen-2016-012785.

Vounzoulaki, E., Khunti, K., Miksza, J.K., et al. (2022) Screening for type 2 diabetes after a diagnosis of gestational diabetes by ethnicity: A retrospective cohort study. *Primary Care Diabetes*, 16 (3). doi:10.1016/j.pcd.2022.03.008.

Vuokko, R., Vakkuri, A. and Palojoki, S. (2023) Systematized Nomenclature of Medicine–Clinical Terminology (SNOMED CT) Clinical Use Cases in the Context of Electronic Health Record Systems: Systematic Literature Review. *JMIR Medical Informatics*. 11. doi:10.2196/43750.

Vusirikala, A., Thomas, T., Bhala, N., et al. (2020) Impact of obesity and metabolic health status in the development of non-alcoholic fatty liver disease (NAFLD): A United Kingdom population-based cohort study using the health improvement network (THIN). *BMC Endocrine Disorders*, 20 (1). doi:10.1186/s12902-020-00582-9.

Wagner, M., Grodstein, F., Leffondre, K., Samieri, C., & Proust-Lima, C. (2021). Time-varying associations between an exposure history and a subsequent health outcome: a landmark approach to identify critical windows. *BMC Medical Research Methodology*, 21(1). <https://doi.org/10.1186/s12874-021-01403-w>

Wall, H., Smith, C. and Hubbard, R. (2012) Body mass index and obstructive sleep apnoea in the UK: a cross-sectional study of the over-50s. *Primary care respiratory journal : journal of*

the General Practice Airways Group, 21 (4): 371–6. doi:10.4104/pcrj.2012.00053.

Wang, A., Mo, T., Li, Q., et al. (2019) The effectiveness of metformin, oral contraceptives, and lifestyle modification in improving the metabolism of overweight women with polycystic ovary syndrome: a network meta-analysis. *Endocrine*, 64 (2): 220–232. doi:10.1007/S12020-019-01860-W.

Wang, C., Xu, F., Chen, M., et al. (2022) Association of Obstructive Sleep Apnea-Hypopnea Syndrome with hearing loss: A systematic review and meta-analysis. *Frontiers in neurology*, 13. doi:10.3389/FNEUR.2022.1017982.

Wang, E.T., Cirillo, P.M., Vittinghoff, E., et al. (2011) Menstrual irregularity and cardiovascular mortality. *Journal of Clinical Endocrinology and Metabolism*, 96 (1). doi:10.1210/jc.2010-1709.

Wang, E.T., Kao, C.N., Shinkai, K., et al. (2013) Phenotypic comparison of Caucasian and Asian women with polycystic ovary syndrome: a cross-sectional study. *Fertility and Sterility*, 100 (1): 214–218. doi:10.1016/J.FERTNSTERT.2013.03.010.

Wang, R., Lagakos, S.W., Ware, J.H., et al. (2007) Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *New England Journal of Medicine*, 357 (21). doi:10.1056/nejmsr077003.

Wang, R. and Mol, B.W.J. (2017) The Rotterdam criteria for polycystic ovary syndrome: Evidence-based criteria? *Human Reproduction*, 32 (2): 261–264. doi:10.1093/humrep/dew287.

Wang, R. and Ware, J.H. (2013) Detecting Moderator Effects Using Subgroup Analyses. *Prevention Science*, 14 (2): 111–120. doi:10.1007/s11121-011-0221-x.

Wang, S. V., Schneeweiss, S., Berger, M.L., et al. (2017) Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0.

Value in Health, 20 (8). doi:10.1016/j.jval.2017.08.3018.

Wang, X., Piantadosi, S., Le-Rademacher, J., et al. (2021) Statistical Considerations for Subgroup Analyses. *Journal of Thoracic Oncology*, 16 (3). doi:10.1016/j.jtho.2020.12.008.

Wardle, M. and Spencer, A. (2017) Implementation of SNOMED CT in an online clinical database. *Future Hospital Journal*, 4 (2). doi:10.7861/futurehosp.4-2-126.

Warlick, R.D. and Tardiff, B.E. (1998) What in the Word? Development of a Structured Clinical Vocabulary Database. *Proceedings of the AMIA Symposium*, p. 1094. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232136/> (Accessed: 23 July 2023).

Warner, J.L., Smith, J. and Wright, A. (2019) It's time to Wikify clinical documentation: How collaborative authorship can reduce the burden and improve the quality of the electronic health record. *Academic Medicine*. 94 (5). doi:10.1097/ACM.0000000000002613.

Weiskopf, N.G., Dorr, D.A., Jackson, C., et al. (2023) Healthcare utilization is a collider: an introduction to collider bias in EHR data reuse. *Journal of the American Medical Informatics Association*. 30 (5). doi:10.1093/jamia/ocad013.

Wiencek, J.R., McCartney, C.R., Chang, A.Y., et al. (2019) Challenges in the assessment and diagnosis of polycystic ovary syndrome. *Clinical Chemistry*. 65 (3). doi:10.1373/clinchem.2017.284331.

Wierman, M.E. (2021) Hyperandrogenic Anovulation: Differential Diagnosis and Evaluation. *Endocrinology and Metabolism Clinics of North America*. 50 (1) pp. 1–10. doi:10.1016/j.ecl.2020.12.003.

Wijeyaratne, C.N., Balen, A.H., Barth, J.H., et al. (2002) Clinical manifestations and insulin resistance (IR) in polycystic ovary syndrome (PCOS) among South Asians and Caucasians: Is there a difference? *Clinical Endocrinology*, 57 (3): 343–350. doi:10.1046/J.1365-2265.2002.01603.X.

Williams, R., Kontopantelis, E., Buchan, I., et al. (2017) Clinical code set engineering for reusing EHR data for research: A review. *Journal of Biomedical Informatics*. 70.

doi:10.1016/j.jbi.2017.04.010.

Witchel, S.F., Oberfield, S.E. and Peña, A.S. (2019) Polycystic Ovary Syndrome: Pathophysiology, Presentation, and Treatment with Emphasis on Adolescent Girls. *Journal of the Endocrine Society*. 3 (8) pp. 1545–1573. doi:10.1210/js.2019-00078.

Wolf, A., Dedman, D., Campbell, J., et al. (2019) Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International Journal of Epidemiology*, 48 (6). doi:10.1093/ije/dyz034.

Wolf, W.M., Wattick, R.A., Kinkade, O.N., et al. (2018) Geographical Prevalence of Polycystic Ovary Syndrome as Determined by Region and Race/Ethnicity. *International journal of environmental research and public health*, 15 (11). doi:10.3390/IJERPH15112589.

World Health Organisation (2011) *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus Abbreviated Report of a WHO Consultation*.

World Health Organization and International Diabetes Federation (2006) *Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation*. 2006th ed.

Xiao, T., Whitmore, G.A., He, X., et al. (2012) Threshold regression for time-to-event analysis: The stthreg package. *Stata Journal*, 12 (2). doi:10.1177/1536867x1201200206.

Xing, Y., Chen, J., Liu, J., et al. (2021) The Impact of Subclinical Hypothyroidism on Patients with Polycystic Ovary Syndrome: A Meta-Analysis. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et métabolisme*, 53 (6): 382–390. doi:10.1055/A-1463-3198.

Yang, R., Li, Q., Zhou, Z., et al. (2022) Changes in the Prevalence of Polycystic Ovary

Syndrome in China Over the Past Decade. *SSRN Electronic Journal*. doi:10.2139/ssrn.4048835.

Yildiz, B.O., Bolour, S., Woods, K., et al. (2010) Visually scoring hirsutism. *Human reproduction update*, 16 (1): 51–64. doi:10.1093/humupd/dmp024.

Yilmaz, B., Vellanki, P., Ata, B., et al. (2018) Metabolic syndrome, hypertension, and hyperlipidemia in mothers, fathers, sisters, and brothers of women with polycystic ovary syndrome: a systematic review and meta-analysis. *Fertility and sterility*, 109 (2): 356–364.e32. doi:10.1016/J.FERTNSTERT.2017.10.018.

Yu, O., Christ, J.P., Schulze-Rath, R., et al. (2023) Incidence, prevalence, and trends in polycystic ovary syndrome diagnosis: a United States population-based study from 2006 to 2019. *American Journal of Obstetrics and Gynecology*. doi:10.1016/j.ajog.2023.04.010.

Yu, Y.H., Filion, K.B., Reynier, P., et al. (2021) Use of levothyroxine among pregnant women with subclinical hypothyroidism in the United Kingdom: A population-based assessment. *Pharmacology Research and Perspectives*, 9 (5). doi:10.1002/prp2.848.

Zeng, X., Ren, Y., Wu, K., et al. (2023) Association Between Smoking Behavior and Obstructive Sleep Apnea: A Systematic Review and Meta-Analysis. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*, 25 (3): 364–371. doi:10.1093/NTR/NTAC126.

Zhang, D., Ma, Y., Xu, J., et al. (2022a) Association between obstructive sleep apnea (OSA) and atrial fibrillation (AF): A dose-response meta-analysis. *Medicine (United States)*, 101 (30). doi:10.1097/MD.00000000000029443.

Zhang, F., Mamtani, R., Scott, F.I., et al. (2016) Increasing use of prescription drugs in the United Kingdom. *Pharmacoepidemiology and Drug Safety*, 25 (6). doi:10.1002/pds.3947.

Zhang, F. peng and Zhou, Y. (2013) Analyzing left-truncated and right-censored data under

Cox models with long-term survivors. *Acta Mathematicae Applicatae Sinica*, 29 (2).

doi:10.1007/s10255-013-0215-5.

Zhang, J., Symons, J., Agapow, P., et al. (2022b) Best practices in the real-world data life cycle. *PLOS Digital Health*, 1 (1). doi:10.1371/journal.pdig.0000003.

Zhang, Y., Peloquin, C.E., Dubreuil, M., et al. (2015) Sleep Apnea and the Risk of Incident Gout: A Population-Based, Body Mass Index-Matched Cohort Study. *Arthritis & rheumatology (Hoboken, N.J.)*, 67 (12): 3298–302. doi:10.1002/art.39330.

Zhang, Z., Reinikainen, J., Adeleke, K.A., et al. (2018) Time-varying covariates and coefficients in Cox regression models. *Annals of Translational Medicine*, 6 (7). doi:10.21037/atm.2018.02.12.

Zhao, X. and Guan, J. (2018) Autonomic nervous system might be related with circadian rhythms and have the intricate effects in obstructive sleep apnea with metabolic syndrome. *Journal of clinical hypertension (Greenwich, Conn.)*, 20 (10): 1553. doi:10.1111/JCH.13378.

Zhou, Y., Wang, X., Jiang, Y., et al. (2017) Association between polycystic ovary syndrome and the risk of stroke and all-cause mortality: insights from a meta-analysis. *Gynecological endocrinology: the official journal of the International Society of Gynecological Endocrinology*, 33 (12): 904–910. doi:10.1080/09513590.2017.1347779.

Zhu, S., Zhang, B., Jiang, X., et al. (2019) Metabolic disturbances in non-obese women with polycystic ovary syndrome: a systematic review and meta-analysis. *Fertility and sterility*, 111 (1): 168–177. doi:10.1016/J.FERTNSTERT.2018.09.013.

Chapter 3 : PCOS and NAFLD

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

<https://journals.plos.org/plosmedicine/article/figures?id=10.1371/journal.pmed.1002542>

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.

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^{1,2‡}, Michael W. O'Reilly^{3,4‡}, Konstantinos N. Manolopoulos^{3,4‡}, Konstantinos A. Toulis¹, Krishna M. Gokhale¹, Alice J. Sitch¹, Chandrika N. Wijeyaratne⁵, Arri Coomarasamy^{3,4}, Wiebke Arlt^{3,4}✉*, Krishnarajah Nirantharakumar^{1,3,4}✉*

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 (WA); 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].

Non-alcoholic 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 non-alcoholic 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 (± 1 year) and BMI (± 2 kg/m²) 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/m²) was categorized into <25, 25 to 30, and >30 kg/m² 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

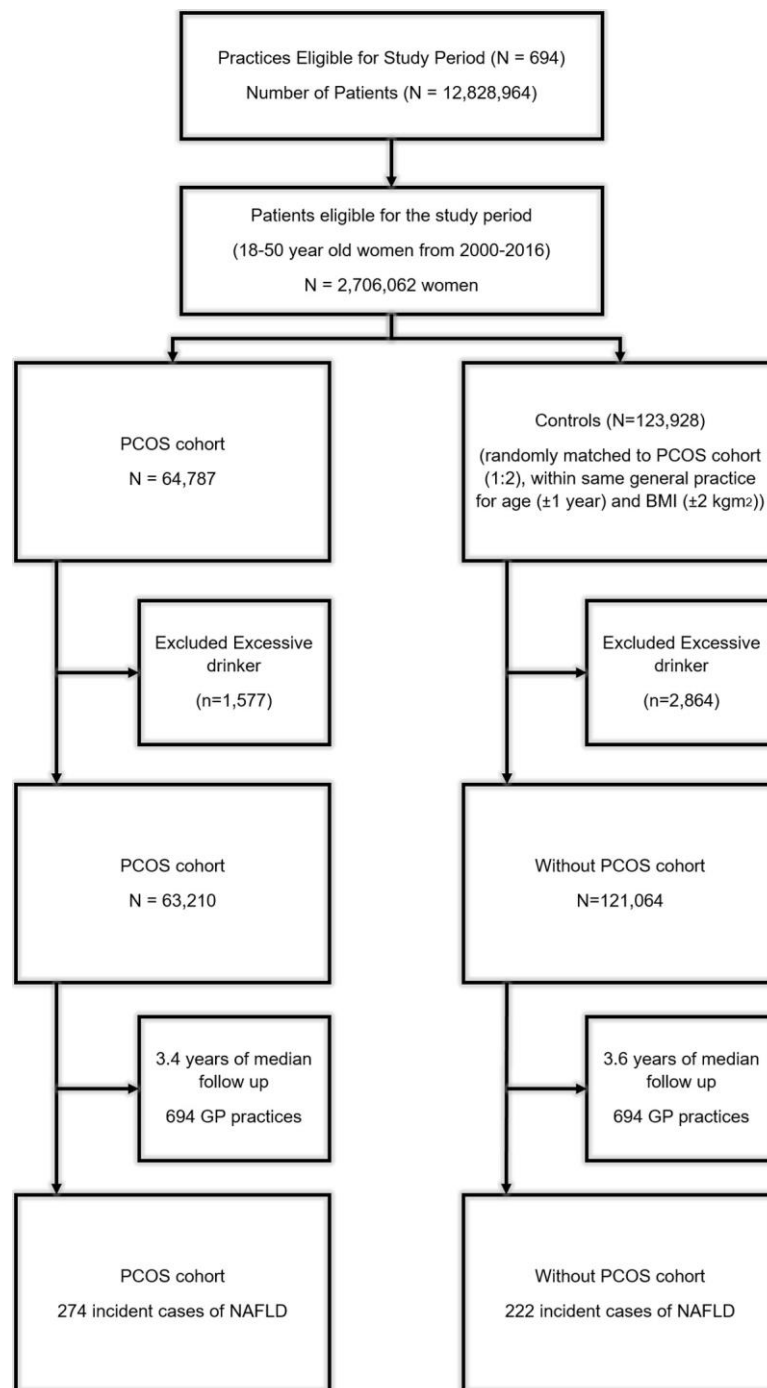


Fig 1. Selection of patients with polycystic ovary syndrome (PCOS) and controls. A total of 694 general practices were eligible for inclusion in the study, with a total of 2,706,062 women for the study period 2000–2016. A total of 64,787 women were listed as having PCOS according to the criteria defined 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 ²) 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 ^ω	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.

^ωIGR, 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 (%)
PCOS features		
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)
Medications		
OCP including an antiandrogenic progestin component ^ω	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.

^ωUse 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) ^ω	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	

^ωAdjusted 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/m², HR = 3.37, 95% CI 2.36–4.83, $p < 0.001$; and BMI >30 kg/m², 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/m²; HR = 2.05, 95% CI 1.37–3.07, $p < 0.001$ for 25–30 kg/m²; and HR = 2.06, 95% CI 1.65–2.57, $p < 0.001$ for >30 kg/m²) (S7 Table). In the analysis with interaction terms for PCOS and BMI, women with PCOS and a BMI of <25 kg/m² 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/m² or >30 kg/m² 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/m² (Fig 2 and S8 Table).

Analysis to explore if androgen excess in women with PCOS confers a risk 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

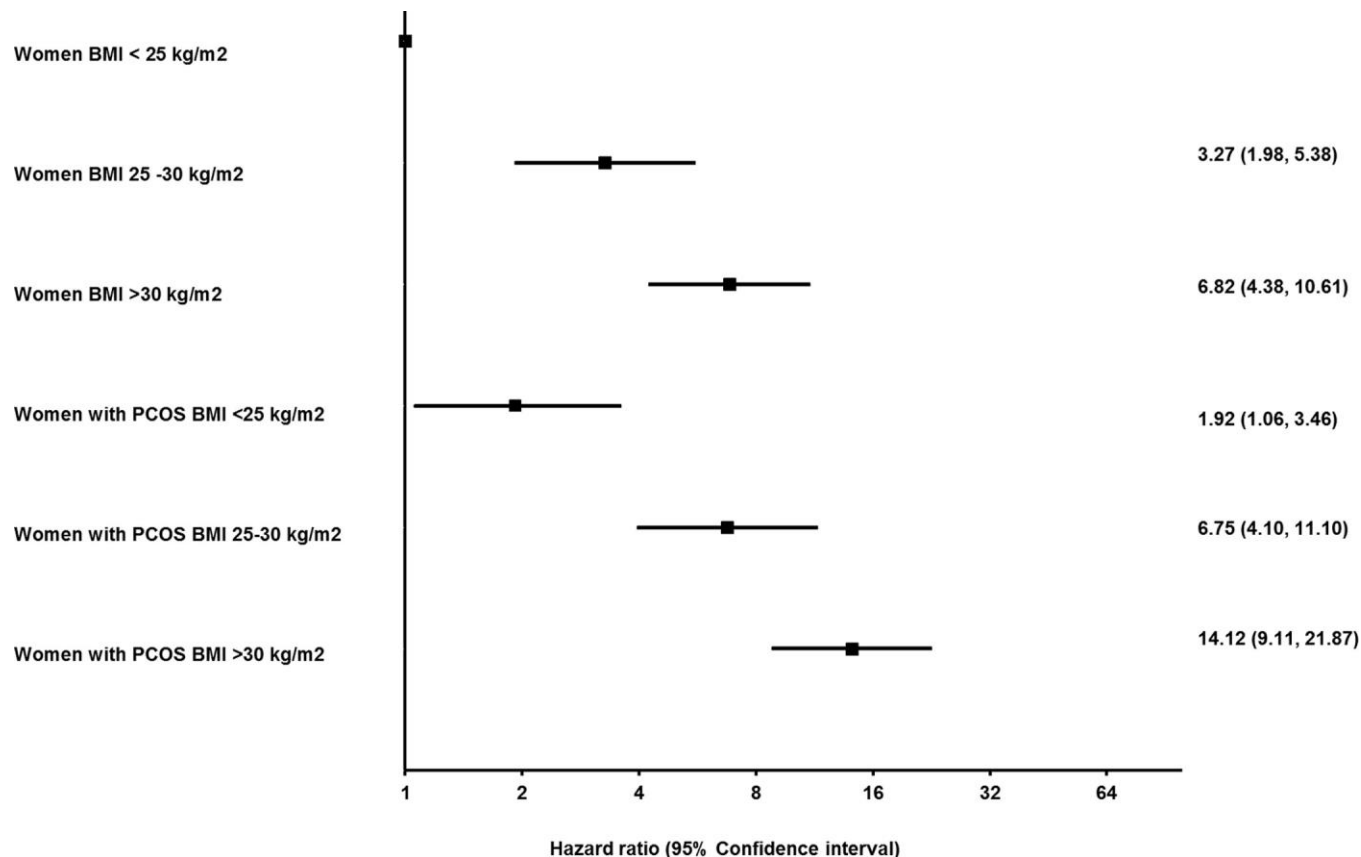


Fig 2. Impact of BMI and polycystic ovary syndrome (PCOS) on the hazard of nonalcoholic fatty liver disease (NAFLD). PCOS and control groups were subcategorized into lean, overweight, and obese groups according to BMI, with lean control women designated as the reference category (unexposed, BMI <25 kg/m²). 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 ²) 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 ^ω	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

^ωIGR, 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.

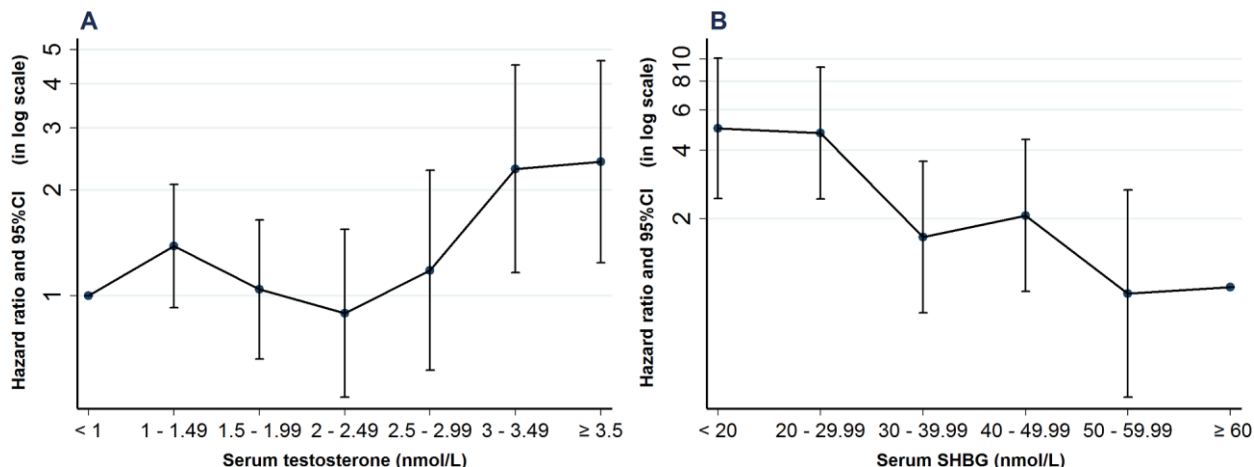


Fig 3. Hazard of nonalcoholic fatty liver disease (NAFLD) according to serum testosterone and sex hormone-binding globulin (SHBG) levels. (A) Serum testosterone (nmol/L) was available in 71,061 unselected women. Using women with testosterone levels below 1.0 nmol/L as a reference cohort, the hazard of 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 ≤ 30 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 ≥ 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 ≥ 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.

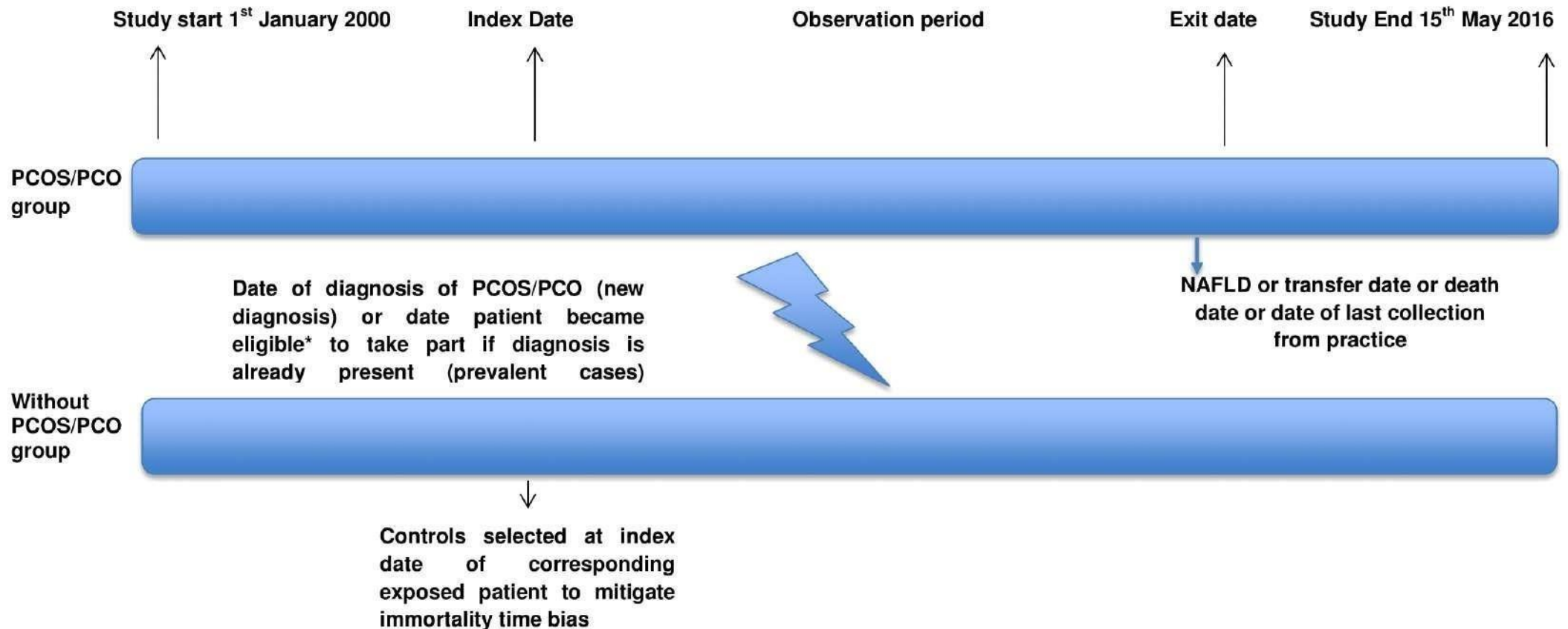
7. O'Reilly MW, Kempegowda P, Jenkinson C, Taylor AE, Quanson JL, Storbeck KH, et al. 11-Oxygenated C19 Steroids Are the Predominant Androgens in Polycystic Ovary Syndrome. *J Clin Endocrinol Metab*. 2017; 102(3):840–8. <https://doi.org/10.1210/jc.2016-3285> PMID: 27901631.
8. Younossi ZM, Koenig AB, 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, Flogel 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. Glinborg 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, Riangwiwat 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.

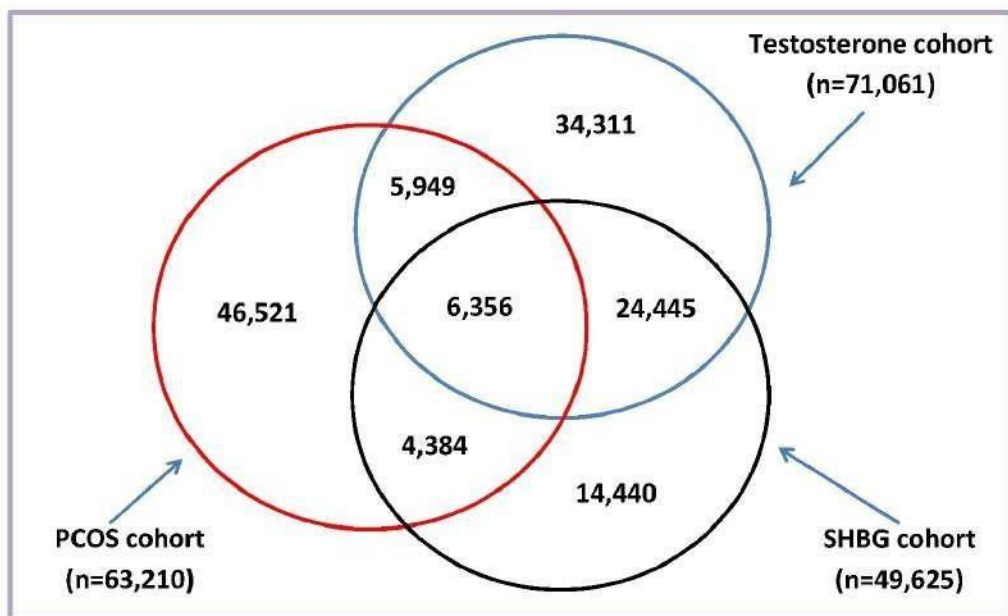
Appendices of Chapter 3 (PCOS and NAFLD)

S1 Fig. Visual representation of study timeline and selection process (Identification of the exposed group (individuals with PCOS/PCO Read codes) and their controls (without PCOS/PCO) matched 1:2 by age at index date, BMI and general practice of origin)

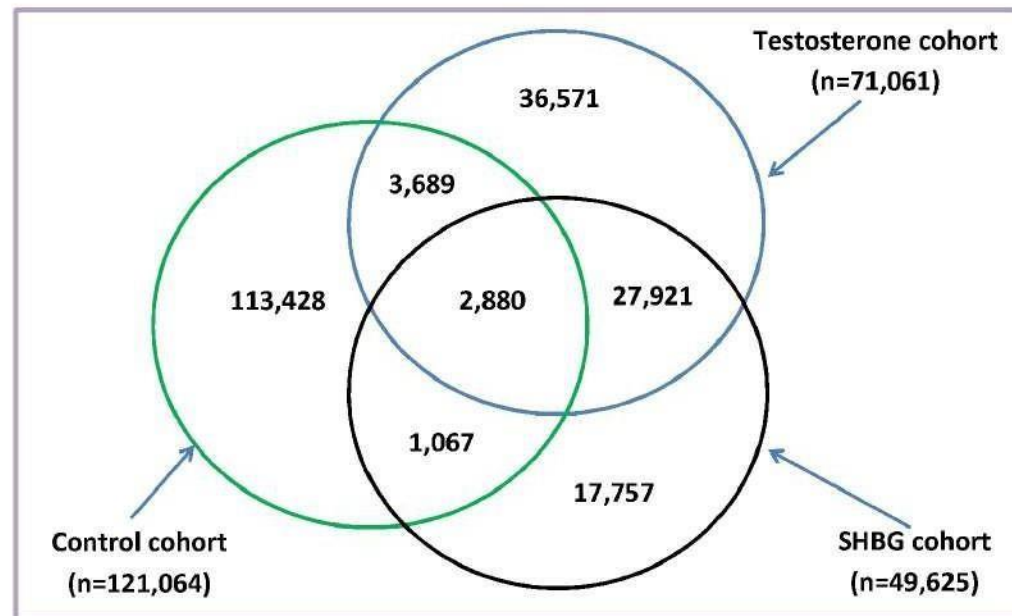


*** To ensure good quality data, a patient is eligible to take part one year after the latest of the following dates: 1) registration in the practice (registration date); 2) introduction of VISION EMR (VISION date); and 3) Acceptable Mortality Recording (AMR) date as an indicator when practices started to record information consistently and in a timely manner. One year latent period is applied to ensure there was sufficient time to record all important covariates.**

A



B



S1: NAFLD among PCOS/PCO women by serum testosterone concentration category (n=11,251)

	Serum testosterone concentration categories (nmol/L)						
	< 1	1 - 1.49	1.5 - 1.99	2 - 2.49	2.5 - 2.99	3 - 3.49	≥ 3.5
Number of participants	1,873	2,174	2,372	1,941	1,266	740	824
Incident NAFLD	9	13	15	5	6	7	6
Person years	10,472	10,805	12,093	10,397	6,684	4,048	4,650
Incidence rate (per 10,000 person years)	8.59	12.03	12.4	4.81	8.98	17.29	12.9
Hazard Ratio	1.0	1.42	1.47	0.57	1.07	2.08	1.52
95% CI of Hazard ratio		(0.61, 3.33)	(0.64, 3.37)	(.19, 1.70)	(0.38, 2.99)	(0.77, 5.59)	(0.54, 4.28)
p-value		0.415	0.359	0.311	0.905	0.147	0.426

S2: NAFLD among PCOS/PCO women by serum SHBG concentration category (n=10,827)

	Serum SHBG concentration categories (nmol/L)					
	<20	20 - 29.99	30 - 39.99	40 - 49.99	50 - 59.99	≥60
Number of participants	1,955	2,497	1,855	1,364	924	2,232
Incident NAFLD	13	15	4	5	1	3
Person years	7,675	9,797	7,261	5,269	3,441	7,753
Incidence rate (per 10,000 person years)	16.94	15.31	5.51	9.49	2.91	3.87
Hazard Ratio	4.43	3.99	1.44	2.45	0.75	1.0
95% CI of Hazard ratio	(1.26, 15.53)	(1.15, 13.78)	(0.32, 6.43)	(0.59, 10.27)	(0.08, 7.23)	-
p-value	0.02	0.03	0.63	0.22	0.80	-

S3: Read codes used in data extraction

Condition	Read Code	Description
Polycystic ovary syndrome	C164.12	Stein - Leventhal syndrome
	C165.00	Polycystic ovarian syndrome
	7E25300	Endoscopic drilling of ovary
Polycystic ovaries	C164.00	Polycystic ovaries
	C164.13	Multi-cystic ovaries
Non- alcoholic fatty liver disease	J61y800	Non-alcoholic steatohepatitis
	J61y100	Non-alcoholic fatty liver
Ovulatory dysfunction	1571	H/O: amenorrhoea
	K590.00	Absence of menstruation
	K590.11	Amenorrhoea
	K590100	Secondary amenorrhoea
	K590z00	Amenorrhoea NOS
	K591.00	Scanty or infrequent menstruation
	K591.11	Infrequent menstruation
	K591000	Hypomenorrhoea
	K591100	Oligomenorrhoea
	K591300	Secondary oligomenorrhoea
	K591z00	Scanty or infrequent menstruation NOS
	K5B0.00	Female infertility of anovulatory origin
	K5B0.11	Anovular cycle
	K5B0100	Secondary anovulatory infertility
	K5B0z00	Female infertility of anovulatory origin NOS
Alopecia	Myu6300	[X]Other androgenic alopecia
	M240.00	Alopecia
	M240z00	Alopecia NOS
	M240H00	Alopecia seborrhoeica
	M240000	Alopecia unspecified
	1N02.00	C/O: hair loss
	M240300	Frontal alopecia of women
	M240012	Hair loss
	M240200	Male pattern alopecia
	M240D00	Marginal alopecia
	22D7.11	O/E – alopecia
	22D4.00	O/E - loss of hair
Hirsutism	22D8.00	O/E – hirsutism
	22D8000	O/E - facial hair
	M241.00	Hirsutism - hypertrichosis

Table S3 (Continued)

Condition	Read codes	Description
Acne	Myu6F00	[X]Acne, unspecified
	2FG5.00	Acne scar
	679g000	Acne management education
	M25y600	Acne keloid
	M261000	Acne vulgaris
	M261100	Acne conglobata
	M261E00	Acne excoriee des jeunes filles
	M261F00	Acne fulminans
	M261H00	Acne keloid
	M261J00	Acne necrotica
	M261K00	Acne keloidalis
	M261X00	Acne, unspecified
	Myu6800	[X]Other acne
	M261.00	Other acne
	M261z00	Other acne NOS
	M261A00	Pustular acne
Diabetes mellitus	C10..00	Diabetes mellitus
	C100.00	Diabetes mellitus with no mention of complication
	C100000	Diabetes mellitus, juvenile type, no mention of complication
	C100011	Insulin dependent diabetes mellitus
	C100100	Diabetes mellitus, adult onset, no mention of complication
	C100111	Maturity onset diabetes
	C100112	Non-insulin dependent diabetes mellitus
	C100z00	Diabetes mellitus NOS with no mention of complication
	C101.00	Diabetes mellitus with ketoacidosis
	C101000	Diabetes mellitus, juvenile type, with ketoacidosis
	C101100	Diabetes mellitus, adult onset, with ketoacidosis
	C101y00	Other specified diabetes mellitus with ketoacidosis
	C101z00	Diabetes mellitus NOS with ketoacidosis
	C102.00	Diabetes mellitus with hyperosmolar coma
	C102000	Diabetes mellitus, juvenile type, with hyperosmolar coma
	C102100	Diabetes mellitus, adult onset, with hyperosmolar coma
	C102z00	Diabetes mellitus NOS with hyperosmolar coma
	C103.00	Diabetes mellitus with ketoacidotic coma
	C103000	Diabetes mellitus, juvenile type, with ketoacidotic coma
	C103100	Diabetes mellitus, adult onset, with ketoacidotic coma
	C103y00	Other specified diabetes mellitus with coma
	C103z00	Diabetes mellitus NOS with ketoacidotic coma
	C104.00	Diabetes mellitus with renal manifestation
	C104.11	Diabetic nephropathy
	C104000	Diabetes mellitus, juvenile type, with renal manifestation
	C104100	Diabetes mellitus, adult onset, with renal manifestation

Condition	Read codes	Description
Diabetes mellitus	C104y00	Other specified diabetes mellitus with renal complications
	C104z00	Diabetes mellitus with nephropathy NOS
	C105.00	Diabetes mellitus with ophthalmic manifestation
	C105000	Diabetes mellitus, juvenile type, + ophthalmic manifestation
	C105100	Diabetes mellitus, adult onset, + ophthalmic manifestation
	C105y00	Other specified diabetes mellitus with ophthalmic complicatn
	C105z00	Diabetes mellitus NOS with ophthalmic manifestation
	C106.00	Diabetes mellitus with neurological manifestation
	C106.11	Diabetic amyotrophy
	C106.12	Diabetes mellitus with neuropathy
	C106.13	Diabetes mellitus with polyneuropathy
	C106000	Diabetes mellitus, juvenile, + neurological manifestation
	C106100	Diabetes mellitus, adult onset, + neurological manifestation
	C106y00	Other specified diabetes mellitus with neurological comps
	C106z00	Diabetes mellitus NOS with neurological manifestation
	C107.00	Diabetes mellitus with peripheral circulatory disorder
	C107.11	Diabetes mellitus with gangrene
	C107.12	Diabetes with gangrene
	C107000	Diabetes mellitus, juvenile +peripheral circulatory disorder
	C107100	Diabetes mellitus, adult, + peripheral circulatory disorder
	C107200	Diabetes mellitus, adult with gangrene
	C107300	IDDM with peripheral circulatory disorder
	C107400	NIDDM with peripheral circulatory disorder
	C107y00	Other specified diabetes mellitus with periph circ comps
	C107z00	Diabetes mellitus NOS with peripheral circulatory disorder
	C108.00	Insulin dependent diabetes mellitus
	C108.11	IDDM-Insulin dependent diabetes mellitus
	C108.12	Type 1 diabetes mellitus
	C108.13	Type I diabetes mellitus
	C108000	Insulin-dependent diabetes mellitus with renal complications
	C108011	Type I diabetes mellitus with renal complications
	C108012	Type 1 diabetes mellitus with renal complications
	C108100	Insulin-dependent diabetes mellitus with ophthalmic comps
	C108111	Type I diabetes mellitus with ophthalmic complications
	C108112	Type 1 diabetes mellitus with ophthalmic complications
	C108200	Insulin-dependent diabetes mellitus with neurological comps
	C108211	Type I diabetes mellitus with neurological complications
	C108212	Type 1 diabetes mellitus with neurological complications
	C108300	Insulin dependent diabetes mellitus with multiple complicatn
	C108311	Type I diabetes mellitus with multiple complications
	C108312	Type 1 diabetes mellitus with multiple complications
	C108400	Unstable insulin dependent diabetes mellitus

Condition	Read codes	Description
Diabetes mellitus	C108411	Unstable type I diabetes mellitus
	C108412	Unstable type 1 diabetes mellitus
	C108500	Insulin dependent diabetes mellitus with ulcer
	C108511	Type I diabetes mellitus with ulcer
	C108512	Type 1 diabetes mellitus with ulcer
	C108600	Insulin dependent diabetes mellitus with gangrene
	C108611	Type I diabetes mellitus with gangrene
	C108612	Type 1 diabetes mellitus with gangrene
	C108700	Insulin dependent diabetes mellitus with retinopathy
	C108711	Type I diabetes mellitus with retinopathy
	C108712	Type 1 diabetes mellitus with retinopathy
	C108800	Insulin dependent diabetes mellitus - poor control
	C108811	Type I diabetes mellitus - poor control
	C108812	Type 1 diabetes mellitus - poor control
	C108900	Insulin dependent diabetes maturity onset
	C108911	Type I diabetes mellitus maturity onset
	C108912	Type 1 diabetes mellitus maturity onset
	C108A00	Insulin-dependent diabetes without complication
	C108A11	Type I diabetes mellitus without complication
	C108A12	Type 1 diabetes mellitus without complication
	C108B00	Insulin dependent diabetes mellitus with mononeuropathy
	C108B11	Type I diabetes mellitus with mononeuropathy
	C108B12	Type 1 diabetes mellitus with mononeuropathy
	C108C00	Insulin dependent diabetes mellitus with polyneuropathy
	C108C11	Type I diabetes mellitus with polyneuropathy
	C108C12	Type 1 diabetes mellitus with polyneuropathy
	C108D00	Insulin dependent diabetes mellitus with nephropathy
	C108D11	Type I diabetes mellitus with nephropathy
	C108D12	Type 1 diabetes mellitus with nephropathy
	C108E00	Insulin dependent diabetes mellitus with hypoglycaemic coma
	C108E11	Type I diabetes mellitus with hypoglycaemic coma
	C108E12	Type 1 diabetes mellitus with hypoglycaemic coma
	C108F00	Insulin dependent diabetes mellitus with diabetic cataract
	C108F11	Type I diabetes mellitus with diabetic cataract
	C108F12	Type 1 diabetes mellitus with diabetic cataract
	C108G00	Insulin dependent diab mell with peripheral angiopathy
	C108G11	Type I diabetes mellitus with peripheral angiopathy
	C108G12	Type 1 diabetes mellitus with peripheral angiopathy
	C108H00	Insulin dependent diabetes mellitus with arthropathy
	C108H11	Type I diabetes mellitus with arthropathy
	C108H12	Type 1 diabetes mellitus with arthropathy
	C108J00	Insulin dependent diab mell with neuropathic arthropathy

Condition	Read codes	Description
Diabetes mellitus	C108J11	Type I diabetes mellitus with neuropathic arthropathy
	C108J12	Type 1 diabetes mellitus with neuropathic arthropathy
	C108y00	Other specified diabetes mellitus with multiple comps
	C108z00	Unspecified diabetes mellitus with multiple complications
	C109.00	Non-insulin dependent diabetes mellitus
	C109.11	NIDDM - Non-insulin dependent diabetes mellitus
	C109.12	Type 2 diabetes mellitus
	C109.13	Type II diabetes mellitus
	C109000	Non-insulin-dependent diabetes mellitus with renal comps
	C109011	Type II diabetes mellitus with renal complications
	C109012	Type 2 diabetes mellitus with renal complications
	C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps
	C109111	Type II diabetes mellitus with ophthalmic complications
	C109112	Type 2 diabetes mellitus with ophthalmic complications
	C109200	Non-insulin-dependent diabetes mellitus with neuro comps
	C109211	Type II diabetes mellitus with neurological complications
	C109212	Type 2 diabetes mellitus with neurological complications
	C109300	Non-insulin-dependent diabetes mellitus with multiple comps
	C109311	Type II diabetes mellitus with multiple complications
	C109312	Type 2 diabetes mellitus with multiple complications
	C109400	Non-insulin dependent diabetes mellitus with ulcer
	C109411	Type II diabetes mellitus with ulcer
	C109412	Type 2 diabetes mellitus with ulcer
	C109500	Non-insulin dependent diabetes mellitus with gangrene
	C109511	Type II diabetes mellitus with gangrene
	C109512	Type 2 diabetes mellitus with gangrene
	C109600	Non-insulin-dependent diabetes mellitus with retinopathy
	C109611	Type II diabetes mellitus with retinopathy
	C109612	Type 2 diabetes mellitus with retinopathy
	C109700	Non-insulin dependent diabetes mellitus - poor control
	C109711	Type II diabetes mellitus - poor control
	C109712	Type 2 diabetes mellitus - poor control
	C109800	Reaven's syndrome
	C109900	Non-insulin-dependent diabetes mellitus without complication
	C109911	Type II diabetes mellitus without complication
	C109912	Type 2 diabetes mellitus without complication
	C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy
	C109A11	Type II diabetes mellitus with mononeuropathy
	C109A12	Type 2 diabetes mellitus with mononeuropathy
	C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy
	C109B11	Type II diabetes mellitus with polyneuropathy
	C109B12	Type 2 diabetes mellitus with polyneuropathy

Condition	Read codes	Description
Diabetes mellitus	C109C00	Non-insulin dependent diabetes mellitus with nephropathy
	C109C11	Type II diabetes mellitus with nephropathy
	C109C12	Type 2 diabetes mellitus with nephropathy
	C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma
	C109D11	Type II diabetes mellitus with hypoglycaemic coma
	C109D12	Type 2 diabetes mellitus with hypoglycaemic coma
	C109E00	Non-insulin depend diabetes mellitus with diabetic cataract
	C109E11	Type II diabetes mellitus with diabetic cataract
	C109E12	Type 2 diabetes mellitus with diabetic cataract
	C109F00	Non-insulin-dependent d m with peripheral angiopath
	C109F11	Type II diabetes mellitus with peripheral angiopathy
	C109F12	Type 2 diabetes mellitus with peripheral angiopathy
	C109G00	Non-insulin dependent diabetes mellitus with arthropathy
	C109G11	Type II diabetes mellitus with arthropathy
	C109G12	Type 2 diabetes mellitus with arthropathy
	C109H00	Non-insulin dependent d m with neuropathic arthropathy
	C109H11	Type II diabetes mellitus with neuropathic arthropathy
	C109H12	Type 2 diabetes mellitus with neuropathic arthropathy
	C109J00	Insulin treated Type 2 diabetes mellitus
	C109J11	Insulin treated non-insulin dependent diabetes mellitus
	C109J12	Insulin treated Type II diabetes mellitus
	C109K00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
	C10A.00	Malnutrition-related diabetes mellitus
	C10A.11	Jamaica type diabetes
	C10A000	Malnutrition-related diabetes mellitus with coma
	C10A100	Malnutrition-related diabetes mellitus with ketoacidosis
	C10A200	Malnutrition-related diabetes mellitus with renal complicatn
	C10A300	Malnutrit-related diabetes mellitus wth ophthalmic complicat
	C10A400	Malnutrition-related diabetes mellitus wth neuro complicatns
	C10A500	Malnutritn-relat diabetes melitus wth periph circul complctn
	C10A600	Malnutrition-related diabetes mellitus with multiple comps
	C10A700	Malnutrition-related diabetes mellitus without complications
	C10AW00	Malnutrit-related diabetes mellitus with unspec complica
	C10AX00	Malnutrit-relat diabetes mellitus with other spec comps
	C10B.00	Diabetes mellitus induced by steroids
	C10B000	Steroid induced diabetes mellitus without complication
	C10C.00	Diabetes mellitus autosomal dominant
	C10C.11	Maturity onset diabetes in youth
	C10C.12	Maturity onset diabetes in youth type 1
	C10D.00	Diabetes mellitus autosomal dominant type 2
	C10D.11	Maturity onset diabetes in youth type 2
	C10E.00	Type 1 diabetes mellitus

Condition	Read codes	Description
Diabetes mellitus	C10E.11	Type I diabetes mellitus
	C10E.12	Insulin dependent diabetes mellitus
	C10E000	Type 1 diabetes mellitus with renal complications
	C10E011	Type I diabetes mellitus with renal complications
	C10E012	Insulin-dependent diabetes mellitus with renal complications
	C10E100	Type 1 diabetes mellitus with ophthalmic complications
	C10E111	Type I diabetes mellitus with ophthalmic complications
	C10E112	Insulin-dependent diabetes mellitus with ophthalmic comps
	C10E200	Type 1 diabetes mellitus with neurological complications
	C10E211	Type I diabetes mellitus with neurological complications
	C10E212	Insulin-dependent diabetes mellitus with neurological comps
	C10E300	Type 1 diabetes mellitus with multiple complications
	C10E311	Type I diabetes mellitus with multiple complications
	C10E312	Insulin dependent diabetes mellitus with multiple complicat
	C10E400	Unstable type 1 diabetes mellitus
	C10E411	Unstable type I diabetes mellitus
	C10E412	Unstable insulin dependent diabetes mellitus
	C10E500	Type 1 diabetes mellitus with ulcer
	C10E511	Type I diabetes mellitus with ulcer
	C10E512	Insulin dependent diabetes mellitus with ulcer
	C10E600	Type 1 diabetes mellitus with gangrene
	C10E611	Type I diabetes mellitus with gangrene
	C10E612	Insulin dependent diabetes mellitus with gangrene
	C10E700	Type 1 diabetes mellitus with retinopathy
	C10E711	Type I diabetes mellitus with retinopathy
	C10E712	Insulin dependent diabetes mellitus with retinopathy
	C10E800	Type 1 diabetes mellitus - poor control
	C10E811	Type I diabetes mellitus - poor control
	C10E812	Insulin dependent diabetes mellitus - poor control
	C10E900	Type 1 diabetes mellitus maturity onset
	C10E911	Type I diabetes mellitus maturity onset
	C10E912	Insulin dependent diabetes maturity onset
	C10EA00	Type 1 diabetes mellitus without complication
	C10EA11	Type I diabetes mellitus without complication
	C10EA12	Insulin-dependent diabetes without complication
	C10EB00	Type 1 diabetes mellitus with mononeuropathy
	C10EB11	Type I diabetes mellitus with mononeuropathy
	C10EB12	Insulin dependent diabetes mellitus with mononeuropathy
	C10EC00	Type 1 diabetes mellitus with polyneuropathy
	C10EC11	Type I diabetes mellitus with polyneuropathy
	C10EC12	Insulin dependent diabetes mellitus with polyneuropathy
	C10ED00	Type 1 diabetes mellitus with nephropathy

Condition	Read codes	Description
Diabetes mellitus	C10ED11	Type I diabetes mellitus with nephropathy
	C10ED12	Insulin dependent diabetes mellitus with nephropathy
	C10EE00	Type 1 diabetes mellitus with hypoglycaemic coma
	C10EE11	Type I diabetes mellitus with hypoglycaemic coma
	C10EE12	Insulin dependent diabetes mellitus with hypoglycaemic coma
	C10EF00	Type 1 diabetes mellitus with diabetic cataract
	C10EF11	Type I diabetes mellitus with diabetic cataract
	C10EF12	Insulin dependent diabetes mellitus with diabetic cataract
	C10EG00	Type 1 diabetes mellitus with peripheral angiopathy
	C10EG11	Type I diabetes mellitus with peripheral angiopathy
	C10EG12	Insulin dependent diab mell with peripheral angiopathy
	C10EH00	Type 1 diabetes mellitus with arthropathy
	C10EH11	Type I diabetes mellitus with arthropathy
	C10EH12	Insulin dependent diabetes mellitus with arthropathy
	C10EJ00	Type 1 diabetes mellitus with neuropathic arthropathy
	C10EJ11	Type I diabetes mellitus with neuropathic arthropathy
	C10EJ12	Insulin dependent diab mell with neuropathic arthropathy
	C10EK00	Type 1 diabetes mellitus with persistent proteinuria
	C10EK11	Type I diabetes mellitus with persistent proteinuria
	C10EL00	Type 1 diabetes mellitus with persistent microalbuminuria
	C10EL11	Type I diabetes mellitus with persistent microalbuminuria
	C10EM00	Type 1 diabetes mellitus with ketoacidosis
	C10EM11	Type I diabetes mellitus with ketoacidosis
	C10EN00	Type 1 diabetes mellitus with ketoacidotic coma
	C10EN11	Type I diabetes mellitus with ketoacidotic coma
	C10EP00	Type 1 diabetes mellitus with exudative maculopathy
	C10EP11	Type I diabetes mellitus with exudative maculopathy
	C10EQ00	Type 1 diabetes mellitus with gastroparesis
	C10EQ11	Type I diabetes mellitus with gastroparesis
	C10ER00	Latent autoimmune diabetes mellitus in adult
	C10F.00	Type 2 diabetes mellitus
	C10F.11	Type II diabetes mellitus
	C10F000	Type 2 diabetes mellitus with renal complications
	C10F011	Type II diabetes mellitus with renal complications
	C10F100	Type 2 diabetes mellitus with ophthalmic complications
	C10F111	Type II diabetes mellitus with ophthalmic complications
	C10F200	Type 2 diabetes mellitus with neurological complications
	C10F211	Type II diabetes mellitus with neurological complications
	C10F300	Type 2 diabetes mellitus with multiple complications
	C10F311	Type II diabetes mellitus with multiple complications
	C10F400	Type 2 diabetes mellitus with ulcer
	C10F411	Type II diabetes mellitus with ulcer

Condition	Read codes	Description
Diabetes mellitus	C10F500	Type 2 diabetes mellitus with gangrene
	C10F511	Type II diabetes mellitus with gangrene
	C10F600	Type 2 diabetes mellitus with retinopathy
	C10F611	Type II diabetes mellitus with retinopathy
	C10F700	Type 2 diabetes mellitus - poor control
	C10F711	Type II diabetes mellitus - poor control
	C10F800	Reaven's syndrome
	C10F811	Metabolic syndrome X
	C10F900	Type 2 diabetes mellitus without complication
	C10F911	Type II diabetes mellitus without complication
	C10FA00	Type 2 diabetes mellitus with mononeuropathy
	C10FA11	Type II diabetes mellitus with mononeuropathy
	C10FB00	Type 2 diabetes mellitus with polyneuropathy
	C10FB11	Type II diabetes mellitus with polyneuropathy
	C10FC00	Type 2 diabetes mellitus with nephropathy
	C10FC11	Type II diabetes mellitus with nephropathy
	C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma
	C10FD11	Type II diabetes mellitus with hypoglycaemic coma
	C10FE00	Type 2 diabetes mellitus with diabetic cataract
	C10FE11	Type II diabetes mellitus with diabetic cataract
	C10FF00	Type 2 diabetes mellitus with peripheral angiopathy
	C10FF11	Type II diabetes mellitus with peripheral angiopathy
	C10FG00	Type 2 diabetes mellitus with arthropathy
	C10FG11	Type II diabetes mellitus with arthropathy
	C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy
	C10FH11	Type II diabetes mellitus with neuropathic arthropathy
	C10FJ00	Insulin treated Type 2 diabetes mellitus
	C10FJ11	Insulin treated Type II diabetes mellitus
	C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
	C10FK11	Hyperosmolar non-ketotic state in type II diabetes mellitus
	C10FL00	Type 2 diabetes mellitus with persistent proteinuria
	C10FL11	Type II diabetes mellitus with persistent proteinuria
	C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria
	C10FM11	Type II diabetes mellitus with persistent microalbuminuria
	C10FN00	Type 2 diabetes mellitus with ketoacidosis
	C10FN11	Type II diabetes mellitus with ketoacidosis
	C10FP00	Type 2 diabetes mellitus with ketoacidotic coma
	C10FP11	Type II diabetes mellitus with ketoacidotic coma
	C10FQ00	Type 2 diabetes mellitus with exudative maculopathy
	C10FQ11	Type II diabetes mellitus with exudative maculopathy
	C10FR00	Type 2 diabetes mellitus with gastroparesis
	C10FR11	Type II diabetes mellitus with gastroparesis

Condition	Read codes	Description
Diabetes mellitus	C10FS00	Maternally inherited diabetes mellitus
	C10G.00	Secondary pancreatic diabetes mellitus
	C10G000	Secondary pancreatic diabetes mellitus without complication
	C10H.00	Diabetes mellitus induced by non-steroid drugs
	C10H000	DM induced by non-steroid drugs without complication
	C10J.00	Insulin autoimmune syndrome
	C10J000	Insulin autoimmune syndrome without complication
	C10K.00	Type A insulin resistance
	C10K000	Type A insulin resistance without complication
	C10L.00	Fibrocalculous pancreatopathy
	C10L000	Fibrocalculous pancreatopathy without complication
	C10M.00	Lipoatrophic diabetes mellitus
	C10M000	Lipoatrophic diabetes mellitus without complication
	C10N.00	Secondary diabetes mellitus
	C10N000	Secondary diabetes mellitus without complication
	C10N100	Cystic fibrosis related diabetes mellitus
	C10P.00	Diabetes mellitus in remission
	C10P000	Type I diabetes mellitus in remission
	C10P011	Type 1 diabetes mellitus in remission
	C10P100	Type II diabetes mellitus in remission
	C10P111	Type 2 diabetes mellitus in remission
	C10y.00	Diabetes mellitus with other specified manifestation
	C10y000	Diabetes mellitus, juvenile, + other specified manifestation
	C10y100	Diabetes mellitus, adult, + other specified manifestation
	C10yy00	Other specified diabetes mellitus with other spec comps
	C10yz00	Diabetes mellitus NOS with other specified manifestation
	C10z.00	Diabetes mellitus with unspecified complication
	C10z000	Diabetes mellitus, juvenile type, + unspecified complication
	C10z100	Diabetes mellitus, adult onset, + unspecified complication
	C10zy00	Other specified diabetes mellitus with unspecified comps
	C10zz00	Diabetes mellitus NOS with unspecified complication
Hypertension	G2...00	Hypertensive disease
	G2...11	BP - hypertensive disease
	G20..00	Essential hypertension
	G20..11	High blood pressure
	G20..12	Primary hypertension
	G200.00	Malignant essential hypertension
	G201.00	Benign essential hypertension
	G202.00	Systolic hypertension
	G203.00	Diastolic hypertension
	G20z.00	Essential hypertension NOS

Condition	Read codes	Description
Hypertension	G20z.11	Hypertension NOS
	G21..00	Hypertensive heart disease
	G210.00	Malignant hypertensive heart disease
	G210000	Malignant hypertensive heart disease without CCF
	G210100	Malignant hypertensive heart disease with CCF
	G210z00	Malignant hypertensive heart disease NOS
	G211.00	Benign hypertensive heart disease
	G211000	Benign hypertensive heart disease without CCF
	G211100	Benign hypertensive heart disease with CCF
	G211z00	Benign hypertensive heart disease NOS
	G21z.00	Hypertensive heart disease NOS
	G21z000	Hypertensive heart disease NOS without CCF
	G21z011	Cardiomegaly - hypertensive
	G21z100	Hypertensive heart disease NOS with CCF
	G21zz00	Hypertensive heart disease NOS
	G22..00	Hypertensive renal disease
	G22..11	Nephrosclerosis
	G220.00	Malignant hypertensive renal disease
	G221.00	Benign hypertensive renal disease
	G222.00	Hypertensive renal disease with renal failure
	G22z.00	Hypertensive renal disease NOS
	G22z.11	Renal hypertension
	G23..00	Hypertensive heart and renal disease
	G230.00	Malignant hypertensive heart and renal disease
	G231.00	Benign hypertensive heart and renal disease
	G232.00	Hypertensive heart&renal dis wth (congestive) heart failure
	G233.00	Hypertensive heart and renal disease with renal failure
	G234.00	Hyperten heart&renal dis+both(congestv)heart and renal fail
	G23z.00	Hypertensive heart and renal disease NOS
	G24..00	Secondary hypertension
	G240.00	Secondary malignant hypertension
	G240000	Secondary malignant renovascular hypertension
	G240z00	Secondary malignant hypertension NOS
	G241.00	Secondary benign hypertension
	G241000	Secondary benign renovascular hypertension
	G241z00	Secondary benign hypertension NOS
	G244.00	Hypertension secondary to endocrine disorders
	G24z.00	Secondary hypertension NOS
	G24z000	Secondary renovascular hypertension NOS
	G24z100	Hypertension secondary to drug
	G24zz00	Secondary hypertension NOS
	G25..00	Stage 1 hypertension (NICE - Nat Ins for Hth Clin Excl 2011)

Condition	Read codes	Description
Hypertension	G25..11	Stage 1 hypertension
	G250.00	Stage 1 hyperten (NICE 2011) without evidnce end organ damage
	G251.00	Stage 1 hyperten (NICE 2011) with evidnce end organ damage
	G26..00	Severe hypertension (Nat Inst for Health Clinical Ex 2011)
	G26..11	Severe hypertension
	G27..00	Hypertension resistant to drug therapy
	G28..00	Stage 2 hypertension (NICE - Nat Ins for Hth Clin Excl 2011)
	G2y..00	Other specified hypertensive disease
	G2z..00	Hypertensive disease NOS
	Gyu2.00	[X]Hypertensive diseases
	Gyu2000	[X]Other secondary hypertension
	Gyu2100	[X]Hypertension secondary to other renal disorders
Hypothyroidism	C03..00	Congenital hypothyroidism
	C03..11	Cretinism
	C030.00	Pendred's syndrome
	C031.00	Goitrous cretin
	C03y.00	Other specified congenital hypothyroidism
	C03y000	Congenital hypothyroidism with diffuse goitre
	C03y100	Congenital hypothyroidism without goitre
	C03z.00	Congenital hypothyroidism NOS
	C03z.11	Congenital thyroid insufficiency
	C03z.12	Cretinism
	C04..00	Acquired hypothyroidism
	C04..11	Myxoedema
	C04..12	Thyroid deficiency
	C04..13	Hypothyroidism
	C040.00	Postsurgical hypothyroidism
	C040.11	Post ablative hypothyroidism
	C041.00	Other postablative hypothyroidism
	C041000	Irradiation hypothyroidism
	C041z00	Postablative hypothyroidism NOS
	C042.00	Iodine hypothyroidism
	C043.00	Other iatrogenic hypothyroidism
	C043000	Hypothyroidism resulting from para-aminosalicylic acid
	C043100	Hypothyroidism resulting from phenylbutazone
	C043200	Hypothyroidism resulting from resorcinol
	C043z00	Iatrogenic hypothyroidism NOS
	C044.00	Postinfectious hypothyroidism
	C045.00	Acquired atrophy of thyroid
	C046.00	Autoimmune myxoedema
	C047.00	Subclinical hypothyroidism
	C04y.00	Other acquired hypothyroidism

Condition	Read codes	Description
Hypothyroidism	C04z.00	Hypothyroidism NOS
	C04z.11	Pretibial myxoedema - hypothyroid
	C04z.12	Thyroid insufficiency
	C04z.13	Hypothyroid goitre, acquired
	C04z000	Premature puberty due to hypothyroidism
	C04z100	Myxoedema coma
	C0A5.00	Subclinical iodine-deficiency hypothyroidism
	Cyu1100	[X]Other specified hypothyroidism
	F381400	Myasthenic syndrome due to hypothyroidism
	Q433700	Neonatal jaundice with congenital hypothyroidism
Impaired glucose regulation	44V2.00	Glucose tolerance. test impaired
	9mX..00	Impaired glucose regulation monitoring invitation
	9mX..00	Impaired glucose regulation monitoring invitation
	9mX1.00	Impaired glucose regulation monitoring invitation 2nd letter
	9mX1.00	Impaired glucose regulation monitoring invitation 2nd letter
	9mX2.00	Impaired glucose regulation monitoring invitation 3rd letter
	9mX2.00	Impaired glucose regulation monitoring invitation 3rd letter
	9mX3.00	Impaired glucose regulation monitoring telephone invitation
	9mX3.00	Impaired glucose regulation monitoring telephone invitation
	9mX4.00	Impaired glucose regulation monitoring verbal invitation
	9mX4.00	Impaired glucose regulation monitoring verbal invitation
	9NS0400	Referral for impaired glucose tolerance management offered
	C11y200	Impaired glucose tolerance
	C11y300	Impaired fasting glycaemia
	C11y400	Impaired glucose regulation
	C11y500	Pre-diabetes
	R102.12	[D]Impaired glucose tolerance test
	R10D000	[D]Impaired fasting glycaemia
	R10D011	[D]Impaired fasting glucose
	R10E.00	[D]Impaired glucose tolerance
	R102.11	[D]Prediabetes
	C313500	Glucose intolerance

S4: Drug codes used to identify antiandrogens, lipid modifying drugs and metformin

Drug	Drug code	Preparation
Cyproterone	85864998	Co-cyprindiol 2000microgram/35microgram tablets
	86466998	Co-cyprindiol 2000microgram/35microgram tablets
	86925998	Co-cyprindiol 2mg+35mcg tablets
	87351998	Co-cyprindiol 2000microgram/35microgram tablets
	90826979	Co-cyprindiol 2000microgram/35microgram tablets
	90833979	Co-cyprindiol 2000microgram/35microgram tablets
	91068998	Co-cyprindiol 2000microgram/35microgram tablets
	91069998	Co-cyprindiol 2000microgram/35microgram tablets
	94832990	Co-cyprindiol 2000microgram/35microgram tablets
	94920998	Co-cyprindiol 2000microgram/35microgram tablets
	95220990	Co-cyprindiol 2000microgram/35microgram tablets
	95396990	Co-cyprindiol 2000microgram/35microgram tablets
	96577998	Co-cyprindiol 2000microgram/35microgram tablets
	97520998	Co-cyprindiol 2000microgram/35microgram tablets
	81667998	Cyproterone 100mg tablets
	81668998	Cyproterone 50mg tablets
	92777990	Cyproterone 50mg tablets
	96269989	Cyproterone 100mg tablets
	96269990	Cyproterone 50mg tablets
	96578997	Cyproterone 100mg tablets
	96578998	Cyproterone 50mg tablets
	97768990	Cyproterone 50mg tablets
	98017989	Cyproterone 100mg tablets
	98017990	Cyproterone 50mg tablets
	98246990	Cyproterone 50mg tablets
	99797997	Cyproterone 100mg tablets
	99797998	Cyproterone 50mg tablets
	99958998	Cyproterone 50mg tablets
Drospirenone	53008979	Ethinylestradiol 20microgram / Drospirenone 3mg tablets
	59254978	Ethinylestradiol 30microgram / Drospirenone 3mg tablets
	74455978	Ethinylestradiol 30microgram / Drospirenone 3mg tablets
	78546978	Ethinylestradiol 30microgram / Drospirenone 3mg tablets
	84583978	Ethinylestradiol 20microgram / Drospirenone 3mg tablets
	86831998	Estradiol 1mg / Drospirenone 2mg tablets
	86832998	Estradiol 1mg / Drospirenone 2mg tablets
	89914979	Ethinylestradiol 30microgram / Drospirenone 3mg tablets
	92571998	Ethinylestradiol 30microgram / Drospirenone 3mg tablets
	98852998	Ethinylestradiol 30microgram / Drospirenone 3mg tablets

Table S4: (Continued)

Drug	Drug code	Preparation
other antiandrogen drugs	82882978	Enzalutamide 40mg capsules
	82127998	Tamsulosin 400microgram / Dutasteride 500microgram capsules
	82128998	Tamsulosin 400microgram / Dutasteride 500microgram capsules
	85997979	Dutasteride 500microgram capsules
	89268998	Dutasteride 500microgram capsules
	89947998	Dutasteride 500microgram capsules
	88336998	Flutamide 250mg tablets
	92415979	Flutamide 250mg tablets
	93504998	Flutamide 250mg tablets
	93505998	Flutamide 250mg tablets
	97211990	Flutamide 250mg tablets
	97862990	Flutamide 250mg tablets
	97867990	Flutamide 250mg tablets
	59259978	Finasteride 1mg tablets
	63181979	Finasteride 5mg/5ml oral suspension
	89063998	Finasteride 1mg tablets
	89990979	Finasteride 1mg tablets
	91684998	Finasteride 1mg tablets
	93144979	Finasteride 5mg tablets
	99249998	Finasteride 5mg tablets
	99250998	Finasteride 5mg tablets
metformin	85622998	Pioglitazone 15mg / Metformin 850mg tablets
	85625998	Pioglitazone 15mg / Metformin 850mg tablets
	85624998	Pioglitazone 15mg / Metformin 850mg tablets
	53006979	Saxagliptin 2.5mg / Metformin 1g tablets
	76993978	Dapagliflozin 5mg / Metformin 1g tablets
	76992978	Dapagliflozin 5mg / Metformin 1g tablets
	76991978	Dapagliflozin 5mg / Metformin 850mg tablets
	78730978	Alogliptin 12.5mg / Metformin 1g tablets
	83401998	Metformin 1g / Sitagliptin 50mg tablets
	78729978	Alogliptin 12.5mg / Metformin 1g tablets
	84008998	Vildagliptin 50mg / Metformin 1g tablets
	84011998	Vildagliptin 50mg / Metformin 850mg tablets
	84009998	Vildagliptin 50mg / Metformin 850mg tablets
	84010998	Vildagliptin 50mg / Metformin 1g tablets
	82068998	Metformin 1g / Sitagliptin 50mg tablets
	53004979	Saxagliptin 2.5mg / Metformin 850mg tablets
	53005979	Saxagliptin 2.5mg / Metformin 850mg tablets
	53007979	Saxagliptin 2.5mg / Metformin 1g tablets
	54904979	Linagliptin 2.5mg / Metformin 850mg tablets
	54907979	Linagliptin 2.5mg / Metformin 1g tablets
	54905979	Linagliptin 2.5mg / Metformin 850mg tablets

Drug	Drug code	Preparation
Metformin	54906979	Linagliptin 2.5mg / Metformin 1g tablets
	82919998	Metformin 500mg oral powder sachets sugar free
	82918998	Metformin 1g oral powder sachets sugar free
	82917998	Metformin hydrochloride 500mg sachets
	82916998	Metformin hydrochloride 1g sachets
	81344998	Metformin 1g modified-release tablets
	83733998	Metformin 750mg modified-release tablets
	83031998	Metformin 1g modified-release tablets
	83732998	Metformin 750mg modified-release tablets
	79512979	Metformin 500mg/5ml oral solution
	83619998	Metformin 500mg modified-release tablets
	81158998	Metformin 500mg modified-release tablets
	81701998	Metformin 500mg modified-release tablets
	83032998	Metformin 1g modified-release tablets
	85673998	Metformin 500mg/5ml oral solution sugar free
	85674998	Metformin 500mg/5ml oral solution
	89129979	Metformin 1g modified-release tablets
	89870979	Metformin 500mg modified-release tablets
	87053998	Metformin 500mg modified-release tablets
	87536998	Metformin 500mg/5ml oral solution
	87054998	Metformin 500mg modified-release tablets
	89868979	Metformin 500mg modified-release tablets
	87883998	Metformin 500mg modified-release tablets
	89128979	Metformin 1g modified-release tablets
	89188979	Metformin 500mg/5ml oral solution
	91221998	Metformin hydrochloride 500mg tablets
	93167990	Metformin 500mg/5ml oral solution
	91221997	Metformin hydrochloride 850mg tablets
	93469979	Metformin 500mg tablets
	92983990	Metformin 500mg/5ml oral solution
	94235992	Metformin 500mg tablets
	94248990	Metformin 500mg tablets
	95271992	Metformin 850mg tablets
	95270992	Metformin 800 mg tab
	96111990	Metformin 500mg tablets
	95880998	Metformin 500mg modified-release tablets
	96270990	Metformin 500mg tablets
	97087998	Metformin 500mg tablets
	97110990	Metformin 500mg tablets
	97110989	Metformin 850mg tablets
	97087997	Metformin 850mg tablets
	95272992	Metformin 250 mg tab

Drug	Drug code	Preparation
Metformin	96850990	Metformin 500mg tablets
	95600990	Metformin 500mg tablets
	99514990	Metformin 500mg tablets
	98125990	Metformin 500mg tablets
	99149990	Metformin 500mg tablets
	99590997	Metformin 850mg tablets
	99514989	Metformin 850mg tablets
	99513990	Metformin 500mg tablets
	99513989	Metformin 850mg tablets
	98125989	Metformin 850mg tablets
	98654989	Metformin 500mg tablets
	99590998	Metformin 500mg tablets
	98493989	Metformin 500mg tablets
	98654990	Metformin 850mg tablets
	98493990	Metformin 850mg tablets
	76199978	Metformin 1g modified-release tablets
	71281979	Metformin 500mg/5ml oral solution
	54786979	Metformin 500mg modified-release tablets
	58558979	Metformin 500mg modified-release tablets
	74441978	Metformin 1g modified-release tablets
	74454978	Metformin 500mg modified-release tablets
	74453978	Metformin 500mg modified-release tablets
	87165998	Metformin with rosiglitazone 1000mg + 4mg tablet
	87166998	Metformin with rosiglitazone 1000mg + 2mg tablet
	87180998	Metformin & rosiglitazone 1g+2mg tablets
	87181998	Rosiglitazone 4mg / metformin 1g tablets
	87182998	Rosiglitazone 2mg / metformin 1g tablets
	87179998	Metformin & rosiglitazone 1g+4mg tablets
	87773998	Rosiglitazone 2mg / metformin 500mg tablets
	87772998	Metformin with rosiglitazone 500mg + 2mg tablet
	87774998	Metformin with rosiglitazone 500mg + 1mg tablet
	87771998	Metformin & rosiglitazone 500mg+1mg tablets
	87775998	Rosiglitazone 1mg / metformin 500mg tablets
	87770998	Metformin & rosiglitazone 500mg+2mg tablets
Lipid modifying drugs	81048998	Atorvastatin 20mg chewable tablets sugar free
	81051998	Atorvastatin 10mg chewable tablets sugar free
	83099998	Simvastatin 40mg/5ml oral suspension sugar free
	82655998	Nicotinic acid & laropiprant 1g+20mg tablets
	83030998	Simvastatin 80mg tablets
	81050998	Atorvastatin 10mg chewable tablets sugar free
	84268998	Colesevelam 625mg tablets
	84267998	Colesevelam 625mg tablets

Drug	Drug code	Preparation
Lipid modifying drugs	83594998	Nicotinic acid 1g / Irapiprant 20mg modified-release tablets
	79254979	Simvastatin 20mg/5ml oral suspension sugar free
	83188998	Bezafibrate 200mg tablets
	83187998	Bezafibrate 400mg modified-release tablets
	81049998	Atorvastatin 20mg chewable tablets sugar free
	82141978	Eicosapentaenoic acid 460mg / Docosahexaenoic acid 380mg capsules
	87853998	Nicotinic acid 1g modified-release tablets
	87852998	Nicotinic acid 500mg modified release tablets
	89154996	Cerivastatin 300microgram tablets
	86791998	Simvastatin 80mg / Ezetimibe 10mg tablets
	87854998	Nicotinic acid 750mg modified-release tablets
	89153996	Cerivastatin sodium 300mcg tablets
	88298997	Fenofibrate micronised 267mg capsules
	88534998	Rosuvastatin 10mg tablets
	86794998	Simvastatin 80mg / Ezetimibe 10mg tablets
	86510979	Ispaghula husk 3.5g sugar free granules
	87025998	Bezafibrate 400mg modified-release tablets
	87418998	Simvastatin 10mg tablets
	87918998	Simvastatin 10mg tablets
	89401998	Bezafibrate 400mg modified-release tablets
	87917998	Simvastatin 20mg tablets
	89089998	Bezafibrate 400mg modified release tablets
	87373998	Simvastatin 10mg tablets
	87760998	Colestipol 5g granules sachets sugar free
	86798998	Simvastatin 20mg / Ezetimibe 10mg tablets
	88297996	Fenofibrate micronised 267mg capsules
	87848998	Nicotinic acid pack
	86796998	Simvastatin 40mg / Ezetimibe 10mg tablets
	87849998	Nicotinic acid 375mg + 500mg + 750mg modified-release tablet
	87850998	Nicotinic acid 1g modified release tablets
	87851998	Nicotinic acid 750mg modified release tablets
	86797998	Simvastatin 20mg / Ezetimibe 10mg tablets
	87916998	Simvastatin 40mg tablets
	89306996	Atorvastatin 40mg tablets
	89311998	Atorvastatin 10mg tablets
	89617998	Ispaghula husk 3.5g sugar free granules
	89154997	Cerivastatin 200microgram tablets
	86795998	Simvastatin 40mg / Ezetimibe 10mg tablets
	89311997	Atorvastatin 20mg tablets
	89306998	Atorvastatin 10mg tablets
	89311996	Atorvastatin 40mg tablets
	88298996	Fenofibrate micronised 200mg capsules

Drug	Drug code	Preparation
Lipid modifying drugs	86788998	Simvastatin 40mg / Ezetimibe 10mg tablets
	86789998	Simvastatin 20mg / Ezetimibe 10mg tablets
	89153998	Cerivastatin sodium 100mcg tablets
	86787998	Simvastatin 80mg / Ezetimibe 10mg tablets
	89154998	Cerivastatin 100microgram tablets
	88297998	Fenofibrate micronised 67mg capsules
	86467998	Rosuvastatin 5mg tablets
	89285979	Nicotinic acid 500mg modified release tablets
	89800998	Eicosapentaenoic acid 460mg / Docosahexaenoic acid 380mg capsules
	89284979	Nicotinic acid 750mg modified release tablets
	88298998	Fenofibrate micronised 67mg capsules
	88297997	Fenofibrate micronised 200mg capsules
	89306997	Atorvastatin 20mg tablets
	87855998	Nicotinic acid 500mg modified-release tablets
	89283979	Nicotinic acid 1g modified release tablets
	89153997	Cerivastatin sodium 200mcg tablets
	86468998	Rosuvastatin 5mg tablets
	92447998	Cerivastatin sodium 400mcg tablets
	90973998	Rosuvastatin 20mg tablets
	93619998	Simvastatin 10mg tablets
	92408998	Rosuvastatin 20mg tablets
	93620996	Simvastatin 40mg tablets
	92448997	Cerivastatin 800microgram tablets
	92410998	Rosuvastatin 40mg tablets
	93620997	Simvastatin 20mg tablets
	93620998	Simvastatin 10mg tablets
	93871990	Simvastatin 40mg tablets
	92409998	Rosuvastatin 10mg tablets
	93243996	Pravastatin 40mg tablets
	91194998	Fluvastatin 80mg modified-release tablets
	93010990	Colestyramine 4g oral powder sachets sugar free
	92549990	Fenofibrate micronised 200mg capsules
	93244998	Pravastatin 10mg tablets
	93244997	Pravastatin 20mg tablets
	93244996	Pravastatin 40mg tablets
	92539998	Rosuvastatin 40mg tablets
	90310998	Atorvastatin 80mg tablets
	93243997	Pravastatin 20mg tablets
	92448998	Cerivastatin 400microgram tablets
	94407990	Simvastatin 20mg tablets
	93838990	Bezafibrate 200mg tablets
	92471998	Simvastatin 80mg tablets

Drug	Drug code	Preparation
Lipid modifying drugs	93851992	Colestipol 5g granules sachets sugar free
	90309998	Atorvastatin 80mg tablets
	92460998	Fenofibrate micronised 160mg tablets
	91316998	Colestyramine sugar free powder
	93243998	Pravastatin 10mg tablets
	92292998	Ezetimibe 10mg tablets
	92154990	Simvastatin 20mg/5ml oral suspension sugar free
	93541998	Colestyramine 4g oral powder sachets
	93542998	Colestyramine 4g oral powder sachets sugar free
	92804997	Fluvastatin 40mg capsules
	94189997	Fenofibrate micronised 200mg capsules
	92293998	Ezetimibe 10mg tablets
	90649998	Fenofibrate 200mg capsules
	92805998	Fluvastatin 20mg capsules
	92804998	Fluvastatin 20mg capsules
	92805997	Fluvastatin 40mg capsules
	93619996	Simvastatin 40mg tablets
	92220998	Simvastatin 80mg tablets
	90653998	Colestyramine 4g oral powder sachets sugar free
	94188997	Fenofibrate micronised 200mg capsules
	94112992	Cholestyramine 325 mg cap
	94188998	Fenofibrate 100mg capsule
	94189998	Fenofibrate 100mg capsules
	92804996	Fluvastatin 80mg modified-release tablets
	93619997	Simvastatin 20mg tablets
	95480990	Simvastatin 10mg tablets
	95479990	Simvastatin 20mg tablets
	95952997	Bezafibrate 400mg modified-release tablets
	95550990	Simvastatin 20mg tablets
	95551990	Simvastatin 10mg tablets
	94925998	Eicosapentaenoic acid 170mg / Docosahexaenoic acid 115mg capsules
	95478990	Simvastatin 40mg tablets
	95471990	Simvastatin 40mg tablets
	95475990	Simvastatin 20mg tablets
	94799998	Fenofibrate micronised 160mg tablets
	96295997	Gemfibrozil 600mg tablets
	95474990	Simvastatin 40mg tablets
	95472990	Simvastatin 20mg tablets
	95451990	Simvastatin 10mg tablets
	95549990	Simvastatin 40mg tablets
	96295998	Gemfibrozil 300mg capsules
	94927990	Simvastatin 80mg tablets

Drug	Drug code	Preparation
Lipid modifying drugs	94827992	Colestyramine 4g oral powder sachets
	95501990	Simvastatin 40mg tablets
	94782990	Pravastatin 20mg tablets
	95185990	Simvastatin 80mg tablets
	95494990	Simvastatin 20mg tablets
	94605998	Colestipol 5g granules sachets sugar free
	95495990	Simvastatin 10mg tablets
	94851990	Pravastatin 10mg tablets
	95500990	Simvastatin 80mg tablets
	94830990	Pravastatin 20mg tablets
	95502990	Simvastatin 20mg tablets
	97078998	Fish oil concentrate 1g capsules
	96685990	Bezafibrate 400mg modified-release tablets
	96685989	Bezafibrate 200mg tablets
	97377979	Cerivastatin sodium 300mcg tablets
	94831990	Pravastatin 10mg tablets
	97078997	Fish oil concentrate oral liquid
	94661998	Colestipol 5g granules sachets sugar free
	95482990	Simvastatin 20mg tablets
	95483990	Simvastatin 10mg tablets
	95486990	Simvastatin 40mg tablets
	95508990	Simvastatin 10mg tablets
	95487990	Simvastatin 20mg tablets
	95952998	Bezafibrate 200mg tablets
	97078996	Fish oil concentrate oral emulsion
	94850990	Pravastatin 20mg tablets
	94849990	Pravastatin 40mg tablets
	94662998	Colestipol 5g granules sachets sugar free
	95493990	Simvastatin 40mg tablets
	95847990	Colestyramine 4g oral powder sachets sugar free
	95098992	Hexopal 200 mg tab
	97455979	Pravastatin 10mg tablets
	95481990	Simvastatin 40mg tablets
	94661997	Colestipol 5g granules sachets sugar free
	97430979	Fluvastatin 20mg capsules
	95805998	Bezafibrate 400mg modified release tablets
	95405990	Simvastatin 40mg tablets
	94789990	Pravastatin 10mg tablets
	95401998	Probucol 250mg tablet
	97247997	Gemfibrozil 600mg tablets
	97247998	Gemfibrozil 300mg capsules
	96134990	Colestyramine 4g oral powder sachets

Drug	Drug code	Preparation
Lipid modifying drugs	95278990	Simvastatin 20mg tablets
	95277990	Simvastatin 40mg tablets
	95372990	Simvastatin 40mg tablets
	95406990	Simvastatin 20mg tablets
	96642998	Clofibrate 500mg capsules
	94925997	Eicosapentaenoic acid 170mg/g / docosahexaenoic acid 115mg/g oral liquid
	95450990	Simvastatin 20mg tablets
	97433979	Ciprofibrate 100mg tablets
	96656990	Bezafibrate 200mg tablets
	97454979	Pravastatin 20mg tablets
	94883998	Acipimox 250mg capsules
	94882998	Acipimox 250mg capsules
	95279990	Simvastatin 10mg tablets
	95445990	Simvastatin 10mg tablets
	96021990	Fenofibrate micronised 200mg capsules
	95442990	Simvastatin 80mg tablets
	95449990	Simvastatin 40mg tablets
	95448990	Simvastatin 80mg tablets
	95443990	Simvastatin 40mg tablets
	94920990	Simvastatin 20mg tablets
	97529979	Fenofibrate micronised 160mg tablets
	99957979	Atorvastatin 80mg tablets
	98066998	Nicotinic acid 25mg tablet
	97518979	Simvastatin 10mg tablets
	97551979	Fenofibrate micronised 200mg capsules
	97476979	Simvastatin 40mg tablets
	97594979	Bezafibrate 400mg modified-release tablets
	97723998	Ciprofibrate 100mg tablets
	97756979	Atorvastatin 40mg tablets
	97495979	Simvastatin 20mg tablets
	97655990	Colestyramine 4g oral powder sachets
	97751998	Ciprofibrate 100mg tablets
	99197998	Colestyramine 4g oral powder sachets
	97559979	Fenofibrate micronised 200mg capsules
	98455998	Clofibrate 500mg capsules
	97482979	Simvastatin 40mg tablets
	97895997	Bezafibrate 400mg modified-release tablets
	98066996	Nicotinic acid 100mg tablet
	97808989	Gemfibrozil 600mg tablets
	97595979	Bezafibrate 400mg modified-release tablets
	97895998	Bezafibrate 200mg tablets
	97800992	Nicotinic acid 500mg modified-release tablets

Drug	Drug code	Preparation
Lipid modifying drugs	97541979	Fenofibrate micronised 267mg capsules
	97494979	Simvastatin 20mg tablets
	97508979	Simvastatin 10mg tablets
	97612979	Cerivastatin sodium 200mcg tablets
	98066997	Nicotinic acid 50mg tablets
	99014998	Bezafibrate 400mg modified-release tablets
	99456998	Probucol 250mg tablets
	97705979	Cerivastatin sodium 100mcg tablets
	99725990	Nicotinic acid 50mg tablets
	58658979	Atorvastatin 10mg tablets
	61212979	Atorvastatin 20mg chewable tablets sugar free
	62570979	Simvastatin 40mg/5ml oral suspension sugar free
	59480979	Fluvastatin 20mg capsules
	57833979	Atorvastatin 10mg tablets
	62597979	Simvastatin 20mg/5ml oral suspension sugar free
	61213979	Atorvastatin 20mg chewable tablets sugar free
	58151979	Atorvastatin 60mg tablets
	58671979	Atorvastatin 40mg tablets
	64839979	Simvastatin 40mg/5ml oral suspension sugar free
	61489979	Simvastatin 20mg tablets
	54777979	Eicosapentaenoic acid 460mg / Docosahexaenoic acid 380mg capsules
	58668979	Atorvastatin 40mg tablets
	58669979	Atorvastatin 40mg tablets
	57838979	Atorvastatin 40mg tablets
	76814978	Eicosapentaenoic acid 460mg / Docosahexaenoic acid 380mg capsules
	72952978	Eicosapentaenoic acid 460mg / Docosahexaenoic acid 380mg capsules
	57832979	Atorvastatin 20mg tablets
	52277979	Eicosapentaenoic acid 460mg / Docosahexaenoic acid 380mg capsules
	58654979	Atorvastatin 20mg tablets
	58711979	Atorvastatin 80mg tablets
	58653979	Atorvastatin 20mg tablets
	61215979	Atorvastatin 10mg chewable tablets sugar free
	58661979	Atorvastatin 10mg tablets
	86020998	Simvastatin 20mg/5ml oral solution sugar free
	92329998	Eicosapentaenoic acid 460mg / docosahexaenoic acid 380mg capsules
	58153979	Atorvastatin 30mg tablets
	64582979	Atorvastatin 40mg/5ml oral suspension
	64584979	Atorvastatin 40mg/5ml oral solution
	64586979	Atorvastatin 10mg/5ml oral suspension
	64588979	Atorvastatin 10mg/5ml oral solution
	64634979	Atorvastatin 20mg/5ml oral suspension
	79256979	Simvastatin 10mg/5ml oral suspension

S5: Hazard of women with PCOS to develop NAFLD compared to women without PCOS using diabetes censored data

	Diabetes censored data		DM and IGR censored data	
	PCOS (Exposed)	Controls (Unexposed)	PCOS (Exposed)	Controls (Unexposed)
Total number of participants	61,841	119,373	61,557	119,077
Person years	291,198	567,775	290,103	566,580
Incident NAFLD n (%)	243 (0.39)	211 (0.18)	239 (0.39)	208 (0.17)
Incidence rates per 10,000 person years	8.34	3.72	8.24	3.67
Hazard Ratio (95% CI)	2.27 (1.89 to 2.73)		2.28 (1.89 to 2.75)	
p-value	<0.001		<0.001	
Adjusted Hazard Ratio (95% CI)*	2.22 (1.85 to 2.68)		2.25 (1.87 to 2.72)	
p-value	<0.001		<0.001	

* Adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation and hypothyroidism at baseline

S6: Regression model estimates for hazard of women with PCOS to develop NAFLD compared to women without PCOS (n= 184,274)

Characteristics		Hazard ratio	95% CI	P value
PCOS		2.23	(1.86, 2.66)	<0.001
Age		1.05	(1.03, 1.06)	<0.001
Townsend index				
	1	1.00		
	2	1.10	(0.81, 1.49)	0.54
	3	1.21	(0.90, 1.61)	0.20
	4	1.47	(1.11, 1.95)	0.01
	5	1.53	(1.13, 2.07)	0.01
	Missing or implausible data	1.52	(0.99, 2.35)	0.06
BMI (kg/m²) Category				
	<25	1.00		
	25-30	3.37	(2.36, 4.83)	<0.001
	>30	6.98	(5.07, 9.60)	<0.001
	Missing or implausible data	1.96	(1.29, 3.00)	<0.001
Diabetes or IGR*		2.39	(1.76, 3.25)	<0.001
Hypothyroidism baseline		1.39	(0.97, 1.98)	0.07

* 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 120min after ingestion of 75g glucose in the oral glucose tolerance test)

S7: Subgroup analysis for hazard of women with PCOS to develop NAFLD compared to women without PCOS stratified by BMI category

	BMI <25kg/m2		25-30Kg/m2		>30Kg/m2	
	PCOS (Exposed)	Controls (Unexposed)	PCOS (Exposed)	Controls (Unexposed)	PCOS (Exposed)	Controls (Unexposed)
Total number of participants	21,689	44,742	12,412	24,656	19,499	32,627
Person years	97,957	205,355	58,842	118,646	93,157	158,056
Incident NAFLD n (%)	21 (0.10)	23 (0.05)	48 (0.39)	47 (0.19)	173 (0.89)	142 (0.44)
Incidence rates per 10,000 person years	2.14	1.12	8.16	4.0	18.57	8.98
Hazard Ratio (95% CI)	1.91 (1.06 to 3.45)		2.05 (1.37 to 3.07)		2.06 (1.65 to 2.57)	
p-value	0.03		<0.001		<0.001	
Adjusted Hazard Ratio (95% CI)*	1.85 (1.02 to 3.34)		2.05 (1.37 to 3.07)		2.06 (1.65 to 2.58)	
p-value	0.043		0.001		<0.001	

* Adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation and hypothyroidism at baseline

S8: Regression model estimates for Impact of BMI and PCOS on hazard of women with PCOS to develop NAFLD compared to women without PCOS (n= 184,274)

Variable	Hazard Ratio	95% CI	P value
Age	1.05	(1.04, 1.06)	<0.001
Townsend index			
1	1.00		
2	1.10	(0.81, 1.49)	0.547
3	1.20	(0.90, 1.61)	0.207
4	1.47	(1.11, 1.96)	0.007
5	1.53	(1.13, 2.07)	0.006
Missing or implausible data	1.53	(0.99, 2.36)	0.057
Diabetes or IGR*	2.29	(1.65, 3.17)	<0.001
Hypothyroidism at baseline	1.39	(0.98, 1.99)	0.069
Body mass index (kg/m²) and PCOS			
Control women and BMI < 25	1.00		
Control and BMI 25-30	3.27	(1.98, 5.38)	<0.001
Control and BMI >30	6.82	(4.38, 10.61)	<0.001
Control and Missing BMI	0.93	(0.44, 1.96)	0.858
PCOS women and <25	1.92	(1.06, 3.46)	0.031
PCOS women and 25-30	6.75	(4.10, 11.10)	<0.001
PCOS women and >30	14.12	(9.11, 21.87)	<0.001
PCOS women and Missing BMI	5.80	(3.40, 9.92)	<0.001

* 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 120min after ingestion of 75g glucose in the oral glucose tolerance test)

S9: Regression model estimates for hazard of women with PCOS to develop NAFLD compared to women without PCOS with addition of PCOS diagnostic features as predictor(n= 184,274)

Characteristics		Hazard ratio	95% CI	P value
PCOS		2.07	(1.72, 2.50)	<0.001
Age		1.05	1.04, 1.06)	<0.001
Townsend index				
	1	1.0		
	2	1.10	(0.81, 1.49)	0.551
	3	1.20	(0.90, 1.60)	0.216
	4	1.46	(1.10, 1.93)	0.009
	5	1.50	(1.10, 2.03)	0.010
	Missing or implausible data	1.51	(0.98, 2.34)	0.062
BMI (kg/m²) Category				
	<25	1.0		
	25-30	3.35	(2.34, 4.79)	<0.001
	>30	6.89	(5.00, 9.49)	<0.001
	Missing or implausible data	1.97	(1.29, 3.00)	0.002
Diabetes or IGR*		2.24	(1.61, 3.10)	<0.001
Hypothyroidism baseline		1.38	(0.96, 1.97)	0.078
Hirsutism		1.37	(1.06, 1.77)	0.016
Alopecia		1.54	(1.16, 2.05)	0.003
Acne		0.99	(0.76, 1.27)	0.910

* 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 120min after ingestion of 75g glucose in the oral glucose tolerance test)

S10:Factors associated with NAFLD amongst Non PCOS women (n=121,064).

Covariates	Hazard Ratio	95% CI	P value
Age	1.06	(1.04, 1.08)	< 0.001
Townsend			
1	1.00		
2	1.30	(0.80, 2.12)	0.29
3	1.74	(1.11, 2.72)	0.02
4	1.72	(1.09, 2.71)	0.02
5	1.93	(1.20, 3.11)	0.01
Missing	2.07	(1.03, 4.14)	0.04
BMI (kg/m²) category			
<25	1.00		
25-30	3.19	(1.93, 5.26)	< 0.001
>30	6.39	(4.09,9.99)	< 0.001
Missing	0.94	(0.44, 1.96)	0.86
Diabetes or IGR*	1.35	(0.57, 3.17)	< 0.001
Hypothyroidism	1.89	(1.14, 3.13)	0.01
Anovulation	1.60	(1.16, 2.20)	0.004
Androgen excess feature			
Hirsutism	1.74	(0.64, 4.75)	0.28
Acne	1.21	(0.76, 1.91)	0.42
Alopecia	1.30	(0.64, 2.64)	0.46
Lipid modifying drugs	0.75	(0.31, 1.79)	0.51
Metformin	1.83	(0.71, 4.75)	0.21
Antiandrogen drug	0.75	(0.39, 1.47)	0.41

S11: Baseline characteristics in cohort of women with available serum testosterone measurement (n=71,061)

Characteristics	Serum testosterone concentration categories (nmol/L)						
	< 1	1 - 1.49	1.5 - 1.99	2 - 2.49	2.5 - 2.99	3 - 3.49	≥ 3.5
Population n (%)	23,958 (33.7)	17,663 (24.9)	13,528 (19.0)	8,168 (11.5)	4,061 (5.7)	1,835 (2.6)	1,848 (2.6)
Age mean (SD)	34.9 (7.8)	32.4 (7.8)	31.0 (7.5)	29.7 (7.4)	28.4 (6.9)	27.6 (6.8)	27.6 (6.9)
Townsend index n (%)							
1	5,597 (23.4)	3,934 (22.3)	2,939 (21.7)	1,645 (20.1)	809 (19.9)	348 (19.0)	329 (17.8)
2	4,693 (19.6)	3,364 (19.0)	2,559 (18.9)	1,498 (18.3)	718 (17.7)	324 (17.7)	296 (16.0)
3	4,988 (20.8)	3,724 (21.1)	2,801 (20.7)	1,730 (21.2)	860 (21.2)	419 (22.8)	378 (20.4)
4	4,261 (17.8)	3,332 (18.9)	2,636 (19.5)	1,617 (19.8)	816 (20.1)	394 (21.5)	413 (22.4)
5	2,775 (11.6)	2,197 (12.4)	1,725 (12.8)	1,135 (13.9)	619 (15.2)	246 (13.4)	327 (17.7)
Missing or implausible data	1,644 (6.9)	1,112 (6.3)	868 (6.4)	543 (6.6)	239 (5.9)	104 (5.7)	105 (5.7)
BMI (kg/m²) category n (%)							
<25	11,326 (47.3)	7,236 (41.0)	5,138 (38.0)	2,875 (35.2)	1,305 (32.1)	556 (30.3)	546 (29.6)
25-30	5,060 (21.1)	3,694 (20.9)	2,743 (20.3)	1,563 (19.1)	789 (19.4)	357 (19.5)	360 (19.5)
>30	4,903 (20.5)	4,282 (24.2)	3,655 (27.0)	2,400 (29.4)	1,259 (31.0)	601 (32.8)	628 (34.0)
Missing or implausible data	2,669 (11.1)	2,451 (13.9)	1,992 (14.7)	1,330 (16.3)	708 (17.4)	321 (17.5)	314 (17.0)

S11 continued

Characteristics	Serum testosterone concentration categories (nmol/L)						
	< 1	1 - 1.49	1.5 - 1.99	2 - 2.49	2.5 - 2.99	3 - 3.49	≥ 3.5
Smoking status n (%)							
Smokers	19,600 (81.8)	13,409 (75.9)	9,685 (71.6)	5,585 (68.4)	2,725 (67.1)	1,200 (65.4)	1,142 (61.8)
Smokers	4,056(16.9)	3,861 (21.9)	3,469 (25.6)	2,288 (28.0)	1,177 (29.0)	567 (30.9)	630 (34.1)
Missing	302 (1.3)	393 (2.2)	374 (2.8)	295 (3.6)	159 (3.9)	68 (3.7)	76 (4.1)
Medical conditions n (%)							
Diabetes mellitus	350 (1.5)	244 (1.4)	149 (1.1)	109 (1.3)	52 (1.3)	27 (1.5)	27 (1.5)
Hypertension	648 (2.7)	394 (2.2)	330 (2.4)	187 (2.3)	95 (2.3)	44 (2.4)	46 (2.5)
Hypothyroidism	1,172 (4.9)	700 (4.0)	416 (3.1)	255 (3.1)	102 (2.5)	42 (2.3)	40 (2.2)
Impaired glucose regulation	129 (0.5)	66 (0.4)	50 (0.4)	26 (0.3)	13 (0.3)	5 (0.3)	6 (0.3)
PCOS	419 (1.8)	426 (2.4)	371 (2.7)	260 (3.2)	159 (3.9)	82 (4.5)	76 (4.1)

S12: Factors associated with NAFLD amongst the cohort of women with available serum testosterone measurement (n=71,061)

Covariates		Hazard ratio	95% CI	P Value
Serum testosterone concentration (nmol/L)				
	< 1.0	1.0		
	1.0 - 1.49	1.39	(0.93, 2.07)	0.112
	1.5 - 1.99	1.04	(0.66, 1.65)	0.856
	2.0 - 2.49	0.89	(0.52, 1.55)	0.684
	2.5 - 2.99	1.18	(0.61, 2.27)	0.616
	3.0 - 3.49	2.30	(1.16, 4.53)	0.017
	>3.5	2.40	(1.24, 4.66)	0.009
Age		1.04	(1.02, 1.06)	<0.001
Townsend index				
	1	1.00		
	2	1.44	(0.91, 2.28)	0.124
	3	1.07	(0.66, 1.73)	0.793
	4	1.62	(1.03, 2.54)	0.037
	5	1.30	(0.77, 2.19)	0.328
	Missing or implausible data	1.55	(0.74, 3.25)	0.247
BMI category (kg/m²)				
	<25	1.00		
	25-30	4.47	(2.45, 8.16)	<0.001
	>30	10.86	(6.30, 18.72)	<0.001
	Missing or implausible data	3.68	(1.85, 7.33)	<0.001
Hypothyroidism		1.20	(0.66, 2.16)	0.547
Diabetes or IGR*		2.64	(1.53, 4.53)	<0.001

* 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 120min after ingestion of 75g glucose in the oral glucose tolerance

S13: Hazard of women with serum testosterone level ≥ 1 nmol/L to develop NAFLD compared to women with serum testosterone level <1 nmol/L (n=71,061)

	Serum testosterone concentration categories (nmol/L)						
	< 1	1 - 1.49	1.5 - 1.99	2 - 2.49	2.5 - 2.99	3 - 3.49	≥ 3.5
Number of participants	23,958	17,663	13,528	8,168	4,061	1,835	1,848
Incident NAFLD n (%)	44 (0.18)	53 (0.3)	34 (0.25)	19 (0.23)	12 (0.30)	11 (0.60)	12 (0.65)
Person years	85,541	73,870	62,685	39,970	19,924	9,187	9,306
Incidence rate per 10,000 person years	5.14	7.17	5.42	4.75	6.02	11.97	12.89
Hazard Ratio	1	1.37	1.03	0.9	1.15	2.28	2.45
95% CI of Hazard ratio		(0.92, 2.05)	(0.66, 1.62)	0.53, 1.55)	(0.61, 2.18)	(1.17, 4.41)	(1.29, 4.64)
p-value		0.12	0.89	0.71	0.67	0.015	0.006
Adjusted Hazard Ratio*	1	1.39	1.04	0.89	1.18	2.3	2.4
95% CI Baseline adjusted Hazard Ratio*		(0.93, 2.07)	(0.66, 1.65)	(0.52, 1.55)	(0.61, 2.28)	(1.16, 4.53)	(1.24, 4.66)
p-value		0.11	0.86	0.68	0.62	0.017	0.009

* Adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation and hypothyroidism at baseline

S14:

Risk of NAFLD in women with available serum testosterone measurement following exclusion of participants with PCOS/PCO codes (n= 58,606)

Covariates	Hazard ratio	95% CI	P Value
Serum testosterone concentration (nmol/L)			
< 1.0	1.0		
1.0 - 1.49	1.47	(0.93, 2.32)	0.101
1.5 - 1.99	0.81	(0.45, 1.45)	0.479
2.0 - 2.49	1.02	(0.53, 1.95)	0.960
2.5 - 2.99	0.98	(0.38, 2.53)	0.958
3.0 - 3.49	0.97	(0.23, 4.08)	0.967
>3.5	2.23	(0.78, 6.38)	0.135

*Patients without a known diagnosis of PCOS who had serum testosterone measurements. A limitation of these data is that many of these women with serum testosterone measurements may have had suspected androgen excess, and hence PCOS, despite the fact that this was not coded.

S15: Baseline characteristics in the cohort of women with available serum SHBG measurement (n=49,625)

Characteristics	Serum SHBG concentration categories (nmol/L)					
	<20	20 - 29.99	30 - 39.99	40 - 49.99	50 - 59.99	≥60
Population n (%)	4,544 (9.2)	7,573 (15.3)	7,929 (16.0)	7,133 (14.4)	5,796 (11.7)	16,650 (33.6)
Age mean (SD)	28.5 (7.5)	29.8 (7.6)	30.6 (7.7)	31.1 (7.7)	31.8 (7.6)	32.4 (7.5)
Townsend index n (%)						
1	783 (17.2)	1,363 (18.0)	1,527 (19.3)	1,472 (20.6)	1,234 (21.3)	3,555 (21.4)
2	718 (15.8)	1,191 (15.8)	1,395 (17.6)	1,250 (17.5)	1,022 (17.6)	3,193 (19.2)
3	950 (20.9)	1,589 (21.0)	1,644 (20.7)	1,461 (20.5)	1,219 (21.0)	3,426 (20.6)
4	1,041 (22.9)	1,621 (21.4)	1,641 (20.7)	1,399 (19.6)	1,139 (19.6)	3,135 (18.8)
5	741 (16.3)	1,247 (16.5)	1,154 (14.6)	1,036 (14.5)	749 (12.9)	1,970 (11.8)
Missing or implausible data	311 (6.8)	562 (7.4)	568 (7.2)	515 (7.2)	433 (7.5)	1,371 (8.2)
BMI (kg/m²) categorised n (%)						
<25	509 (11.2)	1,461 (19.3)	2,307 (29.0)	2,908 (40.8)	2,809 (48.5)	9,896 (59.4)
25-30	802 (17.6)	1,568 (20.7)	1,836 (23.2)	1,597 (22.4)	1,238 (21.4)	2,882 (17.3)
>30	2,578 (56.7)	3,463 (45.7)	2,673 (33.7)	1,591 (22.3)	924 (15.9)	1,642 (9.9)
Missing or implausible data	655 (14.4)	1,081 (14.3)	1,113 (14.0)	1,037 (14.5)	825 (14.2)	2,230 (13.4)
Smoking status n (%)						
Non-smokers	3,376 (74.3)	5,584 (73.7)	5,908 (74.5)	5,379 (75.4)	4,406 (76.02)	13,078 (78.6)
Smokers	1,032 (22.7)	1,799 (23.8)	1,850 (23.3)	1,625 (22.8)	1,274 (22.0)	3,344 (20.1)
Missing or implausible data	136 (3.0)	190 (2.5)	171 (2.2)	129 (1.8)	116 (2.0)	228 (1.4)

S15 continued

Characteristics		Serum SHBG concentration categories (nmol/L)					
		<20	20 - 29.99	30 - 39.99	40 - 49.99	50 - 59.99	≥60
Medical conditions at baseline							
n (%)							
	Diabetes mellitus	167 (3.7)	123 (1.6)	92 (1.2)	79 (1.1)	48 (0.8)	127 (0.8)
	Hypertension	190 (4.2)	255 (3.4)	182 (2.3)	138 (1.9)	107 (1.8)	211 (1.3)
	Hypothyroidism	232 (5.1)	313 (4.1)	298 (3.8)	283 (4.0)	213 (3.7)	608 (3.7)
	Impaired glucose regulation	54 (1.2)	58 (0.8)	39 (0.5)	15 (0.2)	17 (0.3)	37 (0.2)
	PCOS	441 (9.7)	559 (7.4)	404 (5.1)	283 (4.0)	190 (3.3)	478 (2.9)

S16: Factors associated with NAFLD amongst the cohort of women with available serum SHBG measurement (n=49,625)

Covariate	Hazard ratio	95% CI	P Value
SHBG (nmol/L) category			
<20	4.98	(2.45, 10.11)	<0.001
20 - 29.99	4.75	(2.44, 9.25)	<0.001
30 - 39.99	1.66	(0.77, 3.57)	0.199
40 - 49.99	2.06	(0.96, 4.44)	0.065
50 - 59.99	0.94	(0.33, 2.67)	0.903
≥ 60	1.00		
Age	1.07	(1.05, 1.10)	<0.001
Townsend score #			
1	1.00		
2	2.31	(1.24, 4.31)	0.009
3	1.78	(0.94, 3.36)	0.076
4	2.53	(1.37, 4.69)	0.003
5	1.84	(0.92, 3.67)	0.082
Missing or implausible data	1.50	(0.54, 4.15)	0.431
BMI (kg/m ²) Category			
<25	1.00		
25-30	4.08	(1.97, 8.43)	<0.001
>30	5.41	(2.69, 10.88)	<0.001
Missing or implausible data	1.87	(0.73, 4.79)	0.186
Hypothyroidism	1.08	(0.53, 2.23) (0.67, 2.92)	0.856
Diabetes or IGR*	1.39		0.430

Townsend score – presented as quintiles with 1 least deprived and 5 most deprived

* 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 120min after ingestion of 75g glucose in the oral glucose tolerance test)

S17: Hazard of women with serum SHBG level <60 nmol/L to develop NAFLD compared to women with serum SHBG level ≥60 nmol/L (n=49,625)

	Serum SHBG concentration categories (nmol/L)					
	<20	20 - 29.99	30 - 39.99	40 - 49.99	50 - 59.99	≥60
Number of participants	4,544	7,573	7,929	7,133	5,796	16,650
Incident NAFLD n (%)	31 (0.68)	47 (0.62)	16 (0.20)	15 (0.21)	5 (0.09)	12 (0.07)
Person years	17,263	28,470	29,428	26,351	20,233	55,353
Incidence rate per 10,000 years	17.96	16.51	5.44	5.70	2.47	2.17
Hazard Ratio	6.39	5.77	2.18	2.22	0.77	1
95% CI of Hazard ratio	(3.64, 11.23)	(3.38, 9.83)	(1.16, 4.09)	(1.16, 4.22)	(0.29, 2.08)	
p-value	<0.001	<0.001	0.015	0.016	0.61	
Adjusted Hazard Ratio*	4.98	4.75	1.66	2.06	0.94	1
95% CI Adjusted Hazard Ratio*	(2.45, 10.11)	(2.44, 9.25)	(0.77, 3.57)	(0.96, 4.44)	(0.33, 2.67)	
p-value	<0.001	<0.001	0.19	0.06	0.90	

* Adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation and hypothyroidism at baseline

S18:Risk of NAFLD based on available serum SHBG measurement following exclusion of participants with PCOS/PCO codes (n=38,784)

<u>Covariate</u>	<u>Hazard ratio</u>	<u>95% CI</u>	<u>P Value</u>
SHBG (nmol/L) category			
<20	5.42	(2.32, 12.64)	<0.001
20 - 29.99	4.85	(2.22, 10.58)	<0.001
30 - 39.99	1.76	(0.73, 4.26)	0.210
40 - 49.99	1.71	(0.67, 4.35)	0.259
50 - 59.99	0.98	(0.30, 2.21)	0.977
≥ 60	1.00		

*Patients without a known diagnosis of PCOS who had serum SHBG measurements. A limitation of these data is that many of these women with serum SHBG measurements may have had suspected androgen excess, and hence PCOS, despite the fact that this was not coded.

Polycystic ovary syndrome and risk of Non Alcoholic Fatty Liver Disease

Background

Polycystic ovary syndrome is an under-recognised, common and complex endocrine condition (Teede, Gibson-Helm, Norman, & Boyle, 2014) affecting females of reproductive age. The currently adopted diagnostic criteria for PCOS are the Rotterdam 2003 criteria, which require any two of the three criteria to confirm PCOS namely oligo-ovulation and or anovulation, clinical and or biochemical signs of hyperandrogenism and polycystic ovaries, provided secondary causes are excluded (Legro, 2015).

Main clinical manifestations of PCOS are related to hyperandrogenism or /and anovulation. Hyperandrogenism manifests mainly as hirsutism, with acne and alopecia being common associations (Shivaprakash et al., 2013).

There are many long term complications of PCOS such as an increased risk of developing gestational diabetes, type II diabetes, sleep apnoea, cardiovascular disease, endometrial cancer (Barry, Azizia, & Hardiman, 2014).

The prevalence of PCOS among women of reproductive age (15 to 45 years) in many countries vary from 6 to 8% (Diamanti-Kandarakis et al., 1999) (Michelmores, Balen, Dunger, & Vessey, 1999) (Kumarapeli, Seneviratne, Wijeyaratne, Yapa, & Dodampahala, 2008) (Jiao et al., 2014). Therefore, PCOS is considered the most common endocrine abnormality among young women (Livadas & Diamanti-Kandarakis, 2013).

The healthcare related economic burden of PCOS during reproductive life span in the United States in 2005 was estimated to be 4.36 billion US dollars which arose from initial evaluation (2%), hormonal treatment of menstrual dysfunction or abnormal uterine bleeding (31%), providing infertility care (12.1%), PCOS associated diabetes (40.5%) and in treating hirsutism (14.2%) (Azziz, Marin, Hoq, Badamgarav, & Song, 2005). This analysis has not taken into account the potential for premature cardiovascular outcomes, particularly in more vulnerable ethnic groups and regions of the world.

Although prevalence of PCOS has been widely reported, we could not access literature on the incidence of PCOS.

Purpose

The purpose of this doctoral research study is to determine the association between PCOS and Non Alcoholic Fatty Liver Disease (NAFLD).

Methods

Data source

This study will use THIN data. This retrospective study intends to use general practice data as this is one of the most common longitudinal data available in the UK. Although general practice data is available through many sources such as GPRD and THIN, we chose THIN data base as the University of Birmingham has the access to THIN database. Many of the variables included in this study are part of Quality and Outcomes Framework (QoF), quality of maintenance of this data is closely monitored.

The Health Improvement Network (THIN) database, an electronic primary care data base containing anonymised patient data will be used as the data source in this study. THIN database covers 5.7% of

the UK population (Dhalwani, Tata, Coleman, Fleming, & Szatkowski, 2013) and has been proven to be generalizable to UK demographics (Blak, Thompson, Dattani, & Bourke, 2011). More than 800 research papers have been published using the General Practise Research Database which is similar to THIN (Williams, van Staa, Puri, & Eaton, 2012). This database includes approximately two million reproductive aged women (aged 15 to 49 years) and 9.5 million patients from 495 practices in 2013 (Dhalwani et al., 2013). There were around 53 000 prevalent cases and 21 000 incident cases of PCOS in 2012 (Morgan, Jenkins-Jones, Currie, & Rees, 2012). Information available in this dataset includes demography, clinical assessment (diagnoses and health contacts), investigations and prescriptions.

Analysis

Cohorts of exposed and non-exposed groups will be developed. The inclusion criteria for the exposed group will include PCOS patients diagnosed after one year wash in period from the registration at the practice and not being diagnosed with NAFLD at the time of diagnosis of PCOS. The index date for this group will be the date of diagnosis of PCOS and the exit date will be the date of diagnosis of NAFLD or termination of study period or transfer to another practice or death of the patient and whichever the earliest.

For prevalent PCOS patients (diagnosed before study entry) the date of registration at the practice or when the practice becomes eligible to take part will be considered as the index date for this group.

The criteria for selection of non-exposed group for the comparison with the exposed group with incident cases will include 1) those who do not fulfil the criteria to diagnose PCOS, 2) having one year wash in period prior to the index date of the exposed participant 3) not being diagnosed with NAFLD at the time of diagnosis of PCOS 4) randomly chosen to match with the age (same or one year around the year of birth) and clinical practice of the exposed.

The proportion of exposed to non-exposed will be maintained as 1:4 in each cohort. Comparison will be made in the demographic characteristics between those selected and non-selected among the eligible non-exposed registrants in the data set to assess the randomness of selection of non-exposed groups.

Within the PCOS cohort we will explore what factors alter their risk of NAFLD, For example androgen excess features, BMI and antiandrogen medications.

Endpoints

The endpoints will be whichever occur first among any of the following 1) diagnosis of NAFLD, 2) patient leaving the practice 3) death of the patient and 4) the termination of documentation in the THIN database.

Confounding variables

Confounders were chosen through review of literature and controlled using multiple

regression analysis and include Age, BMI, smoking, Townsend index for social deprivation, diabetes mellitus, hypothyroidism and impaired glucose regulation (Loomis et al. 2016).

Reference list

Abdul Sultan, A., Tata, L. J., Grainge, M. J., & West, J. (2013). The incidence of first venous thromboembolism in and around pregnancy using linked primary and secondary care data: a population based cohort study from England and comparative meta-analysis. *PloS One*, 8(7), e70310. <https://doi.org/10.1371/journal.pone.0070310>

Azziz, R., Marin, C., Hoq, L., Badamgarav, E., & Song, P. (2005). Health care-related economic burden of the polycystic ovary syndrome during the reproductive life span. *The Journal of Clinical Endocrinology and Metabolism*, 90(8), 4650–8. <https://doi.org/10.1210/jc.2005-0628>

Barry, J. A., Azizia, M. M., & Hardiman, P. J. (2014). Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction Update*, 20(5), 748–758. <https://doi.org/10.1093/humupd/dmu012>

Blak, B. T., Thompson, M., Dattani, H., & Bourke, A. (2011). Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Informatics in Primary Care*, 19(4), 251–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22828580>

Dhalwani, N. N., Tata, L. J., Coleman, T., Fleming, K. M., & Szatkowski, L. (2013). Completeness of maternal smoking status recording during pregnancy in United Kingdom primary care data. *PloS One*, 8(9), e72218. <https://doi.org/10.1371/journal.pone.0072218>

Diamanti-Kandarakis, E., Kouli, C. R., Bergiele, A. T., Filandra, F. A., Tsianateli, T. C., Spina, G. G., ... Bartzis, M. I. (1999). A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. *The Journal of Clinical Endocrinology and Metabolism*, 84(11), 4006–11. <https://doi.org/10.1210/jcem.84.11.6148>

Jiao, J., Fang, Y., Wang, T., Wang, Z., Zhou, M., & Wang, X. (2014). Epidemiologic investigation of polycystic ovarian syndrome (PCOS) in Han ethnic women of reproductive age in Liaoning Province, China. *Clinical and Experimental Obstetrics & Gynecology*, 41(3), 304–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24992782>

Kumarapeli, V., Seneviratne, R. D. a, Wijeyaratne, C. N., Yapa, R. M. S. C., & Dodampahala, S. H. (2008). A simple screening approach for assessing community prevalence and phenotype of polycystic ovary syndrome in a semiurban population in Sri Lanka. *American Journal of Epidemiology*, 168(3), 321–328. <https://doi.org/10.1093/aje/kwn137>

Legro, R. (2015). Diagnosis and treatment of polycystic ovary syndrome (PCOS): an interview with Richard Legro. *BMC Medicine*, 13, 64. <https://doi.org/10.1186/s12916-015-0299-2>

Livadas, S., & Diamanti-Kandarakis, E. (2013). Polycystic ovary syndrome: Definitions, phenotypes and diagnostic approach. *Frontiers of Hormone Research*, 40, 1–21. <https://doi.org/10.1159/000341673>

Lui, K. J. (1988). Estimation of sample sizes in case-control studies with multiple controls per case: dichotomous data. *American Journal of Epidemiology*, 127(5), 1064–70. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3358407>

Michelmores, K. F., Balen, A. H., Dunger, D. B., & Vessey, M. P. (1999). Polycystic ovaries and associated clinical and biochemical features in young women. *Clinical Endocrinology*, 51(6), 779–86. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10619984>

Morgan, C. L., Jenkins-Jones, S., Currie, C. J., & Rees, D. A. (2012). Evaluation of adverse outcome in young women with polycystic ovary syndrome versus matched, reference controls: a retrospective, observational study. *The Journal of Clinical Endocrinology and Metabolism*, 97(9), 3251–60. <https://doi.org/10.1210/jc.2012-1690>

Shivaprakash, G., A. B., Kamath, A., Shivaprakash, P., Adhikari, P., Up, R., ... Padubidri, J. R. (2013). Acanthosis Nigricans in PCOS Patients and Its Relation with Type 2 Diabetes Mellitus and Body Mass at a Tertiary Care Hospital in Southern India. *Journal of Clinical and Diagnostic Research : JC DR*, 7(2), 317–9. <https://doi.org/10.7860/JC DR/2013/4930.2756>

Teede, H., Gibson-Helm, M., Norman, R. J., & Boyle, J. (2014). Polycystic ovary syndrome: perceptions and attitudes of women and primary health care physicians on features of PCOS and renaming the syndrome. *The Journal of Clinical Endocrinology and Metabolism*, 99(1), E107-11. <https://doi.org/10.1210/jc.2013-2978>

Williams, T., van Staa, T., Puri, S., & Eaton, S. (2012). Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Therapeutic Advances in Drug Safety*, 3(2), 89–99. <https://doi.org/10.1177/2042098611435911>

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	(1.1) Abstract, paragraph 2 (1.2) Abstract, paragraph 2 (1.3) Abstract, paragraph 2
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Introduction, paragraphs 1-3
Objectives	3	State specific objectives, including any prespecified hypotheses			Introduction, paragraph 4
Methods					
Study Design	4	Present key elements of study design early in the paper			Methods, paragraph 2
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up,			Methods, paragraph 3

		and data collection			
Participants	6	(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	(6.1.) Methods, paragraph 3 (6.2) NA (6.3) Methods, paragraph 5 Supp. Table S1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported,	(7.1) Methods, paragraph 6 Supp. Fig S1, S3 Table, S4 Table

				an explanation should be provided.	
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Table 1 S1 Table, S2 Table, S5 Table, S7 Table, S11 Table, S13 Table, S15 Table, S17 Table,
Bias	9	Describe any efforts to address potential sources of bias			Methods, paragraph 5 Methods, paragraph 11
Study size	10	Explain how the study size was arrived at			Methods, paragraph 4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Methods, paragraph 9
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> – If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> – If applicable, explain how matching of cases and controls was addressed			(a) Methods, paragraph 10-15 (b) Methods, paragraph 11 (c) Table 1 (d) Results, paragraph 4

		<i>Cross-sectional study</i> – If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			(e) Results, paragraph 4 Table 3
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	(12.1) Methods, paragraph 16 (12.2) Methods, paragraph 16
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (<i>e.g.</i>, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage.		RECORD 13.1: Describe in detail the selection of the persons included in the study (<i>i.e.</i> , study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	(13.1) Results, paragraphs 1-3 Fig.1

		(c) Consider use of a flow diagram			
Descriptive data	14	(a) Give characteristics of study participants (<i>e.g.</i>, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) <i>Cohort study</i> - summarise follow-up time (<i>e.g.</i>, average and total amount)			(a) Results, paragraph 1-3 Table 1 S1 Table, S2 Table, S5 Table, S7 Table, S11 Table, S13 Table, S15 Table, S17 Table, (b) Table 1 and Table 4 S13 Table, S17 Table (c) Results, paragraph 4 Table 2
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> - Report numbers of outcome events or summary measures			Results, paragraph 4 Table 2 S1 Table and S2 Table
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted			(16.a) Results, paragraph 4-12

		estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			Table 2 and 3 Figures 2 and 3 S5 –S10 Tables S13 and S14 Tables S16-S18 Table (b) Results, paragraph 11-12 Tables 1 and 3 S1 Table S2 Table S6 Table (c) Results, paragraph 4-7, 9-12 S1 Table S2 Table S5 – S10 Tables S12-S18 Tables
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			Results, paragraph 7 Figure 2 Table 2 S7 Table
Discussion					
Key results	18	Summarise key results with reference to study objectives			Discussion, paragraph 1-5
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured	Discussion, paragraph 6

				confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	
Interpretation	20	Give a cautious overall interpretation	of	results considering objectives, limitations, multiplicity of analyses, results	Interpretation
from similar studies, and other relevant evidence			Discussion, paragraph 6-8		from similar studies, and other relevant evidence
Other information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the		Funding	22
Accessibility of protocol, raw data, and programming code				RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	
*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Medicine 2015; in press. *Checklist is protected under Creative Commons Attribution (CC BY) license.					

Chapter 4 : PCOS and OSA

Increased risk of obstructive sleep apnoea in women with polycystic ovary syndrome: a population-based cohort study

Increased risk of obstructive sleep apnoea in women with polycystic ovary syndrome: a population-based cohort study

Balachandran Kumarendran^{1,2}, Dana Sumilo¹, Michael W O'Reilly^{3,4}, Konstantinos A Toulis¹, Krishna M Gokhale¹, Chandrika N Wijeyaratne⁵, Arri Coomarasamy^{3,4}, Wiebke Arlt^{3,4}, Abd A Tahrani^{3,4} and Krishnarajah Nirantharakumar^{1,3,4,6}

¹Institute of Applied Health Research, University of Birmingham, Birmingham, UK, ²Department of Community and Family Medicine, Faculty of Medicine, University of Jaffna, Jaffna, Sri Lanka, ³Institute of Metabolism and Systems Research, University of Birmingham, ⁴Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK, ⁵Department of Obstetrics and Gynecology, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka, and ⁶Health Data Research, Birmingham, UK

Correspondence should be addressed to K Nirantharakumar
Email k.nirantharan@bham.ac.uk

Abstract

Objective: Obesity is very common in patients with obstructive sleep apnoea (OSA) and polycystic ovary syndrome (PCOS). Longitudinal studies assessing OSA risk in PCOS and examining the role of obesity are lacking. Our objective was to assess the risk of OSA in women with vs without PCOS and to examine the role of obesity in the observed findings.

Design: Population-based retrospective cohort study utilizing The Health Improvement Network (THIN), UK.

Methods: 76 978 women with PCOS and 143 077 age-, BMI- and location-matched women without PCOS between January 2000 and May 2017 were identified. Hazard ratio (HR) for OSA among women with and without PCOS were calculated after controlling for confounding variables using multivariate Cox models.

Results: Median patient age was 30 (IQR: 25–35) years; median follow-up was 3.5 (IQR: 1.4–7.1) years. We found 298 OSA cases in PCOS women vs 222 in controls, with incidence rates for OSA of 8.1 and 3.3 per 10 000 person years, respectively. Women with PCOS were at increased risk of developing OSA (adjusted HR = 2.26, 95% CI: 1.89–2.69, $P < 0.001$), with similar HRs for normal weight, overweight and obese PCOS women.

Conclusions: Women with PCOS are at increased risk of developing OSA compared to control women irrespective of obesity. Considering the significant metabolic morbidity associated with OSA, clinicians should have a low threshold to test for OSA in women with PCOS. Whether OSA treatment has an impact on PCOS symptoms and outcomes needs to be examined.

European Journal of
Endocrinology
(2019) 180, 265–272

Introduction

Polycystic ovary syndrome (PCOS) is the commonest endocrine disorder in women of reproductive age (prevalence 8–13%) (1). Chronic anovulation, hyperandrogenism and ovarian polycystic morphology are the defining features of PCOS.

PCOS is associated with multiple comorbidities including obesity, insulin resistance (IR), dyslipidaemia,

gestational diabetes (GDM), type 2 diabetes (T2D), hypertension, non-alcoholic fatty liver disease, impaired quality of life (QOL), cardiovascular disease (CVD) and mortality amongst others (2, 3, 4). Weight loss (by lifestyle intervention or bariatric surgery) remains the only specific treatment for PCOS. The remaining treatments are mainly symptomatic (oral contraceptives

for irregular periods, anti-androgens for hirsutism, ovulation induction for infertility) (5). Hence, there is a need for better understanding of the pathogenesis of PCOS-related metabolic risk and comorbidities in order to develop effective therapies (6).

Obstructive sleep apnoea (OSA) is common, affecting 17–26% of men and 9–28% of women; the prevalence being lower in women of reproductive age compared to men (7). OSA is characterized by instability in the upper airways during sleep, leading to recurrent upper airway obstructions, sleep architecture disruption and cyclical changes in heart rate, blood pressure, sympathetic activity, intrathoracic pressure and oxygen saturations (7, 8). Obesity is very common in patients with OSA and PCOS (9). In addition, OSA is associated with similar comorbidities to PCOS, such as IR, GDM, T2D, hypertension, impaired QOL, CVD and mortality (7, 10). Hence, it is plausible that OSA and PCOS might co-exist and that either condition could contribute to the comorbidities of the other (9). This is further supported by a recent systematic review that showed that OSA prevalence in women with PCOS was 32% (95% CI: 13–55%) (11). However, these studies were of small size ($n < 60$), at risk of selection bias and cross-sectional design, barring the determination of the direction of relationship. Additionally, the study populations exclusively comprised patients with grade II obesity or higher. Hence, there is a need to examine the relationship between OSA and PCOS in a longitudinal population-based study, which allows to assess the impact of obesity.

Several mechanisms, other than obesity and insulin resistance, might increase the likelihood of OSA in women with PCOS compared to women without PCOS of similar adiposity including hyperandrogenism, low progesterone (due to anovulation) and increased oxidative stress (9).

We hypothesized that women with PCOS are at increased risk of developing OSA compared to women without PCOS regardless of the degree of obesity. The primary aim of this study was to assess the risk of incident OSA in women with PCOS vs women without PCOS and examine the role of obesity in the observed relationships. A secondary aim was to assess predictors of incident OSA in women with PCOS.

Subjects and methods

For full description of the methods please refer to the online supplement.

Ethics statement

The use of THIN data for research was approved by the South-East Multicenter Research Ethics Committee in 2003 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).

Study design and setting

We conducted a matched-controlled retrospective cohort study using data from UK general practices contributing to The Health Improvement Network (THIN) electronic database, as we have in previous studies (2, 12, 13).

Study population

The open cohort extended from the 1st January 2000 (study start date) to the 15th May 2017 (study end date).

Inclusion criteria: All women who were aged 18–50 years at the index date (study entry) and had a documentation of PCOS at any time during the study period were included in the exposed group. Patients with any documentation of OSA prior to the index date were excluded. Women without documented PCOS at any time during the study period were included in the unexposed (control) arm. The index date was defined as the date of first documentation of PCOS for newly diagnosed cases and from the date patient became eligible if the first documentation of PCOS was prior to the eligibility date (for existing cases) (Supplementary Fig. E1, see section on [supplementary data](#) given at the end of this article).

Each exposed patient was randomly matched to two unexposed patients (1:2 ratio) for general practice, age at index date and BMI (14, 15).

To minimize the immortal time bias, each randomly matched eligible unexposed patient was assigned the same index date as their corresponding exposed patient (16). Follow-up end date (exit date) was determined from the earliest occurrence of the first documentation of OSA, transfer to another practice, death or study end.

Selection of Read Codes and PCOS definition

Read Codes to define PCOS, OSA and covariates were compiled using a methodical Read Code search strategy (see Supplementary Panel E1 and Table E1) Since there is a possibility of misclassification between PCOS and polycystic ovaries (PCO) due to the resemblance of codes

during data entry, they have been combined in prevalence studies using general practice electronic databases (17).

Statistical analysis

Potential confounders and covariates were chosen based on biological plausibility and links to the exposure and outcome of interest (PCOS and OSA respectively). These included age, Townsend social deprivation index (14), BMI, smoking status, diabetes mellitus, impaired glucose regulation (18), hypertension (19), hypothyroidism, antiandrogen medication and metformin. The critical value for statistical significance was set at 5% and 95% confidence intervals were used in the population estimate of hazard ratios. Further analysis included calculating the number and percentage of incident cases, person years and incidence rates. Unadjusted and adjusted hazard ratios were estimated using Cox regression models.

Sensitivity analysis was carried out to assess selection bias due to case definition (PCOS and PCO vs PCOS only) and survival bias due to inclusion of prevalent cases (who had the documentation of PCOS prior to becoming eligible for the study) (20). In the primary analysis, we did not consider ethnicity as a covariate due to high missing values (48.9%). In a sensitivity analysis we assigned a separate ethnicity category for missing values and included into the Cox model as an additional covariate.

Another model confined to PCOS cases (excluding controls) was used to assess the association of PCOS with OSA while considering PCOS phenotypes and antiandrogen medication as covariates. STATA MP version 14.2 was used for data cleaning and analysis (21).

Results

Study population characteristics

The study population included 76 978 women with PCOS and 143 077 matched control women from 763 general practices registered with THIN (Fig. 1). The median follow-up was 3.5 years (IQR: 1.38 to 7.14) with no significant difference between women with and without PCOS (Fig. 1). There was no obvious difference between women with and without PCOS with regards to age (median 30 years, IQR: 25–35), Townsend index for social deprivation, BMI and smoking status (Table 1). When compared to controls, women with PCOS were more likely to have T2D (2.2 vs 1.0%), hypertension (3.0 vs 2.0%), hypothyroidism (3.9 vs 2.3%) and impaired glucose

regulation (0.6% vs 0.3%) (Table 1). The PCOS group was also more likely to develop diabetes mellitus during the follow-up when compared to control group (5.6 vs 2.6%).

The clinical features of PCOS and medications used in the study cohort are summarized in Table 2. Of the 76 978 women with PCOS, 22 307 (29.0%) were defined using PCOS Read Code while the remainder were defined using PCO Read Code (see Supplementary Table E1).

Primary analysis

There were 298 incident cases of OSA among 76 978 women with PCOS and 222 among 143 077 women without PCOS. Incidence of OSA was significantly higher in women with PCOS vs the matched controls (8.1 vs 3.3 per 10 000 person-years of follow-up, $P < 0.0001$) (Table 3). Women with PCOS had an increased hazard of developing OSA when compared to their matched

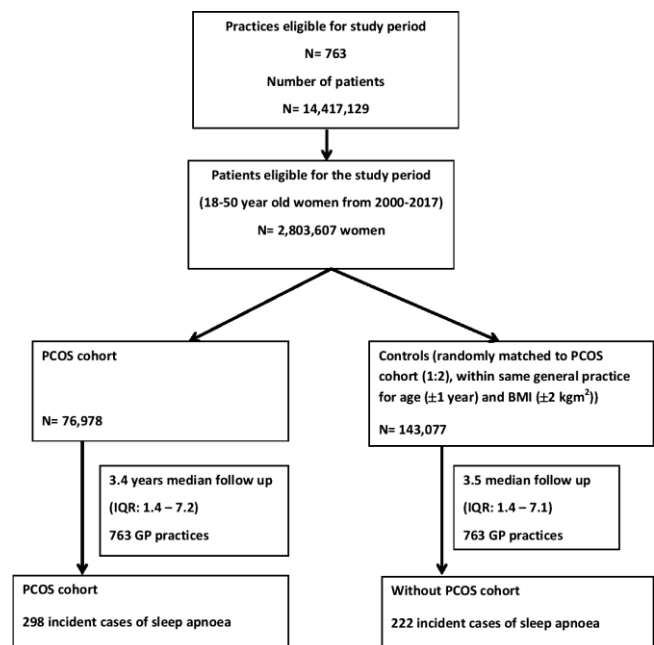


Figure 1

Selection of patients with polycystic ovary syndrome (PCOS) and controls. A total of 14 417 129 patients were registered with 763 eligible practices during 2000 to 2017. Of these, 2 803 607 women were of 18 to 50 years old and were eligible. According to the criteria indicated in 'methods' section, there were 76 978 women with PCOS in PCOS cohort. The control cohort had 143 077 women who were matched within same general practice for age and BMI at a ratio of 1:2. There were 298 and 222 incident cases of sleep apnoea in PCOS and control cohorts respectively.

Clinical Study	B Kumarendran and others	PCOS and OSA	180:4	268

Table 1 Baseline characteristics of women with PCOS and control women without PCOS.

Characteristics	With PCOS (n = 76 978)	Without PCOS (n = 143 077)
Age (years; mean (s.d.))	30.2 (7.4)	30.4 (7.3)
Townsend score*		
1	13 865 (18.0%)	25 951 (18.1%)
2	12 716 (16.5%)	23 910 (16.7%)
3	15 072 (19.6%)	28 172 (19.7%)
4	14 498 (18.8%)	27 341 (19.1%)
5	10 223 (13.3%)	19 675 (13.8%)
Missing data	10 604 (13.8%)	18 028 (12.6%)
BMI (kg/m ² ; mean (s.d.))	28.6 (7.6)	27.4 (6.4)
BMI categorized		
<25 kg/m ²	25 739 (33.4%)	52 666 (36.8%)
25–29.99 kg/m ²	15 065 (19.6%)	29 682 (20.8%)
≥30 kg/m ²	24 139 (31.4%)	37 585 (26.3%)
Missing or implausible data	12 035 (15.6%)	23 144 (16.2%)
Smoking status		
Non-smokers	56 105 (72.9%)	101 370 (70.8%)
Smokers	17 383 (22.6%)	32 778 (22.9%)
Missing or implausible data	3490 (4.5%)	8929 (6.2%)
Medical conditions at baseline		
Diabetes	1694 (2.2%)	1449 (1.0%)
Hypertension	2325 (3.0%)	2805 (2.0%)
Hypothyroidism	2987 (3.9%)	3228 (2.3%)
Impaired glucose regulation†	446 (0.6%)	368 (0.3%)
Ethnicity		
Caucasian-European	36 670 (47.6%)	58 182 (40.7%)
Black-Afro-Caribbean	1692 (2.2%)	3222 (2.3%)
South Asians	3929 (5.1%)	4266 (3.0%)
Others including Chinese, Middle-Eastern	1013 (1.3%)	1781 (1.2%)
Mixed-race	715 (0.9%)	1008 (0.7%)
Missing (no codes found in Medical or AHD file)	32 959 (42.8%)	74 618 (52.2%)

*Townsend score - presented as quintiles with 1 least deprived and 5 most deprived. †(Includes impaired fasting glucose and impaired glucose tolerance).

controls (HR = 2.46, 95% CI: 2.07–2.93, $P < 0.001$). Women with PCOS remained at increased risk of developing OSA compared to women without PCOS following adjustment for age, Townsend score, BMI, hypothyroidism at baseline, baseline and incident diabetes/IGR (adjusted HR = 2.26, 95% CI: 1.89 to 2.69, $P < 0.001$) (Table 3).

In addition to PCOS, we identified older age, increasing BMI and the presence of hypothyroidism at baseline as independent predictors of incident OSA (see Supplementary Table E2).

Sensitivity analysis

Limiting the analysis to incident cases of PCOS (adjusted HR = 2.20, 95% CI: 1.58–3.06, $P < 0.001$) or PCOS Read Codes only (adjusted HR = 2.93, 95% CI: 2.15–3.99, $P < 0.001$) resulted in similar point estimates to those found in the analysis of the whole study population (Table 3). Adding ethnicity as an additional covariate did not alter our findings (adjusted HR 2.21, 95% CI: 1.85–2.64, $P < 0.001$).

Analysis to assess if risk of OSA among women with PCOS is independent of BMI

Increasing BMI was associated with increasing OSA incidence in women with and without PCOS, but the HRs were greater in women with PCOS vs control in all BMI categories (see Fig. 2 and Supplementary Table E2). When compared to women without PCOS, women with PCOS showed a higher risk (adjusted HR (95% CI)) of developing OSA in all three BMI (kg/m²) categories (BMI <25: 1.91 (0.92–3.97), $P = 0.081$; BMI 25–29.99: 2.25 (1.33–3.81), $P = 0.003$; BMI ≥30: 2.10 (1.72–2.56), $P < 0.001$) (see Fig. 2 and Supplementary Table E3).

Analysis to assess the role of factors other than BMI in OSA among PCOS women

In addition to older age, increasing BMI and presence of hypothyroidism, we identified anovulation (adjusted HR (95% CI) 1.33 (1.05–1.69), $P = 0.02$), hirsutism

Clinical Study	B Kumarendran and others	PCOS and OSA	180:4	268

Table 2 PCOS diagnostic features and use of medication. Data are presented as *n* (%)

Characteristics	With PCOS (<i>n</i> = 76 978)	Without PCOS (<i>n</i> = 143 077)
PCOS features at baseline		
Anovulation	20 106 (26.1)	10 463 (7.3)
Any clinical androgen excess feature	21 974 (28.6)	23 231 (16.3)
Acne	14 589 (19.0)	19 371 (13.5)
Alopecia	2749 (3.6)	3820 (2.7)
Hirsutism	7748 (10.1)	1269 (0.9)
Polycystic ovaries	58 099 (75.5)	0 (0)
Medications		
Oral Contraceptive Pill (OCP) including an anti-androgenic progestin component*	22 798 (29.6)	17 005 (11.9)
Cyproterone	17 346 (22.5)	9381 (6.6)
Drospirenone	9685 (12.6)	10 280 (7.2)
Other antiandrogen	192 (0.2)	14 (0.0)
Metformin	18 571 (24.1)	2925 (2.0)

*Use of either 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.

(1.37 (1.05–1.80), $P = 0.02$) and treatment with metformin (1.69, 1.28–2.22, $P < 0.001$) as predictors of incident OSA (Table 4).

Discussion

This is the first large longitudinal study assessing the temporal relationship between PCOS and OSA in a European population and that accounted for the role of obesity and adjusted for a large number of potential confounders. We found that women with PCOS were at increased risk of developing OSA compared women without PCOS regardless of the severity of obesity and independent of a wide range of potential confounders. We also found that hirsutism and chronic anovulation, clinical features of androgen excess, which is closely

linked to insulin resistance in the context of PCOS, were independent predictors of OSA development in women with PCOS.

A recent study from Taiwan also found an increase OSA incidence in women with PCOS, but was of much smaller sample size ($n = 9190$ vs 220 055) and did not match for BMI; in addition, the adjustment for confounders was limited and did not explore the possible underlying mechanisms (22).

The prevalence of OSA in women with PCOS has been investigated in several small-scale cross-sectional studies (9). Two recent meta-analyses showed an estimated OSA prevalence (based on polysomnography or portable level II devices) of 36.1% (95% CI: 22.4–51.0%) and 32% (95% CI: 13–55%) in women with PCOS (23, 24). However, the prevalence of OSA in women with PCOS varied widely from 0 to 69% (9). This big variation in OSA prevalence

Table 3 Risk of women with PCOS developing OSA compared to women without PCOS.

	Primary analysis		Sensitivity analysis (incident cases)		Sensitivity analysis (PCOS specific Read Codes only)	
	Exposed	Unexposed	Exposed	Unexposed	Exposed	Unexposed
Total number of patients	76 978	143 077	23 349	43 106	24 603	44 991
Incident OSA <i>n</i> (%)	298 (0.39)	222 (0.16)	86 (0.37)	60 (0.14)	115 (0.47)	63 (0.14)
Person years	368 203	680 011	112 290	193 646	107 959	187 801
Incidence rates per 10 000 person years	8.1	3.3	7.7	3.1	10.7	3.4
Hazard ratio (95% CI)	2.46 (2.07–2.93)		2.44 (1.76–3.40)		3.11 (2.29–4.23)	
<i>P</i> -value	<0.001		<0.001		<0.001	
Adjusted hazard ratio (95% CI)*	2.26 (1.89–2.69)		2.20 (1.58–3.06)		2.93 (2.15–3.99)	
<i>P</i> -value	<0.001		<0.001		<0.001	
Adjusted hazard ratio (95% CI)†	2.23 (1.87–2.66)		2.14 (1.53–2.99)		2.87 (2.10–3.92)	
<i>P</i> -value	<0.001		<0.001		<0.001	

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

Clinical Study	B Kumarendran and others	PCOS and OSA	180:4	268

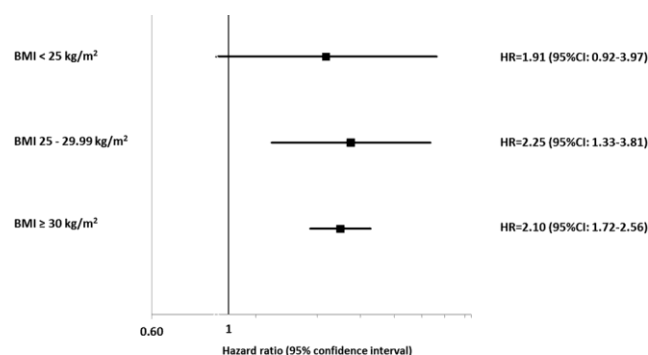


Figure 2

Impact of BMI on the hazard of obstructive sleep apnoea (OSA). Women in both PCOS and control cohorts were subcategorized into lean, overweight and obese groups. The hazard of developing OSA among PCOS cohort in comparison to control cohort was assessed for each BMI sub groups.

between studies is attributed to multiple factors, including the application of different AHI cut-offs to diagnose OSA, small sample sizes, variation in obesity prevalence and potential selection bias due to recruitment from specialized clinics, raising the possibility that the results may not be generalizable (9). Our current study addresses many of these gaps by conducting a population-based longitudinal study of a large sample size that included patients with normal, overweight and obese BMI categories. The impact of PCOS on risk of OSA was stronger in overweight than in obese women; however, this needs to be interpreted with caution given the confidence intervals of effect sizes overlap.

The cross-sectional studies that showed low prevalence rates of OSA in women with PCOS were mainly carried out in leaner cohorts, hence suggesting that the high prevalence of OSA in women with PCOS was only applicable to women with obesity. Whether women with PCOS were truly at increased risk of OSA compared to women with similar obesity but without PCOS was unclear (9, 23, 24). Our study addressed this limitation of earlier studies by including a control group of women matched for BMI amongst other important covariates, and our results show clearly that women with PCOS were at increased risk OSA compared to women without PCOS. In addition, we have shown that this increased risk of OSA in women with PCOS was evident irrespective of whether patients were normal weight, overweight or have obesity. However, the observed relationships were stronger in women who were overweight or had obesity.

Several mechanisms might explain the link between PCOS and OSA and explain the findings of this study

(9, 25). Androgen excess is a cardinal feature of PCOS and androgens are implicated in the pathogenesis of OSA as they affect upper airway stability and ventilatory drive (9). OSA is consistently more common in men than women (26, 27), which may also point to androgens as a contributing factor. Our data support this notion as a documented diagnosis of hirsutism, a clinical manifestation of androgen excess, was an independent predictor of incident OSA. In addition to androgen excess, oxidative stress and low progesterone might play a role in the observed relationship between PCOS and OSA (25). Oxidative stress can contribute to the pathogenesis of OSA by causing dysfunction in the carotid body chemoreceptors resulting in ventilatory instability (28). Progesterone is a respiratory stimulant and can result in lowering upper respiratory resistance (29). Furthermore, OSA is associated with lower progesterone in women (30) and low progesterone affects the ventilatory drive (9, 31, 32). Our results also support this potential mechanism as anovulation predicted incident OSA. Insulin resistance, a feature of OSA and a predictor of incident apnoea can also contribute to the observed relationships in this study (33, 34).

OSA is associated with multiple comorbidities including road traffic accidents, hypertension, T2D, CVD,

Table 4 Predictors of OSA in women with PCOS (n = 76 978).

Covariates	Hazard ratio	95% confidence interval	P value
Age	1.06	(1.04-1.07)	<0.001
Townsend			
1	1.00	(Reference level)	
2	1.26	(0.83-1.89)	0.27
3	1.41	(0.95-2.08)	0.08
4	1.58	(1.07-2.33)	0.02
5	1.67	(1.10-2.53)	0.02
Missing	1.23	(0.77-1.97)	0.38
BMI category			
<25 kg/m ²	1.00	(Reference level)	
25-29.99 kg/m ²	3.01	(1.60-5.69)	0.001
≥30 kg/m ²	12.50	(7.24-21.58)	<0.001
Missing	3.64	(1.91-6.95)	<0.001
Diabetes or IGR*	0.72	(0.44-1.18)	0.19
Hypothyroidism	1.76	(1.20-2.58)	0.004
Anovulation	1.33	(1.05-1.69)	0.02
Androgen excess features			
Hirsutism	1.37	(1.05-1.80)	0.02
Acne	0.85	(0.61-1.19)	0.35
Alopecia	0.97	(0.64-1.46)	0.89
Metformin	1.69	(1.28-2.22)	<0.001
Antiandrogen drug	0.77	(0.57-1.05)	0.10

*Impaired glucose response (includes impaired fasting glucose and impaired glucose tolerance). Values in boldface indicate statistical significance.

and mortality, which are similar to PCOS comorbidities (except the road traffic accidents). However, whether OSA increases the risk of these comorbidities in women with PCOS is largely unknown. A recent systematic review and meta-analysis showed that women with both PCOS and OSA are more likely to have hirsutism, hypertension, insulin resistance, hyperlipidaemia, and dysglycaemia than women with PCOS only (24). However, these findings were based on small-scale cross-sectional studies with a high risk of bias (9). Hence, longitudinal and interventional studies to assess the temporal relationship and causation are needed.

The findings of our study need to be interpreted within the context of the limitations. We carried out an observational cohort study using data from primary care and hence we did not have access to the results of the polysomnography, polygraphy or oximetry such as the AHI, oxygen desaturation index and other hypoxemic measures. In addition, since we and other studies (14, 17) have observed a much lower prevalence of PCOS and OSA in electronic medical records, it is possible that many patients with PCOS and OSA remained undiagnosed. This may be due lack of diagnosis or incomplete recording of PCOS and OSA diagnosis in primary care setting. It is possible that patients with PCOS who come to the attention of the treating clinician represented the more severe phenotype which may have resulted in an overestimate of the effect size. This may limit the generalizability of our findings to patients with milder form of disease. On the other hand, the unexposed population may have patients with undiagnosed PCOS which could potentially lead to an underestimation of the effect size.

Nevertheless, our study has several strengths including the large sample size, the longitudinal nature of the study, the matched controlled design and the adjustment for a wide range of confounders including BMI. In addition, this was a population-based study rather than data from tertiary centres and hence the findings are much more generalizable. The longitudinal design of the study also allowed us to assess the predictors of OSA in women with PCOS.

We conclude that women with PCOS are at increased risk of developing OSA compared to women without PCOS independent of obesity and other potential cofounders. This increased risk was present in patients with normal weight, overweight and obesity. Clinicians treating women with PCOS should have a low threshold to test for OSA. In addition, our findings suggest that clinical features of hyperandrogenism and low progesterone may contribute to the increased risk of OSA in women with PCOS. Whether OSA worsens the long-term sequelae

of PCOS and whether CPAP treatment has favourable impacts on women with PCOS and OSA remains to be examined.

Supplementary data

This is linked to the online version of the paper at <https://doi.org/10.1530/EJE-18-0693>.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this study.

Funding

A A T is a clinician scientist supported by the National Institute for Health Research (NIHR) in the UK (CS-2013-13-029); A C is an NIHR Senior Investigator. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health. W A is supported by the Wellcome Trust (Investigator Award in Science, 209492/Z/17/Z).

Acknowledgements

A A Tahrani and K Nirantharakumar: Joint senior authors, contributed equally to this work.

References

- Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ & International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertility and Sterility* 2018 **110** 364–379. (<https://doi.org/10.1016/j.fertnstert.2018.05.004>)
- Kumarendran B, O'Reilly MW, Manolopoulos KN, Toulis KA, Gokhale KM, Sitch AJ, Wijeyaratne CN, Coomarasamy A, Arlt W & Nirantharakumar K. 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* 2018 **15** e1002542. (<https://doi.org/10.1371/journal.pmed.1002542>)
- Legro RS, Arslanian SA, Ehrmann DA, Hoeger KM, Murad MH, Pasquali R, Welt CK & Endocrine Society. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** 4565–4592. (<https://doi.org/10.1210/jc.2013-2350>)
- Randeva HS, Tan BK, Weickert MO, Lois K, Nestler JE, Sattar N & Lehnert H. Cardiometabolic aspects of the polycystic ovary syndrome. *Endocrine Reviews* 2012 **33** 812–841. (<https://doi.org/10.1210/er.2012-1003>)
- Charalampakis V, Tahrani AA, Helmy A, Gupta JK & Singhal R. Polycystic ovary syndrome and endometrial hyperplasia: an overview of the role of bariatric surgery in female fertility. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 2016 **207** 220–226. (<https://doi.org/10.1016/j.ejogrb.2016.10.001>)
- O'Reilly MW, Kempegowda P, Walsh M, Taylor AE, Manolopoulos KN, Allwood JW, Semple RK, Hebenstreit D, Dunn WB, Tomlinson JW *et al.* AKR1C3-mediated adipose androgen generation drives lipotoxicity in women With polycystic ovary

- syndrome. *Journal of Clinical Endocrinology and Metabolism* 2017 **102** 3327–3339. (<https://doi.org/10.1210/jc.2017-00947>)
- 7 Tahrani AA. Obstructive sleep apnoea in diabetes: does it matter? *Diabetes and Vascular Disease Research* 2017 **14** 454–462. (<https://doi.org/10.1177/1479164117714397>)
 - 8 Kang K, Seo JG, Seo SH, Park KS & Lee HW. Prevalence and related factors for high-risk of obstructive sleep apnea in a large Korean population: results of a questionnaire-based study. *Journal of Clinical Neurology* 2014 **10** 42–49. (<https://doi.org/10.3988/jcn.2014.10.1.42>)
 - 9 Kahal H, Kyrou I, Tahrani AA & Randeva HS. Obstructive sleep apnoea and polycystic ovary syndrome: a comprehensive review of clinical interactions and underlying pathophysiology. *Clinical Endocrinology* 2017 **87** 313–319. (<https://doi.org/10.1111/cen.13392>)
 - 10 Wang X, Ouyang Y, Wang Z, Zhao G, Liu L & Bi Y. Obstructive sleep apnea and risk of cardiovascular disease and all-cause mortality: a meta-analysis of prospective cohort studies. *International Journal of Cardiology* 2013 **169** 207–214. (<https://doi.org/10.1016/j.ijcard.2013.08.088>)
 - 11 Helvaci N, Karabulut E, Demir AU & Yildiz BO. Polycystic ovary syndrome and the risk of obstructive sleep apnea: a meta-analysis and review of the literature. *Endocrine Connections* 2017 **6** 437–445. (<https://doi.org/10.1530/EC-17-0129>)
 - 12 Caleyachetty R, Thomas GN, Toulis KA, Mohammed N, Gokhale KM, Balachandran K & Nirantharakumar K. Metabolically healthy obese and incident cardiovascular disease events Among 3.5 million men and women. *Journal of the American College of Cardiology* 2017 **70** 1429–1437. (<https://doi.org/10.1016/j.jacc.2017.07.763>)
 - 13 Toulis KA, Willis BH, Marshall T, Kumarendran B, Gokhale K, Ghosh S, Thomas GN, Cheng KK, Narendran P, Hanif W *et al.* All-cause mortality in patients with diabetes under treatment with dapagliflozin: a population-based, open-cohort study in the health improvement network database. *Journal of Clinical Endocrinology and Metabolism* 2017 **102** 1719–1725. (<https://doi.org/10.1210/jc.2016-3446>)
 - 14 Wall H, Smith C & Hubbard R. Body mass index and obstructive sleep apnoea in the UK: a cross-sectional study of the over-50s. *Primary Care Respiratory Journal* 2012 **21** 371–376. (<https://doi.org/10.4104/pcrj.2012.00053>)
 - 15 Lim SS, Davies MJ, Norman RJ & Moran LJ. Overweight, obesity and central obesity in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction Update* 2012 **18** 618–637. (<https://doi.org/10.1093/humupd/dms030>)
 - 16 Toulis KA, Willis BH, Marshall T & Gokhale K, Ghosh S, Thomas GN, Cheng KK, Narendran P, Hanif W & Nirantharakumar K. All-cause mortality in patients with diabetes under treatment with dapagliflozin: a population-based, open-cohort study in THIN database. *Journal of Clinical Endocrinology and Metabolism* 2017 **102** 1719–1725. (<https://doi.org/10.1210/jc.2016-3446>)
 - 17 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** e012461. (<https://doi.org/10.1136/bmjopen-2016-012461>)
 - 18 Moran LJ, Misso ML, Wild RA & Norman RJ. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction Update* 2010 **16** 347–363. (<https://doi.org/10.1093/humupd/dmq001>)
 - 19 McEvoy RD. Obstructive sleep apnoea and hypertension: the ESADA study. *European Respiratory Journal* 2014 **44** 835–838. (<https://doi.org/10.1183/09031936.00129614>)
 - 20 Matok I, Azoulay L, Yin H & Suissa S. Immortal time bias in observational studies of drug effects in pregnancy. *Birth Defects Research. Part A, Clinical and Molecular Teratology* 2014 **100** 658–662. (<https://doi.org/10.1002/bdra.23271>)
 - 21 StataCorp. *Stata Statistical Software: Release*. College Station, TX: StataCorp LP, 2015. (<https://doi.org/10.2307/2234838>)
 - 22 Lin TY, Lin PY, Su TP, Li CT, Lin WC, Chang WH, Chen TJ, Bai YM & Chen MH. Risk of developing obstructive sleep apnea among women with polycystic ovarian syndrome: a nationwide longitudinal follow-up study. *Sleep Medicine* 2017 **36** 165–169. (<https://doi.org/10.1016/j.sleep.2016.12.029>)
 - 23 Helvaci N, Karabulut E, Demir AU & Yildiz BO. Polycystic ovary syndrome and the risk of obstructive sleep apnea: a meta-analysis and review of the literature. *Endocrine Connections* 2017 **6** 437–445. (<https://doi.org/10.1530/EC-17-0129>)
 - 24 Kahal H, Kyrou I, Uthman O, Brown A, Johnson S, Wall P, Metcalfe A, Tahrani AA & Randeva HS. The association between obstructive sleep apnoea and metabolic abnormalities in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Sleep* 2018 **41** 1–12. (<https://doi.org/10.1093/sleep/zsy085>)
 - 25 Nitsche K & Ehrmann DA. Obstructive sleep apnea and metabolic dysfunction in polycystic ovary syndrome. *Best Practice and Research. Clinical Endocrinology and Metabolism* 2010 **24** 717–730. (<https://doi.org/10.1016/j.beem.2010.08.001>)
 - 26 Kapsimalis F & Kryger MH. Gender and obstructive sleep apnea syndrome, part 2: mechanisms. *Sleep* 2002 **25** 499–506.
 - 27 Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society* 2008 **5** 136–143. (<https://doi.org/10.1513/pats.200709-155MG>)
 - 28 Fung ML. Pathogenic roles of the carotid body inflammation in sleep apnea. *Mediators of Inflammation* 2014 **2014** 354279. (<https://doi.org/10.1155/2014/354279>)
 - 29 Behan M & Wenninger JM. Sex steroidal hormones and respiratory control. *Respiratory Physiology and Neurobiology* 2008 **164** 213–221. (<https://doi.org/10.1016/j.resp.2008.06.006>)
 - 30 Lee J, Eklund EE, Lambert-Messerlian G, Palomaki GE, Butterfield K, Curran P & Bourjeily G. Serum progesterone levels in pregnant women with obstructive sleep apnea: a case control study. *Journal of Women's Health* 2017 **26** 259–265. (<https://doi.org/10.1089/jwh.2016.5917>)
 - 31 Ruchała M, Bromińska B, Cyrańska-Chyrek E, Kuźnar-Kamińska B, Kostrzevska M & Batura-Gabryel H. Obstructive sleep apnea and hormones – a novel insight. *Archives of Medical Science* 2017 **13** 875–884. (<https://doi.org/10.5114/aoms.2016.61499>)
 - 32 Driver HS, McLean H, Kumar DV, Farr N, Day AG & Fitzpatrick MF. The influence of the menstrual cycle on upper airway resistance and breathing during sleep. *Sleep* 2005 **28** 449–456. (<https://doi.org/10.1093/sleep/28.4.449>)
 - 33 Ehrmann DA. Metabolic dysfunction in PCOS: relationship to obstructive sleep apnea. *Steroids* 2012 **77** 290–294. (<https://doi.org/10.1016/j.steroids.2011.12.001>)
 - 34 Vgontzas AN, Legro RS, Bixler EO, Grayev A, Kales A & Chrousos GP. Polycystic ovary syndrome is associated with obstructive sleep apnea and daytime sleepiness: role of insulin resistance 1. *Journal of Clinical Endocrinology and Metabolism* 2001 **86** 517–520. (<https://doi.org/10.1210/jcem.86.2.7185>)

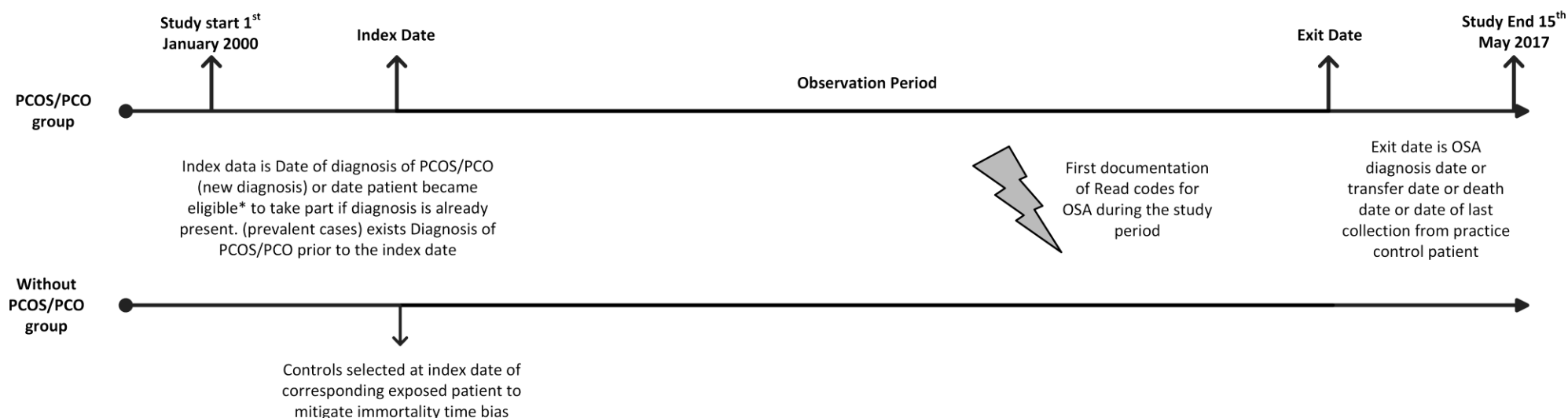
Received 17 August 2018

Revised version received 7 January 2019

Accepted 13 February 2019

Appendices of Chapter 4 (PCOS and OSA)

Visual representation of study timeline and selection process (Identification of the exposed group (individuals with PCOS/PCO Read codes) and their controls (without PCOS/PCO) matched 1:2 by age at index date, BMI and general practice)



* To ensure good quality data, a patient is eligible to take part one year after the latest of the following dates: 1) registration with the general practice (registration date); 2) introduction of VISION Electronic Medical Record (VISION date); and 3) Acceptable Mortality Recording (AMR) date. AMR is an indicator when practices started to record information consistently and in a timely manner^{1,2}. One year latent period is applied to ensure there was sufficient time to record all important covariates.

1. Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf* 2009; **18**(1): 76-83.

2. Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf* 2013; **22**(1): 64-9.

Figure E.1: Visual presentation of study selection process

Detailed methodology

Study design and setting

We conducted a matched-controlled retrospective cohort study using data from UK general practices contributing to The Health Improvement Network (THIN) electronic database, that was used previously in multiple studies of similar design to the current study [1–3]. It contains longitudinal records of clinical, laboratory and therapeutic information of each encounter with the patients registered with these practices using a hierarchical clinical ‘Read Code’ system [4] THIN covers around 6% of UK population (3.7 million) from 763 practices representatively scattered around the UK [5] The generalizability of THIN has been validated for pharmaco-epidemiological research [6] and THIN has been used previously in studies examining PCOS [7] and OSA [8] .

Study population

The open cohort extended from the 1st January 2000 (study start date) to the 15th May 2017 (study end date). All individuals were required to be registered at their practice at least for a year before entry into the study in order to ascertain documentation of concomitant diseases. Their practice was also required to have been using their computer system (Vision) for at least one year prior to the index date (the date at which each participant joined the cohort) and have an acceptable mortality reporting in order to ensure good data quality [9].

All women who fulfilled the above-mentioned criteria and who were aged 18-50 years at the index date and had a documentation of PCOS at any time during the study period were included in the exposed group. Patients with any documentation of OSA prior to the index date were excluded. Women without documented PCOS at any time during the study period were included in the unexposed (control) arm. The index date was defined as the date of first documentation of PCOS for newly diagnosed cases and from the date patient became eligible if the first documentation of PCOS was prior to the eligibility date (for existing cases) (see Figure E1 in the online data supplement.).

Each exposed patient was randomly matched to 2 unexposed patients (1:2 ratio) for general practice, age at index date ($\pm$ one year) and BMI ($\pm$ 2 kg/m²). Matching variables were chosen based on their strong association with PCOS and OSA [8] [10]. When more than 2 participants in the unexposed cohort were available to be matched with a participant in the exposed cohort, two were randomly selected.

To minimize the immortal time bias, each randomly matched eligible unexposed patient was assigned the same index date as their corresponding exposed patient [11]. Follow up end date (exit date) was determined from the earliest occurrence of: the first documentation of OSA, transfer to another practice, death or study-end.

Selection of Read Codes and PCOS definition

Read Codes to define PCOS, OSA and covariates (see Tables E1 to E17 in the online data supplement.) were compiled using a methodical Read Code search strategy (see Panel E1 in the online data supplement.) [12] including codes used in other studies [7] [8]. Since there is a possibility of misclassification between PCOS and Polycystic ovaries (PCO) due to the resemblance of codes during data entry, they have been combined in prevalence studies using general practice electronic databases [7].

Grouping of quantitative variables

BMI (in kg/m²) was categorized into normal (<25), overweight (25 to 29.99) and obese (≥ 30) [13]. Impaired glucose regulation (IGR) included 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 120min after ingestion of 75g glucose in the oral glucose tolerance test).

Statistical analysis

Potential confounders and covariates were chosen based on biological plausibility and links to the exposure and outcome of interest (PCOS and OSA respectively). These included: age, Townsend social deprivation index [8], BMI, smoking status, diabetes mellitus, impaired

glucose regulation [14] hypertension [15], hypothyroidism, antiandrogen medication, and metformin. Baseline characteristics were summarized considering the distribution of data and without statistical testing to comply with the statistical guidelines [16] [17] . The critical value for statistical significance was set at 5% and therefore 95% confidence intervals were used in the population estimate of hazard ratios. Further analysis included calculating the number and percentage of incident cases, person years, and incidence rates. Unadjusted and adjusted hazard ratios were estimated using Cox regression models. The initial adjusted model included age, Townsend score, BMI, diabetes or IGR and hypothyroidism at baseline as covariates. Sensitivity analysis was carried out to assess selection bias due to case definition (PCOS and PCO vs PCOS only) and survival bias due to inclusion of prevalent cases (who had the documentation of PCOS prior to becoming eligible for the study) [18]. Another model confined to PCOS cases (excluding controls) was used to assess the association of PCOS with OSA while considering PCOS phenotypes and antiandrogen medication as covariates. STATA MP version 14.2 was used for data cleaning and analysis [19] 21.

References

1. Kumarendran B, O'Reilly MW, Manolopoulos KN, Toulis KA, Gokhale KM, Sitch AJ, Wijeyaratne CN, Coomarasamy A, Arlt W, Nirantharakumar K. 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.* 2018;15(3).
2. Caleyachetty R, Thomas GN, Toulis KA, Mohammed N, Gokhale KM, Balachandran K, Nirantharakumar K. Metabolically Healthy Obese and Incident Cardiovascular Disease Events Among 3.5 Million Men and Women. *J Am Coll Cardiol* [Internet]. 2017 [cited 2017 Sep 12];70(12). Available from: http://www.onlinejacc.org/content/70/12/1429?sso=1&sso_redirect_count=1&access_token=
3. Toulis KA, Willis BH, Marshall T, Gokhale K, Frpc SG, Thomas GN. All-cause mortality in patients with diabetes under treatment with dapagliflozin: a population-based, open-cohort study in THIN database. *J Clin Endocrinol Metab* [Internet]. 2017; Available from: <https://academic.oup.com/jcem/article/doi/10.1210/jc.2016-3446/3038070/Allcause-mortality-in-patients-with-diabetes-under>
4. Booth N. What are the Read Codes? [Internet]. Vol. 11, Health libraries review. 1994. p. 177–82. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10139676>
5. INPS. About us - The Health Improvement Network (THIN) [Internet]. 2017 [cited 2017 Jun 6]. Available from: <http://www.inps.co.uk/vision/thin/about-us>
6. Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research.

Pharmacoepidemiol Drug Saf [Internet]. 2007 Apr [cited 2016 Jul 1];16(4):393–401. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17066486>

7. Ding T, Baio G, Hardiman PJ, Petersen I, Sammon C. Diagnosis and management of polycystic ovary syndrome in the UK: a retrospective cohort study. 2016;5–8. Available from: <http://dx.doi.org/>
8. Wall H, Smith C, Hubbard R. Body mass index and obstructive sleep apnoea in the UK: a cross-sectional study of the over-50s. *Prim Care Respir J* [Internet]. 2012 Dec 29 [cited 2015 Nov 22];21(4):371–6. Available from: <http://www.nature.com/articles/pcrj201253>
9. Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf* [Internet]. 2009 Jan [cited 2015 Nov 22];18(1):76–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19065600>
10. Lim SS, Davies MJ, Norman RJ, Moran LJ. Overweight, obesity and central obesity in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update* [Internet]. 2012;18(6):618–37. Available from: <http://humupd.oxfordjournals.org/cgi/doi/10.1093/humupd/dms030>
11. Toulis KA, Willis BH, Marshall T, Kumarendran B, Gokhale K, Ghosh S, Thomas GN, Cheng KK, Narendran P, Hanif W, Nirantharakumar K. All-Cause Mortality in Patients With Diabetes Under Treatment With Dapagliflozin: A Population-Based, Open-Cohort Study in The Health Improvement Network Database. *J Clin Endocrinol Metab* [Internet]. 2017 May 1 [cited 2018 Mar 21];102(5):1719–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28323967>
12. Davé S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* [Internet]. 2009 Aug [cited 2017 May 8];18(8):704–7. Available from: <http://doi.wiley.com/10.1002/pds.1770>
13. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. *World Health Organ Tech Rep Ser* [Internet]. 1995 [cited 2018 Feb 14];854:1–452. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8594834>
14. Moran LJ, Misso ML, Wild RA, Norman RJ. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update* [Internet]. 2010;16(4):347–63. Available from: <http://humupd.oxfordjournals.org/cgi/doi/10.1093/humupd/dmq001>
15. McEvoy RD. Obstructive sleep apnoea and hypertension: the ESADA study. *Eur Respir J* [Internet]. 2014 Oct 1 [cited 2017 Sep 24];44(4):835–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25271222>
16. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):1–22.
17. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* [Internet]. 1995 [cited 2017 May 24];310(6973). Available from: <http://www.bmj.com/content/310/6973/170.long>
18. Matok I, Azoulay L, Yin H, Suissa S. Immortal time bias in observational studies of drug effects in pregnancy. *Birth Defects Res A Clin Mol Teratol* [Internet]. 2014 Sep [cited 2015 Nov 11];100(9):658–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24985740>
19. StataCorp. *Stata Statistical Software: Release 14* [Internet]. 2015. TX: StataCorp LP.: College Station; 2015 [cited 2017 May 26]. Available from: <http://www.stata.com/statamp/>

Options for methodical selection of Read codes

- System based grouping
- Use of flags in prioritizing codes into difference hierarchy based on relevance
- Read codes cited in previous studies
- Read codes listed in Clinical codes.org
- Read codes corresponding to Quality Outcome Framework (QoF)
- Applying search for key words
- Discussion within the team (First attempt)
- Discussion with GPs

Panel E1. Options for methodical selection of Read codes

Table E1: Read codes used to identify polycystic ovary syndrome

<u>Read Code</u>	<u>Description</u>
Polycystic ovary syndrome	
C164.11	Iso-sexual virilisation
C164.12	Stein - Leventhal syndrome
C165.00	Polycystic ovarian syndrome
7E25300	Endoscopic drilling of ovary
Polycystic ovaries	
C164.00	Polycystic ovaries
C164.13	Multi-cystic ovaries
Sleep apnoea	
Fy03.00	Sleep apnoea
Fy03.11	Obstructive sleep apnoea
Fy04.11	Ondine's curse
H5B..00	Sleep apnoea
H5B0.00	Obstructive sleep apnoea
R005100	[D]Insomnia with sleep apnoea
R005300	[D]Hypersomnia with sleep apnoea
R005311	[D]Sleep apnoea syndrome
R005312	[D]Syndrome sleep apnoea
R060400	[D]Apnoea

E2 Table: Regression model estimates for hazard of women with PCOS to develop OSA compared to women without PCOS (n=220,055)

Covariate	Hazard ratio	95% Confidence interval	P-value
PCOS	2.26	(1.89 2.69)	<0.001
Age	1.05	(1.04 1.06)	<0.001
Townsend score [#]			
1	1.00	(Reference category)	
2	1.22	(0.89 1.66)	0.220
3	1.29	(0.95 1.73)	0.100
4	1.51	(1.12 2.03)	0.007
5	1.76	(1.30 2.40)	<0.001
Missing or implausible data	1.43	(1.01 2.01)	0.044
BMI Category (kg/m ²)			
<25	1.00	(Reference level)	
25-29.99	2.94	(1.88 4.60)	<0.001
>30	13.69	(9.37 20.0)	<0.001
Missing or implausible data	2.34	(1.41 3.87)	0.001
Diabetes or Impaired glucose regulation*	1.26	(0.87 1.81)	0.223
Hypothyroidism	2.02	(1.50 2.71)	<0.001

Townsend score – presented as quintiles with 1 least deprived and 5 most deprived

* (includes Impaired Fasting Glucose and Impaired Glucose Tolerance)

E3 Table: Subgroup analysis for hazard of women with PCOS to develop OSA compared to women without PCOS stratified by BMI category

	BMI <25kg/m2		25-29.99 Kg/m2		≥30Kg/m2	
	PCOS (Exposed)	Controls (Unexposed)	PCOS (Exposed)	Controls (Unexposed)	PCOS (Exposed)	Controls (Unexposed)
Total number of participants	25,725	52,651	15,035	29,656	23,912	37,409
Person years	117,924	243,863	72,979	144,545	117,242	186,177
Incident OSA n (%)	14 (0.05)	15 (0.03)	30 (0.20)	26 (0.09)	227 (0.94)	176 (0.47)
Incidence rates per 10,000 person years	1.19	0.62	4.1	1.80	19.4	9.45
Hazard Ratio (95% CI)	1.92 (0.92 - 3.97)		2.25 (1.33 - 3.81)		2.05 (1.68 - 2.49)	
p-value	0.080		0.002		<0.001	
Adjusted Hazard Ratio (95% CI)*	1.91 (0.92 - 3.97)		2.25 (1.33 - 3.81)		2.10 (1.72 - 2.56)	
p-value	0.081		0.003		<0.001	

* Adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation and hypothyroidism at baseline

Chapter 5 : PCOS and dysglycaemia

Polycystic Ovary Syndrome, Combined Oral Contraceptives, and the Risk of Dysglycemia: A Population-Based Cohort Study With a Nested Pharmaco-epidemiological Case-Control Study

Author Contributions.

Balachandran Kumarendran (BK), Michael O'Reilly (M.W.O.), Wibke Arlt (W.A), and Krish Nirantharakumar (K.N) developed the research question and designed the study. Dana Sumilo (D.S) and Laurent Azoulay (L.A) contributed to the design of the study. B.K., M.W.O., Anuradha Subramanian (A.S), W.A., and K.N. designed the analysis, interpreted the results, and drafted the manuscript. All authors reviewed and revised the manuscript.



Polycystic Ovary Syndrome, Combined Oral Contraceptives, and the Risk of Dysglycemia: A Population-Based Cohort Study With a Nested Pharmacoepidemiological Case-Control Study

Diabetes Care 2021;44:2758–2766 | <https://doi.org/10.2337/dc21-0437>

Balachandran Kumarendran,^{1,2}
Michael W. O'Reilly,^{3,4}
Anuradha Subramanian,¹ Dana Šumilo,¹
Konstantinos Toulis,¹
Krishna M. Gokhale,¹
Chandrika N. Wijeratne,⁵
Arri Coomarasamy,⁴ Abd A. Tahrani,⁴
Laurent Azoulay,⁶ Wiebke Arlt,^{4,7} and
Krishnarajah Nirantharakumar^{1,4}

OBJECTIVE

Irregular menstrual cycles are associated with increased cardiovascular mortality. Polycystic ovary syndrome (PCOS) is characterized by androgen excess and irregular menses; androgens are drivers of increased metabolic risk in women with PCOS. Combined oral contraceptive pills (COCs) are used in PCOS both for cycle regulation and to reduce the biologically active androgen fraction. We examined COC use and risk of dysglycemia (prediabetes and type 2 diabetes) in women with PCOS.

RESEARCH DESIGN AND METHODS

Using a large U.K. primary care database (The Health Improvement Network [THIN]; 3.7 million patients from 787 practices), we carried out a retrospective population-based cohort study to determine dysglycemia risk (64,051 women with PCOS and 123,545 matched control subjects), as well as a nested pharmacoepidemiological case-control study to investigate COC use in relation to dysglycemia risk (2,407 women with PCOS with [case subjects] and without [control subjects] a diagnosis of dysglycemia during follow-up). Cox models were used to estimate the unadjusted and adjusted hazard ratio, and conditional logistic regression was used to obtain adjusted odds ratios (aORs).

RESULTS

The adjusted hazard ratio for dysglycemia in women with PCOS was 1.87 (95% CI 1.78–1.97, $P < 0.001$; adjustment for age, social deprivation, BMI, ethnicity, and smoking), with increased rates of dysglycemia in all BMI subgroups. Women with PCOS and COC use had a reduced dysglycemia risk (aOR 0.72, 95% CI 0.59–0.87).

CONCLUSIONS

In this study, limited by its retrospective nature and the use of routinely collected electronic general practice record data, which does not allow for exclusion of the impact of prescription-by-indication bias, women with PCOS exposed to COCs had a reduced risk of dysglycemia across all BMI subgroups. Future prospective studies should be considered for further understanding of these observations and potential causality.

¹Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, U.K.

²Department of Community and Family Medicine, Faculty of Medicine, University of Jaffna, Kokkuvil, Sri Lanka

³Department of Medicine, Royal College of Surgeons in Ireland, University of Medicine and Health Sciences, Dublin, Republic of Ireland

⁴Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, U.K.

⁵Department of Obstetrics and Gynecology, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka

⁶Department of Epidemiology, Biostatistics and Occupational Health and Gerald Bronfman Department of Oncology, McGill University, Toronto, Canada

⁷NIHR Birmingham Biomedical Research Centre, University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, Birmingham, U.K.

Corresponding author: Wiebke Arlt, w.ahrt@bham.ac.uk, or Krishnarajah Nirantharakumar, k.nirantharan@bham.ac.uk

Received 22 February 2021 and accepted 8 September 2021

This article contains supplementary material online at <https://doi.org/10.2337/figshare.16608068>.

B.K., M.W.O., and A.S. contributed equally to this work.

W.A. and K.N. are joint senior authors.

© 2021 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/content/license>.

Polycystic ovary syndrome (PCOS) is the most prevalent endocrine disorder in women of reproductive age (1) and is defined by irregular menses and androgen excess. While previously mostly perceived as a reproductive disorder, PCOS is now recognized as a lifelong metabolic disorder with an increased prevalence of the cardiovascular risk factors insulin resistance, dyslipidemia, and hypertension (2–4). An increased risk of type 2 diabetes in women with PCOS has been described in both cross-sectional (5) and cohort studies (5,6), the latter reporting a two- to fourfold increased risk, with type 2 diabetes diagnosed on average 4 years earlier in women with PCOS than in the background population (7). In a population-based cohort study, women with PCOS were also found to have an increased risk of nonalcoholic fatty liver disease (NAFLD), a significant hepatic complication of metabolic syndrome (3). In a recently published prospective cohort study (8) of nearly 80,000 women with a follow-up period of 24 years, investigators described an increased risk of premature mortality, primarily due to cardiovascular disease, in women with irregular and long menstrual cycles, suggestive of PCOS as the major underlying risk factor.

Androgen excess is a cardinal feature of PCOS (1), and its severity has been shown to correlate with insulin resistance in cross-sectional studies (9–11). In the general population, type 2 diabetes risk in women increases with circulating androgen concentrations and decreasing concentrations of sex hormone binding globulin (SHBG) (12–15). Endogenous androgen concentrations in women have also been identified as a risk factor for the development of NAFLD in PCOS, independent of body weight (3). Combined oral contraceptive pills (COCPs) are widely prescribed in women with PCOS for menstrual cycle regulation. In addition, COCPs can exert antiandrogen effects through two distinct mechanisms. The estrogen component in COCPs increases the production of SHBG in the liver, thereby reducing the concentration of free testosterone capable of binding and activating the androgen receptor in target tissues of androgen action (16). Furthermore, some progestins used in COCPs can convey additional antiandrogenic action through androgen receptor blockade, namely, cyproterone and drospirenone, while

other progestins are proandrogenic or exert no effect on the androgen receptor.

Data on the impact of COCP prescription on glucose metabolism in women are conflicting. Clinicians have previously raised concerns that COCP intake may adversely impact on glucose metabolism (17); however, available evidence is limited by a lack of prospective studies and by the confounding issue of higher ethinylestradiol content in historical formulations. Furthermore, a wide diversity of combined oral contraceptive formulations available, including differences in progestin components, makes an accurate assessment of the impact of COCP prescription on glycemia very challenging. In a recent Cochrane library review the authors concluded that current evidence suggests no significant impact on carbohydrate metabolism in women without PCOS, highlighting a paucity of large-scale prospective studies to adequately address the question (18). Conversely, it has been hypothesized that the impact of COCP on carbohydrate metabolism may be protective against incident dysglycemia, due to the impact both of raising SHBG levels and of partial androgen receptor blockade in selected formulations containing antiandrogenic progestins.

Here we tested the hypothesis that the use of COCPs decreases the risk of type 2 diabetes in women with PCOS. To this end, undertaking a population-based cohort study, we first determined the risk of incident dysglycemia, i.e., a composite outcome combining prediabetes and type 2 diabetes, in women with PCOS and then examined in a nested pharmacoepidemiological case-control study whether COCP intake impacts on this risk.

RESEARCH DESIGN AND METHODS

Data Source

Data sets were derived from a U.K. primary care database. There are >17 million patient records from 787 general practices in the IQVIA Medical Research Data (also known as The Health Improvement Network [THIN] database) (19). THIN uses Read codes, a hierarchical coding system for recording symptoms and diagnoses, and is highly suited for assessment of chronic health conditions (3,12,20).

Study Population

Our study population was comprised of women aged 18–50 years during the study period (1 January 2000–31 January 2017). Women were eligible 1 year after registration with their general practice or from the time their practice became eligible for THIN participation (3).

Study Designs

PCOS and Incident Dysglycemia

This matched cohort study considered women with PCOS as exposed. Exposure was ascertained by Read codes for “Polycystic ovary syndrome (PCOS)” or “Polycystic ovaries (PCO)” (3), as this composite code list reflects community prevalence (21). Each exposed woman was matched with up to two women without PCOS within the same general practice for age (± 2 years) and BMI (± 2 kg/m²) (22,23).

Follow-up start date or the index date for the exposed patients was set to PCOS diagnosis date for incident PCOS patients or patient eligibility date for prevalent PCOS patients (patients with a diagnosis ahead of cohort entry). The index date for a matched control subject was set to the corresponding index date of the exposed patient to mitigate immortality time bias (24). Follow-up occurred until the earliest occurrence of 1) outcome, 2) study end, or 3) patient censorship denoted by death, deregistration from the practice, or practice withdrawing from the THIN database.

Outcomes were type 2 diabetes and dysglycemia, with the latter defined as the composite outcome of prediabetes and type 2 diabetes, which were ascertained by Read codes and laboratory results (type 2 diabetes, HbA_{1c} $\geq 6.5\%$ [48 mmol/mol], fasting blood glucose ≥ 7 mmol/L; dysglycemia, HbA_{1c} $\geq 6.0\%$ (42 mmol/mol), fasting blood glucose ≥ 6 mmol/L, random blood glucose ≥ 11.1 mmol/L, and 2-h oral glucose tolerance test [OGTT] result indicated as “abnormal” or “high”). Patients with a recording of the outcome of interest (dysglycemia) or glucose-lowering drug prescription at baseline were not eligible.

Risk Factors of Type 2 Diabetes and Dysglycemia Among Women With PCOS

To identify potential risk factors for the development of type 2 diabetes and dysglycemia within the PCOS cohort, we examined demographic risk factors, BMI, clinical features of androgen excess,

and prescription of COCPs at baseline as candidate risk factors.

Combined Oral Contraceptives and Incident Dysglycemia Among Women With PCOS

To define the impact of COCPs on dysglycemia risk, we conducted a nested case-control study. Women who developed dysglycemia during the follow-up period were case subjects, and the remaining women were potential control subjects. One control subject per case subject was randomly selected after matching for age (± 2 years), BMI (± 2 kg/m²), PCOS diagnosis date (± 2 years), and whether PCOS was diagnosed before or after the patient became eligible to take part in the study. Index date was assigned as the date of diagnosis of dysglycemia for the case subjects, and the same date was assigned to the corresponding control, ensuring a comparable exposure window for matched case-control pairs and, therefore, avoiding time-window bias (25).

The exposure window was prespecified and extended from 1 year prior to cohort entry, to avoid disregarding valid prescriptions in the immediate period after patient registration, and 6 months prior to index date, to exclude prescriptions that cannot be validly attributed to the development of dysglycemia.

COCP prescription was initially considered as a binary variable. COCP prescription was then categorized according to whether the respective progestin component exerts antiandrogen activity. Patients with no prescription of COCP formed the reference groups for both the categorical exposure variables.

Analysis

Crude incidence rates of the primary and secondary composite outcome (type 2 diabetes and dysglycemia) were estimated per 10,000 person-years. Unadjusted and adjusted hazard ratios were obtained using Cox models. Covariates for adjustment were selected based on biological plausibility for confounding. Covariates include age, BMI, socioeconomic status, ethnicity, smoking status, and record of hypertension, hypothyroidism, and prescription of lipid-lowering medications. Socioeconomic status was presented with use of Townsend score (26–28). Ethnicity was categorized based on U.K. 2011 census classification. Smoking status was categorized as currently smoking, discontinued, and never smoked.

Selection of Read code lists exposure, outcome, and covariates was based on methods and codes set out in previous publications (3,12,20,29) (Supplementary Table 1). BMI was categorized as per World Health Organization guidelines, with nonstandard BMI categorization of South Asian women as per the recommended guidelines (30).

Sensitivity Analyses

Sensitivity analyses were performed to assess the extent of misclassification and survival bias. The exposure was restricted, firstly, to women with PCOS-specific diagnostic codes and, secondly, to those with a PCOS diagnosis during the study period (incident patients) (31).

In addition, in the nested-case control study, we investigated whether control selection based on risk set sampling altered our findings, allowing a patient to serve as a control subject for multiple patients diagnosed with dysglycemia, while patients not diagnosed with dysglycemia at a similar time of follow-up could serve as control subjects before they developed dysglycemia.

Subgroup Analyses

To check whether risks of type 2 diabetes and dysglycemia are independent of BMI status, we conducted subgroup analyses within each BMI category.

Analyses for Predictors of Dysglycemia

In the cohort restricted to women with PCOS, Cox regression analysis was used to identify statistically significant predictors of type 2 diabetes and dysglycemia. In addition to covariates mentioned in the primary analysis, prescription of COCPs and variables characteristic of androgen excess and prescription of antiandrogen therapy with single agent drugs were also considered as candidate predictors.

Analysis of Nested Case-Control Study

Conditional logistic regression was performed to obtain unadjusted and adjusted ORs for dysglycemia based on exposure to COCP. The adjusted model included all covariates in the primary analysis, plus prescription of metformin and antiandrogen therapy.

RESULTS

Study Population Characteristics

A total of 64,051 women with PCOS and 123,545 women without PCOS and

matched for age, sex, and general practice were included in the study (Supplementary Fig. 1 and Table 1). The median follow-up period was 3.5 years (interquartile range [IQR] 1.4–7.2). Mean (SD) age of the whole cohort was 30.5 (7.1) years and median BMI 25.6 kg/m² (IQR 22.1–31.4). Age, BMI, deprivation quintiles (Townsend index), and smoking status had no apparent imbalance in distribution between the two groups. Women with PCOS were more likely to be documented as South Asian (4.8% vs. 2.9%), hypothyroid (3.4% vs. 2.1%), and hypertensive (2.2% vs. 1.6%) at baseline (Table 1). COCPs were prescribed for 43.4% of the PCOS-exposed women before the index date; 22.5% of the women with PCOS were prescribed COCPs with an antiandrogenic progestin component (drospirenone or cyproterone acetate) (Table 1).

Risk of Type 2 Diabetes and Dysglycemia

In the primary analysis, the incidence rate of type 2 diabetes among the exposed and the unexposed was 48.7 and 22.8 per 10,000 person-years during a median follow-up of 3.39 years (IQR 1.34–7.16) and 3.47 years (IQR 1.39–7.18), respectively, equating to a doubling in risk of type 2 diabetes among women with PCOS (hazard ratio 2.13, 95% CI 1.98–2.29, $P < 0.001$). Adjustment for age, deprivation quintiles, BMI category, ethnicity, smoking status, and hypothyroidism did not alter the estimated hazard ratio (adjusted hazard ratio [aHR] 2.04, 95% CI 1.89–2.20, $P < 0.001$) (Supplementary Table 2).

In analysis of the effect of PCOS on the composite outcome (dysglycemia), a similar effect was observed (aHR 1.87, 95% CI 1.78–1.97, $P < 0.001$). The incidence rates of dysglycemia were 96.3 and 49.4 per 10,000 person-years among women with and without PCOS during a median follow-up of 3.32 years (IQR 1.32–7.03) and 3.44 years (IQR 1.38–7.11), respectively (Supplementary Table 2).

Sensitivity Analysis

The strength of association between PCOS and type 2 diabetes did not decrease when the analysis was restricted to women with incident diagnosis of PCOS (aHR 1.98, 95% CI 1.70–2.31, $P < 0.001$) and to women with PCOS-specific codes (aHR 2.17, 95% CI 1.88–2.51, $P < 0.001$). This was similarly observed for dysglycemia

Table 1—Baseline characteristics of participants of the population-based cohort study, with stratification by PCOS exposure status

	Women with PCOS (<i>n</i> = 64,051)	Women without PCOS (<i>n</i> = 123,545)
Age, years, mean (SD)	30.4 (7.0)	30.5 (7.1)
BMI, kg/m ² , median (IQR)	25.9 (22.2–31.9)	25.4 (22.0–30.8)
BMI categories, <i>n</i> (%) [*]		
Normal/underweight	23,490 (36.6)	48,360 (39.1)
Overweight	12,734 (19.8)	25,229 (20.4)
Obese	17,591 (27.5)	29,907 (24.2)
Missing	10,236 (16.0)	20,049 (16.2)
Smoking status, <i>n</i> (%)		
Nonsmoker	37,311 (58.3)	71,114 (57.6)
Discontinued	9,044 (14.1)	16,285 (13.2)
Smoker	14,674 (22.9)	28,284 (22.9)
Missing	3,022 (4.7)	7,862 (6.4)
Ethnicity, <i>n</i> (%)		
Caucasian	30,597 (47.8)	50,206 (40.6)
Black	1,464 (2.3)	2,636 (2.1)
Chinese	582 (0.9)	883 (0.7)
South Asian	3,085 (4.8)	3,517 (2.9)
Mixed race	897 (1.4)	1,645 (1.3)
Missing	27,426 (42.8)	64,658 (52.3)
Townsend deprivation score, <i>n</i> (%)		
1 (least deprived)	11,270 (17.6)	21,839 (17.7)
2	10,280 (16.1)	19,866 (16.1)
3	12,064 (18.8)	23,471 (19.0)
4	11,530 (18.0)	22,623 (18.3)
5 (most deprived)	8,182 (12.8)	16,186 (13.1)
Missing	10,725 (16.7)	19,560 (15.8)
Baseline comorbidity, <i>n</i> (%)		
Hypothyroidism	2,172 (3.4)	2,585 (2.1)
Hypertension	1,420 (2.2)	2,030 (1.6)
Baseline medication, <i>n</i> (%)		
Any COCP	27,768 (43.4)	66,332 (53.7)
COCP without antiandrogenic progestin	25,481 (39.8)	64,157 (51.9)
COCP with antiandrogenic progestin	14,437 (22.5)	12,336 (10.0)
Drospirenone	4,944 (7.7)	6,550 (5.3)
Cyproterone	11,069 (17.3)	7,305 (5.9)
Single-agent antiandrogen therapy [†]		
Cyproterone	444 (0.69)	
Other antiandrogen drugs [‡]	42 (0.07)	
Lipid-lowering medication	410 (0.64)	534 (0.43)

^{*}Normal/underweight, <23.5 kg/m² for patients of South Asian ethnicity and <25 kg/m² for patients of all other ethnic groups; overweight, 23.5–27.5 kg/m² for patients of South Asian ethnicity and 25–30 kg/m² for patients of all other ethnic groups; and obese, ≥27.5 kg/m² for patients of South Asian ethnicity and ≥30 kg/m² for patients of all other ethnic groups. [†]Includes dutasteride, enzalutamide, finasteride, flutamide, and spironolactone. [‡]PCOS-relevant variables summarized only for the PCOS-exposed cohort; note that patients with impaired glucose regulation or glucose-lowering drug prescription at baseline were not included in the cohort.

(incident cohort, aHR 1.95, 95% CI 1.76–2.16, *P* < 0.001; PCOS-specific cohort, aHR 1.93, 95% CI 1.75–2.13, *P* < 0.001) (Supplementary Table 2).

Subgroup Analysis With Stratification by BMI

In subgroup analyses, women with PCOS had an increased risk of type 2 diabetes in all BMI categories compared with women without PCOS in the same BMI category (normal/underweight category,

BMI <23 kg/m² among women of South Asian ethnicity and <25 kg/m² among women of all other ethnic groups, aHR 1.88, 95% CI 1.42–2.51, *P* < 0.001; overweight category, BMI 23–27.5 kg/m² among women of South Asian ethnicity and 25–29.9 kg/m² among women of all other ethnic groups, aHR 1.92, 95% CI 1.56–2.35, *P* < 0.001; and obesity category, BMI ≥27.5 kg/m² among women of South Asian ethnicity and ≥30 kg/m² among women of all other ethnic groups,

aHR 1.88, 95% CI 1.72–2.06, *P* < 0.001) (Fig. 1A). Similar findings were observed for the composite outcome (dysglycemia) (Fig. 1A).

Risk Factors for Type 2 Diabetes and Dysglycemia Among Women With PCOS

In analysis of the cohort of women with PCOS to identify risk factors for type 2 diabetes, PCOS-specific variables emerged as significant risk factors, namely, anovulation

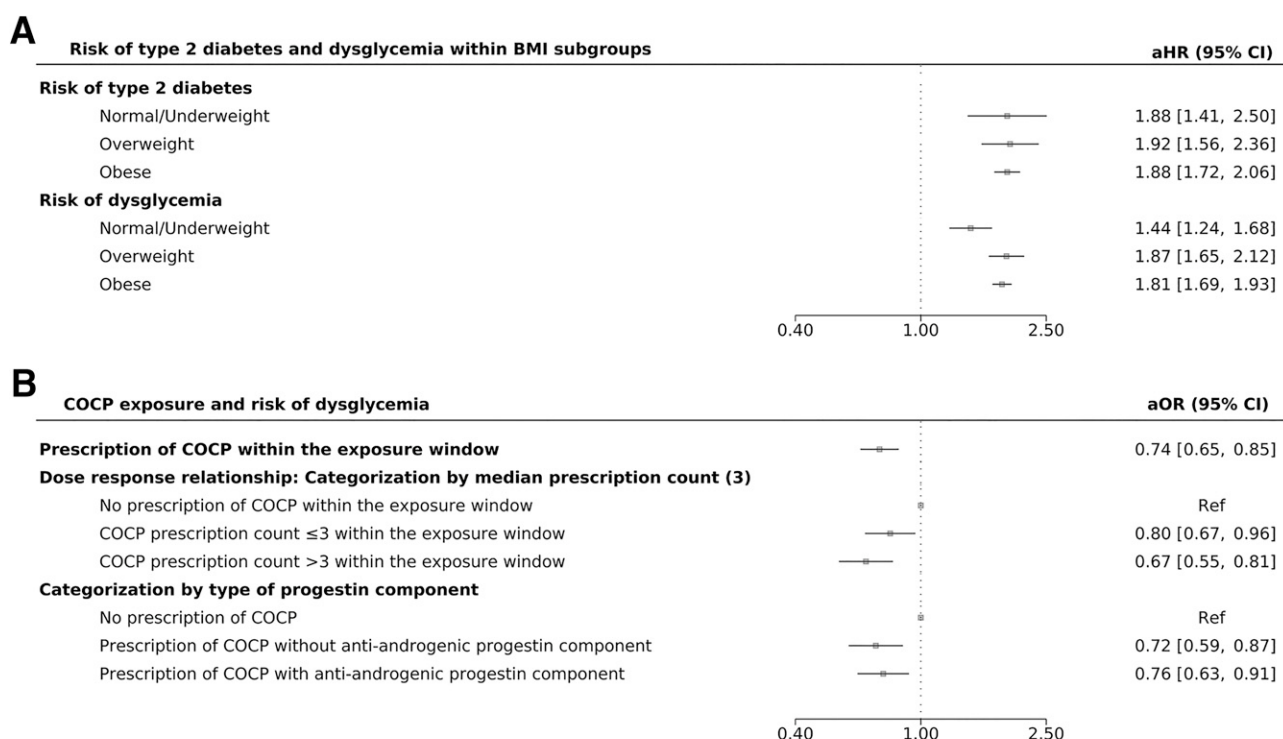


Figure 1—Risk of type 2 diabetes and dysglycemia among 64,051 women with PCOS compared with 123,545 matched control subjects and according to BMI subgroup (population-based cohort study [A]). aOR for risk of dysglycemia according to the prescription of COCPs (B) overall and according to prescription counts and type of progestin component, respectively, in the nested pharmacoepidemiological case-control study (2,407 women with PCOS with a diagnosis of dysglycemia during follow-up [case subjects] and 2,407 women with PCOS without a diagnosis of dysglycemia [control subjects]). Normal/underweight, <23.5 kg/m² for patients of South Asian ethnicity and <25 kg/m² for patients of all other ethnic groups; overweight, 23.5–27.5 kg/m² for patients of South Asian ethnicity and 25–30 kg/m² for patients of all other ethnic groups; and obese, ≥27.5 kg/m² for patients of South Asian ethnicity and ≥30 kg/m² for patients of all other ethnic groups. Ref, reference.

(aHR 1.21, 95% CI 1.08–1.35, $P = 0.001$) and hirsutism (aHR 1.20, 95% CI 1.05–1.36, $P = 0.007$). Conversely, prescription of COCPs emerged as a protective factor, with similar effects observed for COCPs with (aHR 0.84, 95% CI 0.73–0.97, $P = 0.020$) and without (aHR 0.83, 95% CI 0.72–0.94, $P = 0.005$) an antiandrogenic progestin component. The same risk factors and protective factors were observed for the composite dysglycemia outcome (Supplementary Table 3).

Nested Case-Control Analysis: The Effect of Oral Contraceptives on Risk of Dysglycemia

Of the 64,051 women with PCOS in the base cohort, 0.45% ($n = 2,885$) developed dysglycemia during follow-up, who were assigned as the case subjects in the nested case-control study (Table 2). The remaining 61,166 (95.5%) women were considered as potential control subjects. A total of 478 case subjects could not be matched to a control subject on the basis of age, BMI, PCOS

diagnosis date, and incident/prevalent status of PCOS diagnosis. Therefore, our final analysis included 2,407 case subjects and corresponding 2,407 matched control subjects.

Mean (SD) age at index date was 38.9 (8.3) years and mean age at PCOS diagnosis was 28.8 (14.4) years, which were similar between case subjects and control subjects. BMI at cohort entry was similarly distributed between case and control subjects (mean [SD] 32.7 [7.0] vs. 32.6 [7.0] kg/m²). Compared with control subjects, case subjects were more likely to be from a deprived background (Townsend deprivation score level 5 [deprived]: 17.0% vs. 12.3%), smokers (26.6% vs. 20.8%), and of South Asian ethnicity (10.0% vs. 3.2%). At cohort entry, there was also a higher proportion of case subjects with concurrent hypothyroidism (10.6% vs. 7.8% of control subjects). Altogether, 679 case subjects (28.2%) and 815 control subjects (33.9%) were prescribed COCPs during the

exposure window. Among those prescribed COCPs, the median COCP prescription count per person during the exposure window was 3 (IQR 1–7).

With adjustment for age, smoking status, BMI category, ethnicity, Townsend score, baseline hypothyroidism, hypertension, and prescription of isolated antiandrogen drugs, metformin, and lipid-lowering medication at baseline, women with PCOS exposed to COCP were seen to have a reduced risk of dysglycemia (adjusted odds ratio [aOR] 0.74, 95% CI 0.65–0.85, $P < 0.001$). For every issued COCP prescription recorded within the exposure window, there was a 2% reduction in the odds of dysglycemia (aOR 0.98, 95% CI 0.96–0.99, $P = 0.004$) (Fig. 1B).

When COCP prescription issue count was categorized as 1) no prescription, 2) prescription count of three or fewer, and 3) prescription count of more than three within the exposure window, a dose-responsive reduction in the risk of

Table 2—Baseline characteristics of women with PCOS included in the nested case-control study

Variable	Women with PCOS and a diagnosis of dysglycemia (case subjects), <i>n</i> = 2,407	Women with PCOS and without a diagnosis of dysglycemia (control subjects), <i>n</i> = 2,407
Age at index date (dysglycemia diagnosis for cases), years, mean (SD)	38.89 (8.32)	38.84 (8.27)
Age at PCOS diagnosis, years, mean (SD)	28.84 (14.43)	28.76 (14.00)
BMI, kg/m ² , mean (SD)	32.72 (6.98)	32.59 (7.03)
BMI categories, <i>n</i> (%)		
Normal/underweight	270 (11.2)	305 (12.7)
Overweight	439 (18.2)	437 (18.2)
Obese	1,322 (54.9)	1,289 (53.5)
Missing	376 (15.6)	376 (15.6)
Townsend deprivation score, <i>n</i> (%)		
1 (least deprived)	351 (14.6)	481 (20.0)
2	359 (14.9)	436 (18.1)
3	473 (19.7)	457 (19.0)
4	471 (19.6)	420 (17.5)
5 (most deprived)	408 (17.0)	295 (12.3)
Missing	345 (14.3)	318 (13.2)
Smoking status, <i>n</i> (%)		
Nonsmoker	1,306 (54.3)	1,354 (56.3)
Discontinued	295 (12.3)	362 (15.0)
Smoker	639 (26.6)	501 (20.8)
Missing	167 (6.9)	190 (7.9)
Ethnicity, <i>n</i> (%)		
Caucasian	999 (41.5)	1,099 (45.7)
Mixed race	38 (1.6)	21 (0.87)
Chinese/Middle Eastern/other	21 (0.87)	13 (0.54)
Black	80 (3.3)	40 (1.7)
South Asian	241 (10.0)	77 (3.2)
Missing	1,028 (42.7)	1,157 (48.1)
Concurrent conditions at baseline, <i>n</i> (%)		
Hypothyroidism	256 (10.6)	188 (7.8)
Hypertension	623 (25.88)	179 (11.59)
Prescription of drugs within the exposure time window, <i>n</i> (%)		
Contraceptives		
No pill	1,728 (71.8)	1,592 (66.1)
COCP without antiandrogenic progestin	301 (12.5)	389 (16.2)
COCP with antiandrogenic progestin*	378 (15.7)	426 (17.7)
Single-agent antiandrogen therapy^	41 (1.7)	23 (0.96)
Metformin	417 (17.3)	330 (13.7)
Lipid-lowering medication	150 (6.23)	119 (4.94)

Case and control subjects are matched women with and without a diagnosis of dysglycemia during follow-up, respectively. *Normal/underweight, <23.5 kg/m² for patients of South Asian ethnicity and <25 kg/m² for patients of all other ethnic groups; overweight, 23.5–27.5 kg/m² for patients of South Asian ethnicity and 25–30 kg/m² for patients of all other ethnic groups; obese, ≥27.5 kg/m² for patients of South Asian ethnicity and ≥30 kg/m² for patients of all other ethnic groups. *Cyproterone acetate/drospirenone. ^Cyproterone acetate/flutamide/finasteride.

dysglycemia was observed (in reference to no prescription of COCP, aOR of dysglycemia with prescription count of three or fewer = 0.80, 95% CI 0.67–0.96, *P* = 0.017, and aOR with prescription count of more than three = 0.67, 95% CI 0.55–0.81, *P* < 0.001) (Fig. 1B).

Women with PCOS exposed to COCPs had a reduced risk of dysglycemia irrespective of the type of progestin component (COCPs with antiandrogenic progestin, aOR 0.76, 95% CI 0.63–0.91, *P* = 0.003; COCPs with progestin without

antiandrogen activity, aOR 0.72, 95% CI 0.59–0.87; *P* < 0.001) (Fig. 1B and Supplementary Table 4).

Metformin prescription within the exposure window period was associated with increased risk of dysglycemia (aOR 1.50, 95% CI 1.24–1.81, *P* < 0.001), suggestive of possible prescription-by-indication bias for those at increased risk. Findings in the sensitivity analysis with incorporation of a risk set sampling approach showed a similar result (aOR 0.76, 95% CI 0.63–0.91, *P* = 0.003).

CONCLUSIONS

With use of a rigorous nested case-control pharmacoepidemiological analysis, we found that women with PCOS exposed to COCPs had a reduced risk of developing dysglycemia across all BMI subgroups. Our study is also the largest with reporting of glycemic outcomes in a primary care cohort of women with PCOS, demonstrating a twofold increased risk of incident type 2 diabetes and dysglycemia in women with PCOS of any BMI.

Our finding of an increased type 2 diabetes risk in women with PCOS is consistent with results of recent population studies from Denmark and Finland (7,32) and hospitalization data from Australia (33), all reporting a two- to fourfold increased type 2 diabetes risk in PCOS. Using the Australian Longitudinal Study on Women's Health, Kakoly et al. (34) demonstrated that a diagnosis of PCOS was one of the most influential predictors of incident type 2 diabetes in women, even after adjusting for BMI and family history. Few population studies have looked specifically at the composite outcome of dysglycemia, which takes into account a spectrum of impaired glucose regulation ranging from impaired glucose tolerance and impaired fasting glucose through to overt hyperglycemia (35,36). Crucially, our data highlight that normal-weight women with PCOS were also at increased risk of type 2 diabetes and dysglycemia. This parallels our previous finding of increased NAFLD risk in normal-weight women with PCOS (3), further challenging the notion that PCOS-related metabolic complications are only relevant in the context of obesity.

These data suggest that, rather than obesity in isolation, PCOS-specific factors, including androgen excess, underpin the increased metabolic risk. We found that those women with PCOS and hirsutism, a clinical feature of androgen excess, had a further increased risk of dysglycemia. In a population-based cohort study, using the same primary care population database, we previously documented an independent link between serum testosterone and incident diabetes risk in women (12). We demonstrated that the risk of incident type 2 diabetes increased significantly in women with serum testosterone level >1.5 nmol/L compared with the reference cohort with levels <1 nmol/L; the risk was twofold higher in women with serum testosterone values >3.5 nmol/L. We also demonstrated in a small cross-sectional cohort study that women with increased circulating androgen concentrations had a higher risk of an abnormal OGTT result, with the OGTT-derived insulin sensitivity index correlating inversely with circulating androgen burden (9). In a recent meta-analysis (37) it was demonstrated that women with increased serum testosterone had a 60% higher risk of type 2 diabetes than women with normal

testosterone levels. Furthermore, a recent large-scale genome association study in 425,097 participants of the UK Biobank demonstrated that the risk of type 2 diabetes in women increased in line with increasing circulating testosterone concentrations (15).

The association of female androgen excess, insulin resistance, and type 2 diabetes is undoubtedly complex. Insulin resistance promotes androgen excess by upregulating ovarian androgen generation and peripheral androgen activation in adipose tissue (38,39); the latter increases lipid accumulation in the adipocyte and, once adipocyte lipid storage capacity is exhausted, fatty acid overspill (39), which is intricately linked to metabolic dysfunction. Abnormalities in skeletal muscle metabolic function have also been described in PCOS, with altered muscle mitochondrial energy biogenesis in the context of androgen excess likely to drive disturbances in glucose metabolism (40,41). Rodent-based studies also support a direct role for androgens in pancreatic β -cell dysfunction, driving insulin hypersecretion, oxidative injury, and consequent β -cell failure (42). These data have recently been underpinned by a study using human pancreatic islets, demonstrating that intracrine activation of testosterone to the most potent androgen, 5 α -dihydrotestosterone, increases glucose-stimulated insulin secretion (43).

A recent cohort study (8) in nearly 80,000 women with a follow-up period of 24 years described an increased risk of premature mortality, primarily due to cardiovascular disease, in women with irregular cycles. COCPs are routinely used for menstrual cycle regulation in women with PCOS. Our study is the first population-based study investigating the hypothesis that COCPs might mitigate the risk of dysglycemia in women with PCOS, with antiandrogen activity conferred by an estrogen-mediated increase in SHBG as the proposed mechanism. Studies examining the impact of COCPs on glucose metabolism have reported conflicting results, and most are limited by small participant numbers and significant heterogeneity in COCP use. In a 2016 Korean population study of 6,554 postmenopausal women, investigators found that those who took the COCP during their reproductive years for >6 months had a 37% increased risk of type 2 diabetes (44). However, in a more recent study

with examination of the National Health and Nutrition Examination Survey (NHANES) database between 2007 and 2018 it was found that COCP use in $>6,000$ women aged 35–50 years who met matching criteria was associated with a 29% reduced risk of type 2 diabetes compared with the risk of never users (45). A further limitation is the tendency in previous studies to extrapolate data from otherwise healthy female patient groups to women with PCOS, who are likely to manifest a biologically distinct set of risk factors for dysglycemia. For the first cohort of the Nurses' Health Study, 2,276 healthy women were followed for a median of 12 years from 1976, with findings that risk of type 2 diabetes was increased by 10% in women with previous COCP use compared with those who never took the medication (46); however, these data reflect the use of older COCP preparations with higher ethinylestradiol concentrations between the 1970s and 1990s. In a recent Cochrane Library review no convincing evidence was found of glycemic risk associated with COCP prescription among women without PCOS (18), while a 2011 meta-analysis of the limited evidence in women with PCOS suggested neither adverse nor beneficial impact of COCPs on glucose homeostasis (47). A 2017 systematic review and meta-analysis highlighted the urgent need for further studies to understand the relationship between glucose metabolism and COCP use in both lean and obese women with PCOS (48). The results of our study improve our understanding in this regard and indicate the need for prospective, randomized controlled trials on the impact of COCPs on the risk of type 2 diabetes and dysglycemia. We found that following adjustment for confounding factors, women with PCOS and COCP use had a 27% reduction in the relative risk of incident dysglycemia, with the highest reduction in patients receiving higher numbers of COCP prescriptions. When analyzed separately, women with PCOS and COCP use had a similarly reduced risk of dysglycemia when exposed to COCPs with and without antiandrogenic progestin components, suggesting that the estrogen-induced increase in SHBG may be the primary driver of the risk-mitigating effect. However, this finding is potentially limited by the lower number of patients receiving antiandrogenic COCPs. Cyproterone acetate and drospirenone are progestins with antiandrogenic properties, as opposed to progestins such as desogestrel or levonorgestrel, which have neutral or

proandrogenic effects (49). While cyproterone acetate and drospirenone exert antiandrogen activity via androgen receptor blockade, their antiandrogen activity is considerably lesser than that of recently approved novel antiandrogens mainly used in the treatment of prostate cancer (50).

Our finding that women using metformin and women using single-agent antiandrogen therapy had an increased risk of incident dysglycemia is very likely reflective of a confounding-by-indication bias (51). Accordingly, the women with PCOS at highest risk of dysglycemia based on metabolic or androgen phenotype may have been systematically prescribed metformin and single agent antiandrogen therapy. It is possible that our observation of reduced dysglycemia risk in women with PCOS on COCPs may also reflect a prescription-by-indication bias, whereby those women with cardiovascular risk factors such as obesity, dyslipidemia, and hypertension were less likely to have been prescribed the COCP. However, we believe that this is less likely from closer review of the data; in our nested pharmacoepidemiological study, 26% of women had a BMI in the obese range and one-quarter of women with a BMI >35 kg/m² took COCPs during the follow-up period. We also carefully adjusted our analysis for metabolic phenotype by including BMI, hypertension, and dyslipidemia as variables.

Our study has a number of notable limitations, including the above-mentioned prescription-by-indication bias issues and others that are common to studies of retrospective data with use of electronic general practice databases. The definition of women with no PCOS was based on the absence of any Read code in relation to PCOS and not on systematic diagnostic assessment to exclude PCOS. Therefore, the proportion of women with PCOS was also much lower than the published community prevalence data for PCOS (52). Another limitation is that we used the Read code for polycystic ovaries (PCO) as indicative of PCOS. However, in a sensitivity analysis limited to women with PCOS Read codes, we documented similar findings, excluding the use of the PCO Read code as a significant limitation. Higher testing rates for type 2 diabetes among women with PCOS may also have resulted in overestimating the effect size; however, the effect size observed for type 2 diabetes in our study is similar to that of

existing literature (53). It was also not possible to adjust for more specific lifestyle factors such as physical activity, energy intake, or fiber consumption within a large population database as used in the current study. To explore the possibility of right censoring bias, the median follow-up and the loss to follow-up pattern was compared between patients with and without PCOS. There was no systematic difference observed between the two groups, and therefore the assumption of noninformative censoring was reasonable for the time-to-event analysis in this study, limiting the possibility of right censoring bias.

In conclusion, we demonstrated that women with PCOS have a significantly increased risk of dysglycemia that persisted after adjustment for BMI, corroborating the recommendation that women with PCOS should be systematically screened for type 2 diabetes irrespective of body weight category. In our nested pharmacoepidemiology study, we found that women with PCOS and exposure to COCPs had a lower risk of incident dysglycemia. Though the limitations of our study design preclude ascertainment of causality, we hypothesize that a beneficial effect of COCPs might be conveyed by an estrogen-induced increase in hepatic SHBG production. This increase would result in a decrease in the biologically active, unbound circulating androgen fraction, and this reduction in androgen excess could have metabolically beneficial effects including a decrease in risk of dysglycemia. However, to definitively establish causality a large-scale randomized trial evaluating the efficacy of COCPs in reducing the risk of dysglycemia in women with PCOS would be required, with careful comparison of the potential additional benefit of COCPs containing antiandrogenic progestin components.

Funding. This work was supported by the Wellcome Trust (Investigator Grant WT209492/Z/17/Z [to W.A.]) and the Health Research Board (Emerging Clinician Scientist Award ECSA-FA-2020-001 [to M.W.O.]). K.N. is a UK Research and Innovation (UKRI)/Health Data Research (HDR) UK Innovation Clinical Fellow. W.A. receives support from the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham (grant BRC-1215-20009).

The views expressed are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care of the U.K.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. B.K., M.W.O., W.A., and K.N. developed the research question and designed the study. D.Š. and L.A. contributed to the design of the study. B.K., M.W.O., A.S., W.A., and K.N. designed the analysis, interpreted the results, and drafted the manuscript. D.Š. and L.A. contributed to the design of the study. All authors reviewed and revised the manuscript. W.A. and K.N. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Teede HJ, Misso ML, Costello MF, et al.; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertil Steril* 2018;110:364–379
2. Rande HS, Tan BK, Weickert MO, et al. Cardiometabolic aspects of the polycystic ovary syndrome. *Endocr Rev* 2012;33:812–841
3. Kumarendran B, O'Reilly MW, Manolopoulos KN, et al. 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* 2018;15:e1002542
4. Legro RS, Arslanian SA, Ehrmann DA, et al.; Endocrine Society. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2013;98:4565–4592
5. Pelan R, Mellembakken JR, Sundström-Poromaa I, et al. The prevalence of type 2 diabetes is not increased in normal-weight women with PCOS. *Hum Reprod* 2017;32:2279–2286
6. Morgan CL, Jenkins-Jones S, Currie CJ, Rees DA. Evaluation of adverse outcome in young women with polycystic ovary syndrome versus matched, reference controls: a retrospective, observational study. *J Clin Endocrinol Metab* 2012;97:3251–3260
7. Rubin KH, Glintborg D, Nybo M, Abrahamsen B, Andersen M. Development and risk factors of type 2 diabetes in a nationwide population of women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2017;102:3848–3857
8. Wang YX, Arvizu M, Rich-Edwards JW, et al. Menstrual cycle regularity and length across the reproductive lifespan and risk of premature mortality: prospective cohort study. *BMJ* 2020;371:m3464
9. O'Reilly MW, Taylor AE, Crabtree NJ, et al. Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *J Clin Endocrinol Metab* 2014;99:1027–1036
10. Münzker J, Hofer D, Trummer C, et al. Testosterone to dihydrotestosterone ratio as a new biomarker for an adverse metabolic phenotype in the polycystic ovary syndrome. *J Clin Endocrinol Metab* 2015;100:653–660
11. Luotola K, Piltunen TT, Puurunen J, Morin-Papunen LC, Tapanainen JS. Testosterone is associated with insulin resistance index independently of adiposity in women with polycystic ovary syndrome. *Gynecol Endocrinol* 2018;34:40–44
12. O'Reilly MW, Glisic M, Kumarendran B, et al. Serum testosterone, sex hormone-binding globulin

- and sex-specific risk of incident type 2 diabetes in a retrospective primary care cohort. *Clin Endocrinol (Oxf)* 2019;90:145–154
13. Ding EL, Song Y, Malik VS, Liu S. Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA* 2006;295:1288–1299
 14. Rasmussen JJ, Selmer C, Frøssing S, et al. Endogenous testosterone levels are associated with risk of type 2 diabetes in women without established comorbidity. *J Endocr Soc* 2020;4:a050
 15. Ruth KS, Day FR, Tyrrell J, et al.; Endometrial Cancer Association Consortium. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med* 2020;26:252–258
 16. Amiri M, Nahidi F, Bidhendi-Yarandi R, Khalili D, Tohidi M, Ramezani Tehrani F. A comparison of the effects of oral contraceptives on the clinical and biochemical manifestations of polycystic ovary syndrome: a crossover randomized controlled trial. *Hum Reprod* 2020;35:175–186
 17. Diamanti-Kandaraki E, Baillargeon JP, Luorino MJ, Jakubowicz DJ, Nestler JE. A modern medical quandary: polycystic ovary syndrome, insulin resistance, and oral contraceptive pills. *J Clin Endocrinol Metab* 2003;88:1927–1932
 18. Lopez LM, Grimes DA, Schulz KF. Steroidal contraceptives: effect on carbohydrate metabolism in women without diabetes mellitus. *Cochrane Database Syst Rev* 2012;18:CD006133
 19. 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;19:251–255
 20. Jadhakhan F, Marshall T, Ryan R, Gill P. Risk of chronic kidney disease in young adults with impaired glucose tolerance/impaired fasting glucose: a retrospective cohort study using electronic primary care records. *BMC Nephrol* 2018;19:42
 21. 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:e012461
 22. Azziz R. Does the risk of diabetes and heart disease in women with polycystic ovary syndrome lessen with age? *Fertil Steril* 2017;108:959–960
 23. Echiburú B, Pérez-Bravo F, Galgani JE, et al. Enlarged adipocytes in subcutaneous adipose tissue associated to hyperandrogenism and visceral adipose tissue volume in women with polycystic ovary syndrome. *Steroids* 2018;130:15–21
 24. Karim ME, Gustafson P, Petkau J; Long-Term Benefits and Adverse Effects of Beta-Interferon for Multiple Sclerosis (BeAMS) Study Group. Comparison of statistical approaches for dealing with immortal time bias in drug effectiveness studies. *Am J Epidemiol* 2016;184:325–335
 25. Suissa S, Dell'aniello S, Vahey S, Renoux C. Time-window bias in case-control studies: statins and lung cancer. *Epidemiology* 2011;22:228–231
 26. Townsend P. Deprivation. *J Soc Policy* 1987;16:125–146
 27. Adams J, Ryan V, White M. How accurate are Townsend Deprivation Scores as predictors of self-reported health? A comparison with individual level data. *J Public Health (Oxf)* 2005;27:101–106
 28. Riley J, Antza C, Kempegowda P, et al. Social deprivation and incident diabetes-related foot disease in patients with type 2 diabetes: a population-based cohort study. *Diabetes Care* 2021;44:731–739
 29. Davé S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* 2009;18:704–707
 30. Nishida C, Barba C, Cavalli-Sforza T, et al.; WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;363:157–163
 31. Benchimol EI, Smeeth L, Guttman A, et al.; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12:e1001885
 32. Ollila M-ME, West S, Keinänen-Kiukaanniemi S, et al. Overweight and obese but not normal weight women with PCOS are at increased risk of Type 2 diabetes mellitus—a prospective, population-based cohort study. *Hum Reprod* 2017;32:423–431
 33. Hart R, Doherty DA. The potential implications of a PCOS diagnosis on a woman's long-term health using data linkage. *J Clin Endocrinol Metab* 2015;100:911–919
 34. Kakoly NS, Earnest A, Teede HJ, Moran LJ, Joham AE. The impact of obesity on the incidence of type 2 diabetes among women with polycystic ovary syndrome. *Diabetes Care* 2019;42:560–567
 35. Li HWR, Lam KSL, Tam S, et al. Screening for dysglycaemia by oral glucose tolerance test should be recommended in all women with polycystic ovary syndrome. *Hum Reprod* 2015;30:2178–2183
 36. Cheung LP, Ma RCW, Lam PM, et al. Cardiovascular risks and metabolic syndrome in Hong Kong Chinese women with polycystic ovary syndrome. *Hum Reprod* 2008;23:1431–1438
 37. Yao Q-M, Wang B, An X-F, Zhang J-A, Ding L. Testosterone level and risk of type 2 diabetes in men: a systematic review and meta-analysis. *Endocr Connect* 2018;7:220–231
 38. Ehrmann DA, Schneider DJ, Sobel BE, et al. Troglitazone improves defects in insulin action, insulin secretion, ovarian steroidogenesis, and fibrinolysis in women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 1997;82:2108–2116
 39. O'Reilly MW, Kempegowda P, Walsh M, et al. AKR1C3-mediated adipose androgen generation drives lipotoxicity in women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2017;102:3327–3339
 40. Skov V, Glinborg D, Knudsen S, et al. Reduced expression of nuclear-encoded genes involved in mitochondrial oxidative metabolism in skeletal muscle of insulin-resistant women with polycystic ovary syndrome. *Diabetes* 2007;56:2349–2355
 41. Hansen SL, Svendsen PF, Jeppesen JF, et al. Molecular mechanisms in skeletal muscle underlying insulin resistance in women who are lean with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2019;104:1841–1854
 42. Navarro G, Allard C, Morford JJ, et al. Androgen excess in pancreatic b cells and neurons predisposes female mice to type 2 diabetes. *JCI Insight* 2018;3:e98607
 43. Xu W, Schiffer L, Qadir MMF, et al. Intracrine testosterone activation in human pancreatic b-cells stimulates insulin secretion. *Diabetes* 2020;69:2392–2399
 44. Kim SW, Jeon JH, Lee WK, et al. Long-term effects of oral contraceptives on the prevalence of diabetes in post-menopausal women: 2007–2012 KNHANES. *Endocrine* 2016;53:816–822
 45. Yao W, Dong X, Yu X, Luo J, Zhang D. The use of oral contraceptive is inversely associated with the risk of type 2 diabetes mellitus among middle-aged women. *Gynecol Endocrinol* 2021;37:758–763
 46. Rimm EB, Manson JE, Stampfer MJ, et al. Oral contraceptive use and the risk of type 2 (non-insulin-dependent) diabetes mellitus in a large prospective study of women. *Diabetologia* 1992;35:967–972
 47. Halperin IJ, Kumar SS, Stroup DF, Laredo SE. The association between the combined oral contraceptive pill and insulin resistance, dysglycemia and dyslipidemia in women with polycystic ovary syndrome: a systematic review and meta-analysis of observational studies. *Hum Reprod* 2011;26:191–201
 48. Luque-Ramírez M, Nattero-Chávez L, Ortiz Flores AE, Escobar-Morreale HF. Combined oral contraceptives and/or antiandrogens versus insulin sensitizers for polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update* 2018;24:225–241
 49. Mathur R, Levin O, Azziz R. Use of ethinylestradiol/drospirenone combination in patients with the polycystic ovary syndrome. *Ther Clin Risk Manag* 2008;4:487–492
 50. Rice MA, Malhotra SV, Stoyanova T. Second-generation antiandrogens: from discovery to standard of care in castration resistant prostate cancer. *Front Oncol* 2019;9:801
 51. Daniel S, Koren G, Lunenfeld E, Levy A. NSAIDs and spontaneous abortions - true effect or an indication bias? *Br J Clin Pharmacol* 2015;80:750–754
 52. 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;31:2841–2855
 53. Kakoly NS, Khomami MB, Joham AE, et al. Ethnicity, obesity and the prevalence of impaired glucose tolerance and type 2 diabetes in PCOS: a systematic review and meta-regression. *Hum Reprod Update* 2018;24:455–467

Appendices of Chapter 5 (PCOS and dysglycaemia)

Suppl. Table 1: Read codes used in data extraction

Condition	Read Code	Description
Polycystic ovary syndrome	C164.12	Stein - Leventhal syndrome
	C165.00	Polycystic ovarian syndrome
	7E25300	Endoscopic drilling of ovary
Polycystic ovaries	C164.00	Polycystic ovaries
	C164.13	Multi-cystic ovaries
Ovulatory dysfunction	1571	H/O: amenorrhoea
	K590.00	Absence of menstruation
	K590.11	Amenorrhoea
	K590100	Secondary amenorrhoea
	K590z00	Amenorrhoea NOS
	K591.00	Scanty or infrequent menstruation
	K591.11	Infrequent menstruation
	K591000	Hypomenorrhoea
	K591100	Oligomenorrhoea
	K591300	Secondary oligomenorrhoea
	K591z00	Scanty or infrequent menstruation NOS
	K5B0.00	Female infertility of anovulatory origin
	K5B0.11	Anovular cycle
	K5B0100	Secondary anovulatory infertility
	K5B0z00	Female infertility of anovulatory origin NOS
Alopecia	Myu6300	[X]Other androgenic alopecia
	M240.00	Alopecia
	M240z00	Alopecia NOS
	M240H00	Alopecia seborrhoeica
	M240000	Alopecia unspecified
	1N02.00	C/O: hair loss
	M240300	Frontal alopecia of women
	M240012	Hair loss
	M240200	Male pattern alopecia
	M240D00	Marginal alopecia
	22D7.11	O/E – alopecia
	22D4.00	O/E - loss of hair
Hirsutism	22D8.00	O/E – hirsutism
	22D8000	O/E - facial hair
	M241.00	Hirsutism - hypertrichosis

Suppl. Table 1 (Continued)

Condition	Read Code	Description
Type 2 diabetes mellitus	C106100	Diabetes mellitus, adult onset, + neurological manifestation
	C109611	Type II diabetes mellitus with retinopathy
	C10K000	Type A insulin resistance without complication
	C109.13	Type II diabetes mellitus
	C109.12	Type 2 diabetes mellitus
	C104.00	Diabetes mellitus with renal manifestation
	C10FQ00	Type 2 diabetes mellitus with exudative maculopathy
	C109.11	NIDDM - Non-insulin dependent diabetes mellitus
	C10F111	Type II diabetes mellitus with ophthalmic complications
	C109F00	Non-insulin-dependent d m with peripheral angiopath
	C10F911	Type II diabetes mellitus without complication
	C10FA11	Type II diabetes mellitus with mononeuropathy
	C105y00	Other specified diabetes mellitus with ophthalmic complicatn
	C100112	Non-insulin dependent diabetes mellitus
	C107200	Diabetes mellitus, adult with gangrene
	C109500	Non-insulin dependent diabetes mellitus with gangrene
	C10FJ11	Insulin treated Type II diabetes mellitus
	C104.11	Diabetic nephropathy
	C10FR11	Type II diabetes mellitus with gastroparesis
	C10z.00	Diabetes mellitus with unspecified complication
	C109F12	Type 2 diabetes mellitus with peripheral angiopathy
	C10D.00	Diabetes mellitus autosomal dominant type 2
	C10F100	Type 2 diabetes mellitus with ophthalmic complications
	C109F11	Type II diabetes mellitus with peripheral angiopathy
	C10F900	Type 2 diabetes mellitus without complication
	C10FA00	Type 2 diabetes mellitus with mononeuropathy
	C106z00	Diabetes mellitus NOS with neurological manifestation
	C109612	Type 2 diabetes mellitus with retinopathy
	C109511	Type II diabetes mellitus with gangrene
	C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy
	C109512	Type 2 diabetes mellitus with gangrene
	C10FP00	Type 2 diabetes mellitus with ketoacidotic coma
	C10F011	Type II diabetes mellitus with renal complications
	C101z00	Diabetes mellitus NOS with ketoacidosis
	C109E00	Non-insulin depend diabetes mellitus with diabetic cataract

Suppl. Table 1 (Continued)

No.	Read Code	Description
Type 2 Diabetes Mellitus	C10F811	Metabolic syndrome X
	C107100	Diabetes mellitus, adult, + peripheral circulatory disorder
	C109400	Non-insulin dependent diabetes mellitus with ulcer
	C105.00	Diabetes mellitus with ophthalmic manifestation
	C10FQ11	Type II diabetes mellitus with exudative maculopathy
	C10F000	Type 2 diabetes mellitus with renal complications
	C109E11	Type II diabetes mellitus with diabetic cataract
	C109E12	Type 2 diabetes mellitus with diabetic cataract
	C106y00	Other specified diabetes mellitus with neurological comps
	C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
	C109012	Type 2 diabetes mellitus with renal complications
	C109011	Type II diabetes mellitus with renal complications
	C10FS00	Maternally inherited diabetes mellitus
	C10F311	Type II diabetes mellitus with multiple complications
	C10y.00	Diabetes mellitus with other specified manifestation
	C109H00	Non-insulin dependent d m with neuropathic arthropathy
	C10C.00	Diabetes mellitus autosomal dominant
	C10FC11	Type II diabetes mellitus with nephropathy
	C10zy00	Other specified diabetes mellitus with unspecified comps
	C105100	Diabetes mellitus, adult onset, + ophthalmic manifestation
	C107400	NIDDM with peripheral circulatory disorder
	C10FD11	Type II diabetes mellitus with hypoglycaemic coma
	C109H12	Type 2 diabetes mellitus with neuropathic arthropathy
	C10F300	Type 2 diabetes mellitus with multiple complications
	C10FL11	Type II diabetes mellitus with persistent proteinuria
	C10H.00	Diabetes mellitus induced by non-steroid drugs
	C109H11	Type II diabetes mellitus with neuropathic arthropathy
	C104y00	Other specified diabetes mellitus with renal complications
	C10FC00	Type 2 diabetes mellitus with nephropathy
	C10FJ00	Insulin treated Type 2 diabetes mellitus
	C10FR00	Type 2 diabetes mellitus with gastroparesis
	C10P111	Type 2 diabetes mellitus in remission
	C10F211	Type II diabetes mellitus with neurological complications
	C109G00	Non-insulin dependent diabetes mellitus with arthropathy
	C105z00	Diabetes mellitus NOS with ophthalmic manifestation
	C10FB11	Type II diabetes mellitus with polyneuropathy
	C109700	Non-insulin dependent diabetes mellitus - poor control

Suppl. Table 1 (Continued)

No.	Read Code	Description
Type 2 Diabetes Mellitus	C109600	Non-insulin-dependent diabetes mellitus with retinopathy
	C10FK11	Hyperosmolar non-ketotic state in type II diabetes mellitus
	C10N.00	Secondary diabetes mellitus
	C109.00	Non-insulin dependent diabetes mellitus
	C10P100	Type II diabetes mellitus in remission
	C10F200	Type 2 diabetes mellitus with neurological complications
	C109G11	Type II diabetes mellitus with arthropathy
	C109G12	Type 2 diabetes mellitus with arthropathy
	C100z00	Diabetes mellitus NOS with no mention of complication
	C10..00	Diabetes mellitus
	C100100	Diabetes mellitus, adult onset, no mention of complication
	C10FB00	Type 2 diabetes mellitus with polyneuropathy
	C109711	Type II diabetes mellitus - poor control
	C109712	Type 2 diabetes mellitus - poor control
	C109212	Type 2 diabetes mellitus with neurological complications
	C109211	Type II diabetes mellitus with neurological complications
	C10FE00	Type 2 diabetes mellitus with diabetic cataract
	C109J00	Insulin treated Type 2 diabetes mellitus
	C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria
	C10G.00	Secondary pancreatic diabetes mellitus
	C10F511	Type II diabetes mellitus with gangrene
	C103z00	Diabetes mellitus NOS with ketoacidotic coma
	C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy
	C10A500	Malnutritn-relat diabetes melitus wth periph circul complctn
	C103100	Diabetes mellitus, adult onset, with ketoacidotic coma
	C10L.00	Fibrocalculous pancreatopathy
	C10N100	Cystic fibrosis related diabetes mellitus
	C107.00	Diabetes mellitus with peripheral circulatory disorder
	C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps
	C10FF11	Type II diabetes mellitus with peripheral angiopathy
	C109J12	Insulin treated Type II diabetes mellitus
	C109J11	Insulin treated non-insulin dependent diabetes mellitus
	C10A.11	Jamaica type diabetes
	C10FN11	Type II diabetes mellitus with ketoacidosis
	C10H000	DM induced by non-steroid drugs without complication
	C109B12	Type 2 diabetes mellitus with polyneuropathy
	C10F500	Type 2 diabetes mellitus with gangrene
	C109B11	Type II diabetes mellitus with polyneuropathy
	C108y00	Other specified diabetes mellitus with multiple comps

Suppl. Table 1 (Continued)

No.	Read Code	Description
Type 2 Diabetes Mellitus	C10yz00	Diabetes mellitus NOS with other specified manifestation
	C102.00	Diabetes mellitus with hyperosmolar coma
	C104100	Diabetes mellitus, adult onset, with renal manifestation
	C107.11	Diabetes mellitus with gangrene
	C109111	Type II diabetes mellitus with ophthalmic complications
	C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma
	C10M.00	Lipoatrophic diabetes mellitus
	C109112	Type 2 diabetes mellitus with ophthalmic complications
	C107.12	Diabetes with gangrene
	C10FL00	Type 2 diabetes mellitus with persistent proteinuria
	C10F411	Type II diabetes mellitus with ulcer
	C103y00	Other specified diabetes mellitus with coma
	C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy
	C109900	Non-insulin-dependent diabetes mellitus without complication
	C10zz00	Diabetes mellitus NOS with unspecified complication
	C10z100	Diabetes mellitus, adult onset, + unspecified complication
	C10N000	Secondary diabetes mellitus without complication
	C10FE11	Type II diabetes mellitus with diabetic cataract
	C109000	Non-insulin-dependent diabetes mellitus with renal comps
	C10B.00	Diabetes mellitus induced by steroids
	C10FM11	Type II diabetes mellitus with persistent microalbuminuria
	C10F400	Type 2 diabetes mellitus with ulcer
	C104z00	Diabetes mellitus with nephropathy NOS
	C109A11	Type II diabetes mellitus with mononeuropathy
	C109A12	Type 2 diabetes mellitus with mononeuropathy
	C109911	Type II diabetes mellitus without complication
	C109912	Type 2 diabetes mellitus without complication
	C10yy00	Other specified diabetes mellitus with other spec comps
	C109412	Type 2 diabetes mellitus with ulcer
	C109411	Type II diabetes mellitus with ulcer
	C10K.00	Type A insulin resistance
	C10FG00	Type 2 diabetes mellitus with arthropathy
	C10G000	Secondary pancreatic diabetes mellitus without complication
	C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma
	C10F711	Type II diabetes mellitus - poor control
	C107z00	Diabetes mellitus NOS with peripheral circulatory disorder
	C10FH11	Type II diabetes mellitus with neuropathic arthropathy

Suppl. Table 1 (Continued)

No.	Read Code	Description
Type 2 Diabetes Mellitus	C109300	Non-insulin-dependent diabetes mellitus with multiple comps
	C10FP11	Type II diabetes mellitus with ketoacidotic coma
	C10P.00	Diabetes mellitus in remission
	C109D12	Type 2 diabetes mellitus with hypoglycaemic coma
	C102z00	Diabetes mellitus NOS with hyperosmolar coma
	C109D11	Type II diabetes mellitus with hypoglycaemic coma
	C10F700	Type 2 diabetes mellitus - poor control
	C100.00	Diabetes mellitus with no mention of complication
	C102100	Diabetes mellitus, adult onset, with hyperosmolar coma
	C10y100	Diabetes mellitus, adult, + other specified manifestation
	C109311	Type II diabetes mellitus with multiple complications
	C109312	Type 2 diabetes mellitus with multiple complications
	C106.00	Diabetes mellitus with neurological manifestation
	C10FF00	Type 2 diabetes mellitus with peripheral angiopathy
	C10F.11	Type II diabetes mellitus
	C109K00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
	C10FN00	Type 2 diabetes mellitus with ketoacidosis
	C109C00	Non-insulin dependent diabetes mellitus with nephropathy
	C10F611	Type II diabetes mellitus with retinopathy
	C107y00	Other specified diabetes mellitus with periph circ comps
	C106.13	Diabetes mellitus with polyneuropathy
	C101.00	Diabetes mellitus with ketoacidosis
	C109200	Non-insulin-dependent diabetes mellitus with neuro comps
	C106.11	Diabetic amyotrophy
	C106.12	Diabetes mellitus with neuropathy
	C10FG11	Type II diabetes mellitus with arthropathy
	C10F.00	Type 2 diabetes mellitus
	C109C11	Type II diabetes mellitus with nephropathy
	C10B000	Steroid induced diabetes mellitus without complication
	C109C12	Type 2 diabetes mellitus with nephropathy
	C10F600	Type 2 diabetes mellitus with retinopathy
	C108z00	Unspecified diabetes mellitus with multiple complications
	44V2.00	Glucose tolerance test impaired
Impaired glucose regulation	8HIS.00	Referral for management of impaired glucose tolerance
	9NS0400	Referral for impaired glucose tolerance management offered
	9mX3.00	Impaired glucose regulation monitoring telephone invitation
	9mX4.00	Impaired glucose regulation monitoring verbal invitation
	9mX0.00	Impaired glucose regulation monitoring invitation 1st letter
	C313500	Glucose intolerance

	9mX2.00	Impaired glucose regulation monitoring invitation 3rd letter
	9mX1.00	Impaired glucose regulation monitoring invitation 2nd letter
	C11y300	Impaired fasting glycaemia
	9mX..00	Impaired glucose regulation monitoring invitation
	C11y200	Impaired glucose tolerance
	R102.11	[D]Prediabetes
	C11y500	Pre-diabetes
	R102.12	[D]Impaired glucose tolerance test
	C11y400	Impaired glucose regulation
	R10E.00	[D]Impaired glucose tolerance
	R10D000	[D]Impaired fasting glycaemia
	R10D011	[D]Impaired fasting glucose

Suppl. Table 2: Risk of type 2 diabetes and dysglycemia (composite outcome combining prediabetes and type 2 diabetes) in women with PCOS compared to population-based controls matched for sex, age, BMI and location of general practice.

	Type 2 Diabetes		Dysglycemia	
	Women with PCOS	Controls	Women with PCOS	Controls
Outcome events, n (%)	1483 (2.32)	1348 (1.09)	2885 (4.51)	2894 (2.34)
Person-years	304379	590335	299,551	585,307
Crude Incidence Rate/10,000 person years	48.7	22.8	96.3	49.4
Follow-up years, Median (IQR)	3.39 (1.34-7.16)	3.47 (1.39-7.18)	3.32 (1.32-7.03)	3.44 (1.38-7.11)
Unadjusted hazard Rate ratio (95% CI)	2.13 (1.98-2.29); p<0.001		1.95 (1.85-2.05); p<0.001	
Adjusted hazard Rate ratio (95% CI)	2.04 (1.89-2.20); p<0.001		1.87 (1.78-1.97); p<0.001	
Adjusted hazard Rate ratio (95% CI)*	1.98 (1.70-2.31); p<0.001		1.95 (1.76-2.16); p<0.001	
Adjusted hazard Rate ratio (95% CI)†	2.17 (1.88-2.51); p<0.001		1.93 (1.75-2.13); p<0.001	

Adjustments for age, ethnicity, Townsend deprivation quintile, BMI category, smoking status and baseline diagnosis of hypothyroidism, hypertension and prescription of lipid lowering drugs

*Sensitivity Analysis 1: Analysis of women with incident diagnosis of PCOS and their corresponding controls

†Sensitivity Analysis 2: Analysis of women with PCOS-specific diagnosis and their corresponding controls

Suppl. Table 3: Risk factors of type 2 diabetes and dysglycemia (composite outcome combining prediabetes and type 2 diabetes) among women with PCOS (n=64,051).

Risk factors	Type 2 Diabetes	Dysglycemia
Age	1.06 (1.05-1.07); p<0.001	1.06 (1.05-1.06); p<0.001
BMI Categories*		
Underweight/Normal weight	Ref	Ref
Overweight	3.10 (2.41-3.99); p<0.001	2.68 (2.31-3.11); p<0.001
Obese	11.74 (9.46-14.57); p<0.001	7.13 (6.27-8.10); p<0.001
Missing	4.97 (3.85-6.42); p<0.001	3.15 (2.68-3.70); p<0.001
Townsend deprivation score category		
1 (least deprived)	Ref	Ref
2	1.17 (0.97-1.42); p=0.103	1.14 (1.00-1.30); p=0.054
3	1.47 (1.23-1.75); p<0.001	1.27 (1.12-1.44); p<0.001
4	1.54 (1.28-1.84); p<0.001	1.39 (1.22-1.58); p<0.001
5 (most deprived)	1.89 (1.54-2.24); p<0.001	1.67 (1.46-1.91); p<0.001
Missing	1.37 (1.12-1.66); p=0.002	1.28 (1.11-1.46); p<0.001
Smoker categories		
Non-Smoker	Ref	Ref
Discontinued	1.06 (0.91-1.25); p=0.440	1.05 (0.93-1.17); p=0.430
Smoker	1.36 (1.20-1.54); p<0.001	1.30 (1.19-1.42); p<0.001
Missing	0.97 (0.78-1.22); p=0.816	0.86 (0.73-1.02); p=0.093
Ethnicity		
Caucasian	Ref	Ref
Mixed Race	1.75 (1.09-2.81); p=0.019	1.93 (1.40-2.66); p<0.001
Chinese/middle eastern/others	1.27 (0.68-2.37); p=0.456	1.57 (1.04-2.35); p=0.030
Black	1.15 (0.82-1.61); p=0.408	1.66 (1.35-2.05); p<0.001
South Asian	1.92 (1.56-2.37); p<0.001	2.41 (2.09-2.78); p<0.001
Missing	1.00 (0.89-1.11); p=0.945	1.01 (0.93-1.09); p=0.866
Concurrent Conditions at baseline		
Hypothyroidism	1.07 (0.85-1.34); p=0.560	1.17 (1.00-1.37); p=0.056
Hypertension	2.11 (1.79-2.49); p<0.001	1.99 (1.76-2.26); p<0.001
PCOS relevant variables		
Anovulation	1.21 (1.08-1.35); p=0.001	1.17 (1.08-1.27); p<0.001
Hirsutism	1.20 (1.05-1.36); p=0.007	1.15 (1.05-1.27); p=0.004
Hair loss	0.93 (0.77-1.13); p=0.493	0.97 (0.85-1.11); p=0.687
Prescriptions at baseline		
Contraceptives		
No Pill	Ref	Ref
COCP without anti-androgen	0.83 (0.72-0.94); p=0.005	0.77 (0.70-0.85); p<0.001
COCP with anti-androgenic progestin*	0.84 (0.73-0.97); p=0.020	0.85 (0.77-0.94); p=0.002
Single agent anti-androgen therapy^	1.20 (0.84-1.71); p=0.314	0.99 (0.74-1.31); p=0.926
Lipid lowering drugs	1.37 (0.95-1.96); p<0.001	1.35 (1.02-1.77); p=0.036

*Normal/Underweight: $<23.5 \text{ kg/m}^2$ for patients of South Asian ethnicity & $<25 \text{ kg/m}^2$ for patients of all other ethnic groups

Overweight: $23.5\text{-}27.5 \text{ kg/m}^2$ for patients of South Asian ethnicity & $25\text{-}30 \text{ kg/m}^2$ for patients of all other ethnic groups

Obese = $\geq 27.5 \text{ kg/m}^2$ for patients of South Asian ethnicity & $\geq 30 \text{ kg/m}^2$ for patients of all other ethnic groups

*cyproterone acetate/drospirenone

^cyproterone acetate/flutamide/finasteride

Suppl. Table 4. Risk and protective factors among 2407 women with dysglycemia and 2407 matched controls (n=4814)

Covariates	Adjusted OR (95% CI)
Combined oral contraceptive use	
No COCP	Ref
COCP without anti-androgen	0.72 (0.59-0.87); p<0.001
COCP with anti-androgenic progestin*	0.76 (0.63-0.91); p=0.003
Age	1.05 (0.99-1.12); p=0.106
Smoking Status	
Non-Smoker	Ref
Discontinued Smoker	0.96 (0.79-1.16); p=0.663
Smoker	1.46 (1.25-1.71); p<0.001
Missing	0.93 (0.69-1.24); p=0.613
BMI category*	
Normal Weight	Ref
Overweight	1.45 (0.91-2.31); p=0.118
Obese	1.72 (0.93-3.16); p=0.082
Missing	.
Ethnicity	
Caucasian	Ref
Mixed Race	2.40 (1.33-4.34); p=0.009
Others	1.71 (0.83-3.53); p=0.107
Black	2.03 (1.32-3.12); p<0.001
South Asians	3.39 (2.51-4.59); p<0.001
Missing	0.98 (0.85-1.12); p=0.854
Townsend deprivation score category	
1 (least deprived)	Ref
2	1.19 (0.96-1.46); p=0.111
3	1.40 (1.14-1.73); p=0.001
4	1.41 (1.14-1.74); p=0.001
5 (most deprived)	1.75 (1.39-2.20); p<0.001
Missing	1.53 (1.22-1.92); p<0.001
Concurrent Conditions at baseline	
Hypothyroidism	1.35 (1.09-1.68); p=0.006
Hypertension	2.97 (2.48-3.56); p<0.001
Prescriptions at baseline	
Metformin	1.50 (1.24-1.81); p<0.001
Single agent anti-androgen therapy^	1.74 (0.97-3.13); p=0.063
Lipid lowering medication	1.21 (0.58-2.52); p=0.607

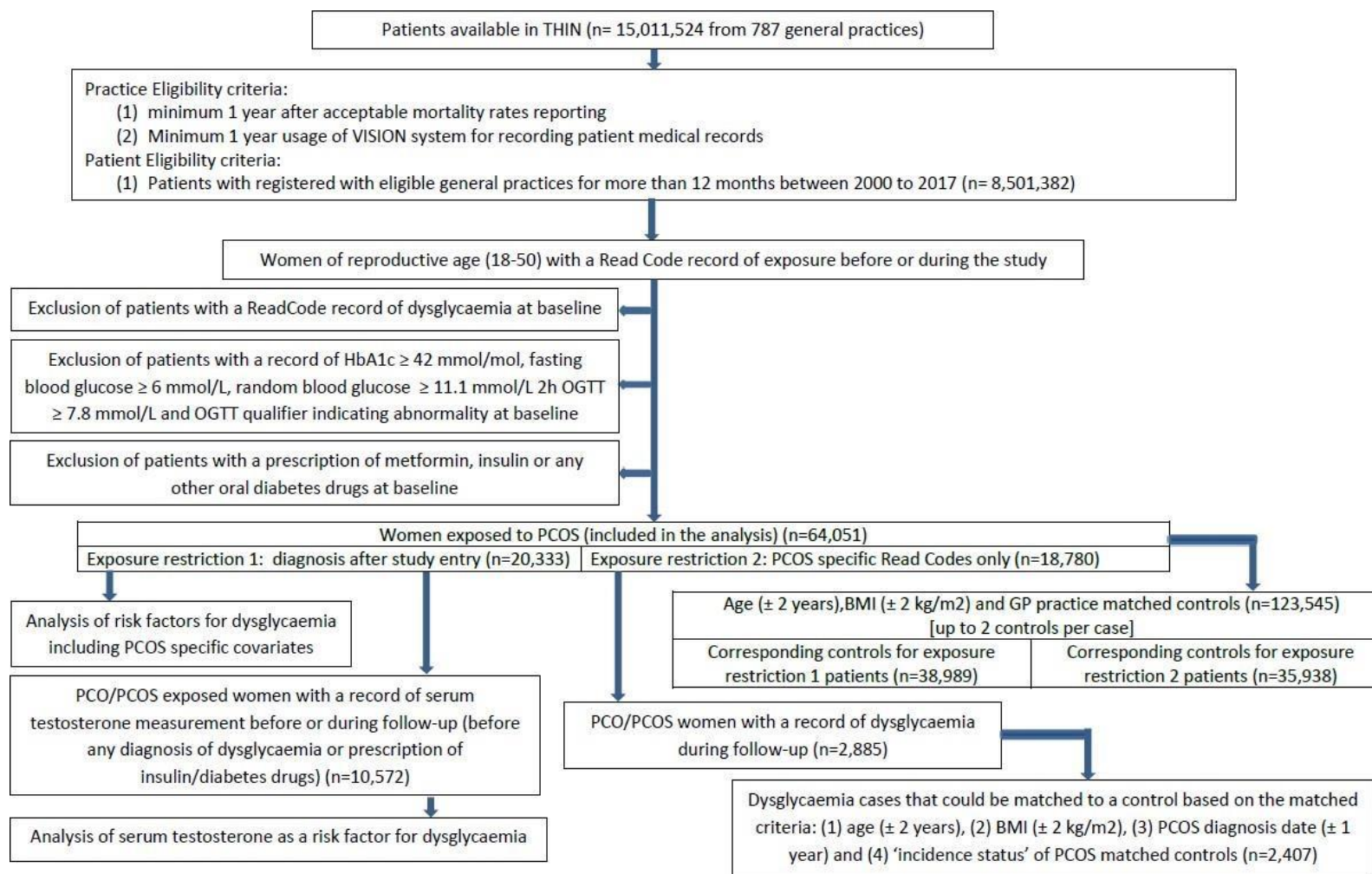
*Normal/Underweight: <23.5 kg/m² for patients of South Asian ethnicity & <25 kg/m² for patients of all other ethnic groups

Overweight: 23.5-27.5 kg/m² for patients of South Asian ethnicity & 25-30 kg/m² for patients of all other ethnic groups

Obese = ≥ 27.5 kg/m² for patients of South Asian ethnicity & ≥ 30 kg/m² for patients of all other ethnic groups

*cypoterone acetate/drospirenone

^cypoterone acetate/flutamide/finasteride

Suppl. Figure 1: Flowchart

Chapter 6 : General Discussion

6.1. Summary of findings

In this thesis, I primarily aimed to determine the risk of NAFLD, OSA, and dysglycemia among women with PCOS compared to women without PCOS using routinely collected UK electronic general practice open retrospective cohort data.

In my article published in PLOS Medicine on the risk of NAFLD, I aimed 1) to determine the unadjusted risk of prevalent and incident NAFLD among women with PCOS when compared with women without PCOS and to derive the risk estimates adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation up to end of follow-up, and hypothyroidism at baseline, 2) To determine if the risk is independent of their BMI status, 3) to determine the factors associated with risk of NAFLD among the women with PCOS, 4) to determine the risk of NAFLD by different levels of serum testosterone and SHBG in independent cohorts.

My article published in the European Journal of Endocrinology on the risk of OSA, I aimed 1) to determine the unadjusted risk of prevalent and incident OSA among women with PCOS when compared with women without PCOS and to derive the risk estimates adjusted for age, Townsend score, BMI, diabetes or impaired glucose regulation up to end of follow-up, and hypothyroidism at baseline. 2) to determine if the risk is independent of their BMI status and 3) to determine the factors associated with risk of OSA among the women with PCOS.

In my article published in Diabetes Care, I aimed 1) To determine the unadjusted risk of prevalent and incident dysglycaemia among women with PCOS when compared with women without PCOS and to derive the risk estimates adjusted for age, Townsend score, BMI, ethnicity, smoking status, hypothyroidism, hypothyroidism, and prescription with lipid lowering medications. 2) To determine if the risk is independent of their BMI status, 3) To

determine the factors associated with risk of dysglycaemia among the women with PCOS. 4) To determine the risk of dysglycaemia by different levels of testosterone and SHBG in independent cohorts.

6.2. Comparison with literature

6.2.1. The risk of NAFLD among women with PCOS

A meta-analysis of 17 studies found that women with PCOS have an increased risk of NAFLD (hazard ratio [HR] = 2.23, 95% CI 1.86 to 2.66) and a higher prevalence of NAFLD (odds ratio [OR] = 2.54, 95% CI 2.19 to 2.95) compared to women without PCOS (Rocha et al., 2017). Another meta-analysis of 17 observational studies found that women with PCOS had a higher risk of NAFLD (odds ratio [OR] = 2.25, 95% CI 1.95 to 2.60) compared to women without PCOS, even after adjusting for potential confounding factors (Wu et al., 2018). Another meta-analysis which included 7148 subjects from 23 studies (Shengir et al., 2021) estimated 2.5 times higher risk of NAFLD (OR=2.49, 95% CI: 2.20–2.82) among women with PCOS compared to their controls. Increased risk of NAFLD among women with PCOS is also proven in a histopathological study (Sarkar et al., 2020) observing severe hepatocyte ballooning (OR=3.4, 95% CI: 1.1-10.6) and advanced fibrosis (OR=7.1, 95% CI: 1.3-39) following adjustment for age and BMI. These findings support our own observation of an increased risk of NAFLD among women with PCOS.

6.2.2. The risk of OSA among women with PCOS

The pathophysiological plausibility of the role of PCOS in the development of OSA in our study is supported by the findings of a study by Simon et al. (2019) that assessed the circadian rhythm and metabolic function in adolescent girls with PCOS. Compared to controls, the girls with PCOS had a later offset timing of morning melatonin. Although the confidence interval was wide, suggesting that the study was underpowered, Vgontzas et al.

(2001) found that sleep disordered breathing was nearly 30 times more common in women with PCOS (OR = 30.6, 95% CI: 7.2-139.4) than in women without PCOS. The Apnoea Hypopnea Index (AHI), an index that quantifies the severity of sleep disordered breathing, was also significantly higher in 18 women with PCOS (22.5 ± 6.0 vs 6.7 ± 1.0) than in their age- and BMI-matched controls (Fogel et al., 2011).

6.2.3. The risk of dysglycaemia among women with PCOS

A meta-analysis of 23 observational studies found that obese women with PCOS had a 3.85 times higher risk of developing type 2 diabetes than non-obese women with PCOS. The study included 319,780 participants, of whom 60,336 had PCOS and 8847 had type 2 diabetes (Anagnostis et al., 2021).

PCOS was included as a risk factor for type 2 diabetes in the 2017 update of NICE guidelines and in the 2018 version of QDiabetes-2018, a revised prediction algorithm to predict 10-year risk of diabetes. This supports our findings, which are also consistent with a systematic review and meta-regression of 40 studies that found an increased prevalence of impaired glucose tolerance (3.26, 95% CI: 2.17–4.90) and type 2 diabetes (OR=2.87, 95% CI: 1.44–5.72) in women with PCOS. The higher estimate in the meta-regression compared to our study may be due to the inclusion of global data with a higher risk of diabetes, such as Asians, the higher degree of heterogeneity in the meta-analysis ($I^2=85.9\%$), systematic differences in the diagnosis of PCOS and dysglycemia in the general practice data, and the potential selection of women with high risk of diabetes, such as those with obesity or subfertility. Additionally, the primary studies did not adjust for confounders.

A 24-year follow-up of 27 women with PCOS and 94 controls in Sweden observed an elevated prevalence of T2DM among women with PCOS (19%) compared with women without PCOS (1%). The authors attributed this to the selection of women with subfertility or hirsutism and the smaller sample size. (Forslund et al., 2020).

6.2.4. The role of BMI in the NAFLD, OSA and dysglycaemia among women with PCOS

We demonstrated that the risk of NAFLD, OSA, and dysglycemia was independent of BMI in women with PCOS. However, the risk of these conditions was higher in women with PCOS and higher BMI than in women with PCOS and lower BMI. This association was independent of BMI and increased with increasing BMI, as supported by other studies. For example, a mediation analysis in a 25-year cohort found that visceral adipose tissue volume, as measured by year-25 CT scan, accounted for 41.0% of the association between free testosterone and prevalent NAFLD. Similarly, a meta-analysis of 17 observational studies found that women with PCOS and obesity had a 3.01 times higher risk of NAFLD than women with PCOS but without obesity. Finally, waist-hip ratio was significantly associated with Apnoea Hypopnea Index (AHI) in 18 women with PCOS, when compared with their age and BMI matched controls.

6.2.5. Role of androgen excess in NAFLD, and dysglycaemia among women with PCOS

The pathological plausibility of the role of sex hormones in the development of NAFLD in our study is supported by the findings of a nonhuman primate polycystic ovary syndrome model (Roy et al., 2022). This model showed that hepato-specific androgen receptors contribute to hepatic steatosis; androgen levels affect hepatic genes and metabolic pathways; the hepatic circadian timing system is affected by epigenetic modification.

Our findings on the role of androgen excess in the development of NAFLD are supported by many primary studies and meta-analyses. For example, Jones et al. (2012) found that women

with hyperandrogenic PCOS had a significantly higher hepatic fat percentage (% CH₂/H₂O) than women with PCOS without hyperandrogenic features or women without PCOS. Similarly, a meta-analysis of 12 studies found that the prevalence of NAFLD was nearly twice as high in postmenopausal women as in premenopausal women after adjusting for age and metabolic factors (Jaroenlapnopparat et al., 2023).

Studies have shown that hyperandrogenism is a major risk factor for NAFLD in women with PCOS. A meta-analysis of 17 studies found that women with both NAFLD and PCOS had significantly higher levels of total testosterone and Free Androgen Index (FAI) than women with PCOS without NAFLD (Rocha et al., 2017). Similarly, another meta-analysis of 17 studies found that women with hyperandrogenism and PCOS had a 3.31 times higher risk of NAFLD than women with PCOS and no hyperandrogenism (Wu et al., 2018). These findings were supported by a study of adolescent girls with PCOS, which found that hyperandrogenism was significantly associated with NAFLD (Kara et al., 2023).

A cohort study with a 72% retention rate after 25 years of follow-up found that free testosterone levels measured in year 2 were associated with an increased prevalence of NAFLD in year 25. This association remained significant after adjusting for baseline and changes in waist circumference, HOMA-IR, and triglycerides. Even after excluding women who met the criteria for hyperandrogenism, the association between free testosterone and NAFLD was still significant, suggesting that testosterone may play a role in the development of NAFLD in women without androgen excess likely through raising insulin resistance, inflammation and oxidative stress and increased expression of sex hormone receptors in liver cells would adversely affect liver health even in the absence of hyperandrogenism even with low grade androgen signalling (Jaruvongvanich et al., 2017).

A positive correlation was observed between melatonin variables (later offset timing of morning melatonin and longer duration of melatonin secretion) and serum testosterone

concentrations in adolescent girls with PCOS, as compared to age-, activity-, and BMI-matched controls (Simon et al., 2019).

A positive correlation was observed between the severity of androgen phenotype and dysglycaemia in women with PCOS (O'Reilly et al., 2014).

Although combined oral contraceptive pills reduced the risk of dysglycaemia, the reason for not observing a larger effect size for would be due to unadjusted confounders such as lifestyle factors.

6.2.6. Interrelationships between NAFLD, OSA, and Dysglycemia in PCOS and Non-PCOS Women

In this section, I delve into the intricate interrelationships between NAFLD, OSA, dysglycemia, and androgen excess, exploring their connections not only within the context of PCOS but also among women without PCOS, with a focus on the underlying pathophysiological basis.

Sex hormones, oxidative stress, endothelial dysfunction, sympathetic activity, obesity, insulin resistance, hyper-insulinemia, androgen excess, microbiome, and genetic susceptibility are all believed to contribute to NAFLD in women with PCOS (Kahal et al., 2017). Steroids such as androgens, oestrogen, and progesterone, as well as body fat, are thought to contribute to OSA in women with PCOS (Doycheva & Ehrmann, 2022). A positive correlation was found between melatonin variables (later offset timing of morning melatonin and longer duration of melatonin secretion) and markers of insulin resistance (HOMA-IR) in adolescent girls with PCOS when compared to age, activity, and BMI-matched controls (Simon et al., 2019). Inadequate sleep duration was associated with an increased risk of NAFLD in a meta-analysis of 261,554 people (Yang et al., 2022) and another meta-analysis (Wijarnpreecha et al., 2022).

NAFLD has a bidirectional relationship with cardiovascular diseases, acting as both a risk factor and an outcome. A study by Targher et al. (2008) found that the risk of cardiovascular disease was 2.2 times higher among persons with NAFLD independent of age, sex, adiposity measures, and other common metabolic risk factors. Mantovani et al. (2021) also found a causal relationship between NAFLD and incident diabetes, with the risk of diabetes being 2.2 times higher among persons with NAFLD. Patel-Sanchez et al. (2023) found that the prevalence of NAFLD was 2.5 times higher among adolescent women with obesity/overweight, T2DM, and PCOS when compared with adolescent women with obesity/overweight and PCOS but not with T2DM. In line with the findings of these studies, we observed an increased risk of dysglycemia among women with NAFLD when compared with women without NAFLD. The hazard ratio (aHR) was 2.57, with a 95% confidence interval of 1.65 to 4.01.

Our findings are also supported by a multivariable Mendelian randomisation analysis (Liu et al., 2023) among women with PCOS, which found a causal association between NAFLD (OR=1.11, 95% CI: 1.05 to 1.17), fasting insulin (OR=3.11, 95% CI: 1.68 to 5.76), and serum bioavailable testosterone (OR=1.90, 95% CI: 1.27 to 2.85). Similarly, a meta-analysis and meta-regression of 2120 participants (Jullian-Desayes et al., 2021) found that age, apnea-hypopnea index (AHI) > 5/h, body mass index (BMI) > 26 kg/m², and type 2 diabetes are risk factors for steatosis. These findings support our study's findings of an interplay between NAFLD, OSA, and dysglycaemia. Resembling to our study, a meta-analysis and meta-regression among women with PCOS (Manzano-Nunez et al., 2023) estimated similar risk factors for NAFLD namely, BMI (OR=1.35, 95%CI: 1.28 to 1.43), Free androgen index (OR=1.06, 95%CI: 1.03 to 1.1) and hyperandrogenism (OR=1.35, 95%CI: 1.28 to 1.43).

The association of tamoxifen use with incident and prevalent non-alcoholic fatty liver disease (NAFLD) (Lee et al., 2020) is worth considering, as tamoxifen is one of the agents

used for ovulation induction in women with polycystic ovary syndrome (PCOS) (Ignatov & Ortmann, 2020).

The potential interaction of insulin resistance in the association between polycystic ovary syndrome (PCOS) and non-alcoholic fatty liver disease (NAFLD) (Sarkar et al., 2017) was accounted for in our analysis. This is because our adjusted estimates included diabetes or impaired glucose regulation. In addition, among women with PCOS, fasting plasma insulin levels were significantly higher (306 ± 52 vs. 177 ± 18 pmol/L) in those with sleep disordered breathing (SDB) when compared with those without SDB (Vgontzas et al., 2001). These findings support the need to adjust for diabetes or impaired glucose regulation when determining the association between PCOS and OSA.

6.3. The implications of this research for practice and policy

Our research has important implications for the prevention of NAFLD, OSA, and dysglycemia in women with PCOS across all levels of care.

Our research findings could help raise awareness among the public, primary care staff, health administrators, and policy makers. Specific awareness of the additional risk of high BMI, higher levels of serum testosterone, and lower levels of serum SHBG could be useful in designing targeted interventions for these high-risk categories. Creating awareness about PCOS, NAFLD, OSA, and dysglycemia could help with early detection, increased motivation for lifestyle changes, improved health literacy, reduced stigma, improved health-seeking behavior, personalized care plans, enhanced training for healthcare staff, public health initiatives, improved quality of life, integration of health policy, and research and innovation. The interrelationship of these conditions underscores the need for a multidisciplinary team approach to their detection and management and clinical practice guidelines for PCOS can consider broader complications of PCOS.

The observations of our study are relevant to improving documentation in general practices in the UK, as they reveal that BMI is missing in one in six women and ethnicity codes in half of the sample.

6.4. The impact of this research

The impact of this research is evident in the widespread citation of this study in clinical guidelines, overviews, meta-analyses, narrative reviews, expert committee reports, cohort studies, book chapters, and other articles.

One of the clinical guidelines, which cites our study, includes diabetes as a complication of polycystic ovary syndrome (PCOS). The guideline recommends screening for diabetes at baseline and follow-up screening at intervals of one to three years, depending on the level of risk (Teede et al., 2018).

Our article on the risk of NAFLD in PCOS has been cited 155 times, including in overviews (Muthiah et al., 2022), systematic reviews (Shahbaz et al., 2022; Teede et al., 2023; Zaman et al., 2022), meta-analyses (Lavrentaki et al., 2019; Liao & Cao, 2020; Ramezani Tehrani et al., 2021), narrative reviews (Allen et al., 2022; Glintborg & Andersen, 2020), expert committee reports, and clinical practice guidelines.

The clinical practice guidelines that cited our article include the American Association of Clinical Endocrinology (AACE) Clinical Practice Guideline for the Diagnosis and Management of Non-alcoholic Fatty Liver Disease in Primary Care and Endocrinology Clinical Settings (also known as the AASLD guidelines) (Cusi et al., 2022) and the Korean Association for the Study of the Liver (KASL) clinical practice guidelines for the management of non-alcoholic fatty liver disease (Kang et al., 2021). The Multidisciplinary Androgen Excess and PCOS Committee report on female adult acne and androgen excess

(Carmina et al., 2022), an expert committee report, also cited our study.

Our paper on the risk of OSA in PCOS has been cited 155 times including in systematic review and meta-analysis (Delanerolle et al., 2022; Ramezani Tehrani et al., 2021; J. Zhang et al., 2022), overview (Hirsch & Gaultney, 2023), narrative reviews (Barber et al., 2019; Barber & Franks, 2021; Islam et al., 2022; Sam & Ehrmann, 2019; Witchel et al., 2020), expert committee report such as Multidisciplinary Androgen Excess and PCOS Committee (Carmina et al., 2022), cohort studies (Subramanian et al., 2021) and chapter in a book (Shaib& Attarian, 2023).

Our paper on risk of dysglycaemia among women with PCOS has five citations (Gunn, 2022;McDonnell et al., 2022; Singh et al., 2022; Subramanian et al., 2022).

While these citations would have predominantly reached researchers and health workers, wider publicity in the lay media would have contributed sensitisation among policy makers and improvement in health literacy and health seeking behaviour among the patients and public.

6.5. Strengths and limitations

We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (von Elm et al., 2007) in preparing our three articles. We chose to use a retrospective cohort design (Talari & Goyal, 2020), which is the best possible study design to use in general practice data. We minimized the limitations of this design by 1) matching comparators by age, BMI, and practice, 2) using stringent selection criteria, such as age range for recruitment and follow-up start and end dates, 3) selecting various data sources within the THIN database, namely patient, therapy, medical, and additional health data (AHD) files (Turner, 2016), 4) relying on documentation on electronic records throughout the follow-up period, rather than on recall from the participants in conventional retrospective

cohort studies, 5) performing sensitivity analysis based on different case definitions and prevalent versus incident cases, rather than filling the missing variables using approximation or assumptions, and 6) matching controls with the same general practice and age to minimize discrepancy in the care provider and care protocol, including systematic difference in referral over the follow-up period.

Despite these measures to minimise the limitations, our study is still likely to have the limitations of retrospective cohort studies such as 1) systematic differences in health seeking, including severity of PCOS, to a general practice, 2) changes in the diagnostic criteria and clinical guidelines over the follow-up period, and 3) the inability to generalize to community-based selection of participants.

Although general practices in the UK are generally situated within a 20-minute walk on average (Todd et al., 2015), there are still considerable spatial accessibility issues (Bauer et al., 2018). These include barriers to registration, such as the perceived lack of proof of address (Hodson et al., 2019), limited access during routine office hours (Reilly et al., 2020), and barriers for access by specific groups, such as homeless people (Hirst & Cuthill, 2021), vulnerable migrants (Blane & Ciftci, 2022), commercial sex workers (Potter et al., 2022), ethnic minorities (Magadi & Magadi, 2022), young adults (Davey et al., 2013), and women (Dixon et al., 2021). Ethnicity is documented for only half of the participants. Among those with recorded ethnicity, the relatively higher Caucasian among women with PCOS (47.6%) compared to the control group (40.7%) could be attributed to Caucasians having systematically increased prevalence of PCOS, health seeking behaviour, healthcare provision such as establishing diagnosis or follow-up care, chance of getting documented of their ethnicity by their GP.

We derived the data from the 'The Health Improvement Network' (THIN) database (Cai et al., 2012), which has been used in 760 publications listed in PubMed. The use of general

practice data can have many limitations, especially in the context of PCOS. For example, an Australian retrospective cohort study tracing a birth cohort noted that only one in three (32%) women with PCOS had a prior diagnosis of PCOS (March et al., 2010). This suggests that general practice data may miss women with milder forms of PCOS, especially those with polycystic morphology.

The THIN database is generated from routine general practice data and is not specific for research purposes. This means that it is possible that some women with PCOS may have been missed, as the database does not include all the information necessary to make a diagnosis. For example, polycystic ovaries are one of the three criteria used to diagnose PCOS, but this requires an ultrasound scan, which may not have been performed in all cases. Sensitivity analysis based on the PCOS phenotypes could not be performed due to potential missing data required to construct phenotypes of PCOS. Therefore, I provided sensitivity analysis based on individual features, namely, anovulation, clinical androgen excess, and polycystic ovaries.

Our findings of 12.9% hirsutism among women with PCOS are comparable with the estimates of 13% (95% CI: 10 to 17) and 13% (95% CI: 8 to 20) in two meta-analyses (Amiri et al., 2022; Bozdag et al., 2016). However, our observation of 76.5% polycystic ovaries is slightly lower than the estimate of 82.5% among women with PCOS who were recruited from a specialist endocrine clinic in Sri Lanka (Wijeyaratne et al., 2011). This difference is likely due to the fact that some women with PCOS may have been missed in the THIN database, as they may not have undergone an ultrasound scan.

Overall, our findings suggest that the THIN database is a valuable resource for research on PCOS, but it is important to be aware of its limitations. Further research is needed to confirm our findings and to investigate the reasons for the differences in prevalence of polycystic ovaries between the THIN database and other studies.

The adverse impact on the estimates of this study would be relatively lower, as this study does not aim to estimate the prevalence of PCOS. The estimates may be generalisable to women with PCOS who are routinely diagnosed in general practice or in hospitals, as hospital-based care is accessed through referrals from general practice (Grosios et al., 2010; Madan et al., 2017).

Constructing a case definition of PCOS using Read codes from general practice data had some constraints. Although codes to represent androgen excess, anovulation, and polycystic ovaries would likely have been entered separately, Read codes specific to the diagnosis of PCOS might not have been entered. Additionally, there is a likelihood that codes with close resemblance would be chosen interchangeably, such as Read Code PCOS (Polycystic ovary syndrome) and Read Code PCO (Polycystic ovaries), when selecting an appropriate code from a drop-down list.

Therefore, we developed two case definitions, namely, 1) "PCOS," which includes a broader category that includes polycystic ovaries and 2) "PCOS specific Read codes only".

We performed sensitivity analysis for scenario-based analysis. A similar approach was used by Ding et al. in their study on the diagnosis and management of PCOS in the UK using the THIN database, in which they derived case definitions separately for diagnosed cases and probable cases (Ding et al., 2016).

Despite these limitations, general practice data can provide real-world evidence to obtain the best possible estimates. Models developed using routinely available variables were able to infer the likelihood of NASH or advanced fibrosis in adult diabetic patients with NAFLD with a sensitivity of 90% but a specificity of 57% (Bazick et al., 2015).

Another important strength of our study is the larger sample size, with women with possible

PCOS (around 60,000 possible PCOS, 20,000 with PCOS Read codes, and 20,000 incident cases) and more than 120,000 women without PCOS. The study also included a relatively larger number of outcomes, namely, NAFLD (n=496), OSA (n=520), T2DM (n=2831), and Dysglycaemia (n=5779). This larger sample size enabled us to perform sensitivity and stratified analyses to assess the role of bias and confounders.

We excluded participants who had documented evidence of excessive alcohol consumption, as per the guidance on the diagnosis of NAFLD (Westfall et al., 2020), to minimize sampling bias. Controls were randomly matched to women with PCOS at a ratio of 2:1 to increase the power of the study (Setia, 2016). The median follow-up was around 3.5 years for both groups of women in our study, which is believed to be sufficient for the development of the outcome.

For the reporting of follow-up time, we reported only the median follow-up. However, recent evidence suggests that the measure with the narrower confidence interval should be chosen. A retrospective cohort study in Korea with 3-year median follow-up (Won et al., 2021) observed metabolic syndrome and hyperandrogenism as risk factors for NAFLD among women with PCOS. A Taiwanese study (Lin et al., 2017) with a slightly longer median follow-up of 4.5 and 5.2 years, respectively, found an adjusted estimate for the risk of OSA of 2.63 (95% CI: 1.57 to 4.04). This is closer to the risk estimates of 2.87 (2.10–3.92) in our study.

A study in Hong Kong, China (Ng et al., 2023) with 6.3-year follow-up estimated an adjusted hazard ratio of 4.4 (95% CI: 3.9–5.0) for diabetes. Another study conducted in Iran (Kazemi Jaliseh et al., 2017) with a median follow-up of 12.9 years estimated adjusted hazard ratios of 4.9 (95% CI: 2.5–9.3) for diabetes and 1.7 (95% CI: 1.1–2.6) for OSA. The higher hazard ratios in these two studies could be attributed to longer duration of follow-up and ethnic differences. For example, Iranian and Chinese women have a higher risk of

diabetes than women in other countries (Peykari et al., 2023; W. Yang et al., 2010). The median age of participants in our study (30 years, IQR 25 ± 35 years) indicates that the majority of the participants had follow-up before 40 years. This is because the risk of diabetes among women with PCOS sharply declines closer to the level of women without PCOS when the age increases above 40 years (Kazemi Jaliseh et al., 2017).

We performed a baseline comparison of age, BMI, medical conditions, smoking, and Townsend index. Age is an important risk factor for the development of NAFLD (Doycheva et al., 2017). A retrospective analysis of whole-night polysomnography of 23,806 patients in Israel (Gabbay & Lavie, 2012) revealed a linear increase in OSA with increasing age and BMI. The prevalence of OSA among women with PCOS increases with advancing age (Kahal et al., 2020). Among women stratified by age (18-24, 25-29, and 30-34 years), the hazard ratio of developing T2DM attributable to PCOS decreased from 10.4 to 5.3 to 4.1, respectively (W.-T. Liao et al., 2022). BMI is a risk factor for NAFLD (Schwimmer et al., 2005) and T2DM (Kakoly et al., 2019). Smoking is a risk factor for NAFLD (Rezayat et al., 2018) and OSA (Zeng et al., 2023). There is no gender-based difference in the risk of T2DM (Yuan et al., 2019). Hypertension shows a bidirectional association with NAFLD (Hu et al., 2023), OSA (Floras, 2015), and T2DM (Sun et al., 2019). Hypothyroidism is associated with NAFLD (Ludwig et al., 2015), OSA (Thavaraputta et al., 2019; M. Zhang et al., 2016), and T2DM (Rong et al., 2021). Diabetes or impaired glucose regulation is associated with NAFLD (Hazlehurst et al., 2016) and OSA (Subramanian et al., 2019). Social deprivation is associated with NAFLD (Cho et al., 2021), OSA (Etindele Soso & Matos, 2021), and T2DM (Derks et al., 2017).

Although using BMI at baseline in this retrospective cohort with 3.5 years of median follow-up would increase the likelihood of misclassification bias, the impact would have been lower, as there is lack of evidence for effective weight reduction programs in the UK based

on women's experience (Hillman et al., 2020) and a national survey on management of weight reduction in the UK (Sharma et al., 2010).

With regards to the evaluation of outcome, general practice data lacked documentation of severity of NAFLD and OSA and we could not determine the factors associated with the phenotypes of NAFLD and OSA. Although ascertainment bias due to systematically increased case detection of NAFLD, OSA and T2DM among women with higher BMI (Katrina Loomis et al., 2016), this bias would have been minimised by matching based on BMI, analysis stratified by BMI and adjustment for BMI in regression analysis. Quality of documentation could also be a limitation as one in seven practices were found to have poor quality documentation of diabetes related codes (Tate et al., 2017).

Limitations in obtaining ethnicity data in PCOS studies are relatively common. A meta-analysis of 40 studies noted that most studies did not include ethnicity/race data (Kakoly et al., 2018).

With regards to the risk measures in our study, an underestimate is possible due to the inclusion of women with Read codes for polycystic ovaries under women with PCOS in the primary analysis, and the inclusion of women with undiagnosed outcomes (NAFLD, OSA, or T2DM) in the general practice setting.

An overestimate is possible due to women with mild forms of PCOS either not seeking care or not being diagnosed, or their diagnosis not being documented in the general practice. The approach to mitigate the issue of missing the diagnoses among those with mild PCOS would involve conducting a prospective cohort study with a random sample from community. However, such a study would be resource intensive and require longer duration to achieve adequate statistical power.

6.6. Future investigations

Future research could focus on determining the effectiveness of antiandrogen medication in reducing NAFLD, OSA, and T2DM in women with PCOS. This could be done initially through nested case-control based pharmaco-epidemiological studies, and if the evidence is convincing, through randomised controlled trials. Similarly, the role of lipid-lowering drugs in reducing the risk of T2DM in women with PCOS could be explored. The role of androgen excess on the severity of hepatic manifestations and OSA in women with PCOS could also be investigated. The effectiveness of existing prevention and treatment interventions in women with PCOS could be compared with women without PCOS. Determining the multimorbidity of NAFLD, OSA, and T2DM in women with PCOS in a single cohort would help to understand the interlink between these conditions.

6.7. Conclusion

Women with polycystic ovary syndrome have a two-fold increased risk of developing complications namely, NAFLD, OSA and T2DM/Dysglycaemia. Contrary to the common belief that complications of PCOS are only seen in women with increased BMI, we found that women with PCOS and normal BMI are also at an increased risk of developing these complications. Hence, the risk for these complications is independent of BMI.

Considering women without PCOS and with BMI $<25 \text{ kg/m}^2$ as the reference category, women with PCOS had a higher risk of developing NAFLD regardless of BMI category. The risk was incrementally higher in each BMI category respectively. Among the women with PCOS, risk factor for NAFLD include not only increasing age, higher BMI and dysglycaemia, but also anovulation and clinical features of androgen excess namely alopecia and hirsutism but not acne. Among the women with PCOS, risk factors for OSA include increasing age, higher BMI, hypothyroidism, anovulation, hirsutism, and metformin, which is a possible marker of severity of PCOS. Among the women with PCOS, risk for T2DM and

dysglycaemia was increased by increasing age, higher BMI, hypertension, anovulation, non-Caucasian ethnicities, anovulation and hirsutism and reduced by prescription with combined oral contraceptives and lipid lowering drugs. Like clinical androgen excess, women with biochemical androgen excess (high testosterone, low SHBG) had an increased rate of NAFLD and dysglycaemia with a gradient. Overall, androgen excess emerges as an overarching risk factor for NAFLD, OSA and dysglycaemia among women with PCOS independent of their BMI.

We recommend adopting a lower threshold for screening for NAFLD, OSA and T2DM among women with PCOS irrespective of their BMI.

6.8. References of Chapter 6 : General Discussion

- Allen, L. A., Shrikrishnapalasuriyar, N., & Rees, D. A. (2022). Long-term health outcomes in young women with polycystic ovary syndrome: A narrative review. In *Clinical Endocrinology* (Vol. 97, Issue 2). <https://doi.org/10.1111/cen.14609>
- Amiri, M., Fallahzadeh, A., Sheidaei, A., Mahboobifard, F., & Ramezani Tehrani, F. (2022). Prevalence of idiopathic hirsutism: A systematic review and meta-analysis. In *Journal of Cosmetic Dermatology* (Vol. 21, Issue 4). <https://doi.org/10.1111/jocd.14313>
- Anagnostis, P., Paparodis, R. D., Bosdou, J. K., Bothou, C., Macut, D., Goulis, D. G., & Livadas, S. (2021). Risk of type 2 diabetes mellitus in polycystic ovary syndrome is associated with obesity: a meta-analysis of observational studies. *Endocrine*, 74(2), 245–253. <https://doi.org/10.1007/S12020-021-02801-2>
- Barber, T. M., & Franks, S. (2021). Obesity and polycystic ovary syndrome. In *Clinical Endocrinology* (Vol. 95, Issue 4, pp. 531–541). John Wiley and Sons Inc. <https://doi.org/10.1111/cen.14421>
- Barber, T. M., Hanson, P., Weickert, M. O., & Franks, S. (2019). Obesity and Polycystic Ovary Syndrome: Implications for Pathogenesis and Novel Management Strategies. *Clinical Medicine Insights: Reproductive Health*, 13, 117955811987404. <https://doi.org/10.1177/1179558119874042>
- Bauer, J., Müller, R., Brüggmann, D., & Groneberg, D. A. (2018). Spatial Accessibility of Primary Care in England: A Cross-Sectional Study Using a Floating Catchment Area Method. *Health Services Research*, 53(3). <https://doi.org/10.1111/1475-6773.12731>
- Bazick, J., Donithan, M., Neuschwander-Tetri, B. A., Kleiner, D., Brunt, E. M., Wilson, L., Doo, E., Lavine, J., Tonascia, J., & Loomba, R. (2015). Clinical model for NASH and

advanced fibrosis in adult patients with diabetes and NAFLD: Guidelines for referral in NAFLD. *Diabetes Care*, 38(7). <https://doi.org/10.2337/dc14-1239>

Blane, D. N., & Ciftci, Y. (2022). Editorials Improving GP registration and access for migrant health. In *British Journal of General Practice* (Vol. 72, Issue 715). <https://doi.org/10.3399/bjgp22X718301>

Bozdag, G., Mumusoglu, S., Zengin, D., Karabulut, E., & Yildiz, B. O. (2016). The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction (Oxford, England)*, 31(12), 2841–2855. <https://doi.org/10.1093/HUMREP/DEW218>

Cai, B., Xu, W., Bortnichak, E., & Watson, D. J. (2012). An algorithm to identify medical practices common to both the General Practice Research Database and The Health Improvement Network database. *Pharmacoepidemiology and Drug Safety*, 21(7), 770–774. <https://doi.org/10.1002/PDS.3277>

Carmina, E., Dreno, B., Lucky, W. A., Agak, W. G., Dokras, A., Kim, J. J., Lobo, R. A., Ramezani Tehrani, F., & Dumesic, D. (2022). Female Adult Acne and Androgen Excess: A Report From the Multidisciplinary Androgen Excess and PCOS Committee. *Journal of the Endocrine Society*, 6(3). <https://doi.org/10.1210/jendso/bvac003>

Cho, J., Lee, I., Park, D. H., Kwak, H. B., & Min, K. (2021). Relationships between socioeconomic status, handgrip strength, and non-alcoholic fatty liver disease in middle-aged adults. *International Journal of Environmental Research and Public Health*, 18(4). <https://doi.org/10.3390/ijerph18041892>

Cusi, K., Isaacs, S., Barb, D., Basu, R., Caprio, S., Garvey, W. T., Kashyap, S., Mechanick, J. I., Mouzaki, M., Nadolsky, K., Rinella, M. E., Vos, M. B., & Younossi, Z. (2022). American Association of Clinical Endocrinology Clinical Practice Guideline for the Diagnosis and

Management of Nonalcoholic Fatty Liver Disease in Primary Care and Endocrinology Clinical Settings: Co-Sponsored by the American Association for the Study of Liver Diseases (AASLD). *Endocrine Practice*, 28(5), 528–562.

<https://doi.org/10.1016/J.EPRAC.2022.03.010>

Das, A., Lin, T. A., Lin, C., Meirson, T., McCaw, Z. R., Tian, L., & Ludmir, E. B. (2023). Assessment of Median and Mean Survival Time in Cancer Clinical Trials. *JAMA Network Open*, 6(4), e236498. <https://doi.org/10.1001/jamanetworkopen.2023.6498>

Davey, A., Carter, M., & Campbell, J. L. (2013). Priorities for young adults when accessing UK primary care: literature review. In *Primary health care research & development* (Vol. 14, Issue 4). <https://doi.org/10.1017/S1463423612000497>

Delanerolle, G., Ayis, S., Barzilova, V., Phiri, P., Zeng, Y., Ranaweera, S., Shetty, A., Haque, N., Kar, D., Majumder, K., Rathod, S., Raymont, V., Shi, J. Q., & Hapangama, D. K. (2022). A Systematic Review and Meta-analysis of Polycystic Ovary Syndrome and Mental Health among Black Asian Minority Ethnic populations. *MedRxiv*.

Derks, I. P. M., Koster, A., Schram, M. T., Stehouwer, C. D. A., Dagnelie, P. C., Groffen, D. A. I., & Bosma, H. (2017). The association of early life socioeconomic conditions with prediabetes and type 2 diabetes: results from the Maastricht study. *International Journal for Equity in Health*, 16(1). <https://doi.org/10.1186/s12939-017-0553-7>

Ding, T., Baio, G., Hardiman, P. J., Petersen, I., & Sammon, C. (2016). Diagnosis and management of polycystic ovary syndrome in the UK: a retrospective cohort study. *BMJ Open*, 6, 5–8. <https://doi.org/10.1136/bmjopen-2016-012461>

Dixon, S., McNiven, A., Connolly, A., & Hinton, L. (2021). Women's health and primary care: Time to get it right for the life course. In *British Journal of General Practice* (Vol. 71, Issue 713). <https://doi.org/10.3399/bjgp21X717713>

Doycheva, I., & Ehrmann, D. A. (2022). Nonalcoholic fatty liver disease and obstructive sleep apnea in women with polycystic ovary syndrome. In *Fertility and Sterility* (Vol. 117, Issue 5, pp. 897–911). Elsevier Inc. <https://doi.org/10.1016/j.fertnstert.2022.03.020>

Doycheva, I., Watt, K. D., & Alkhouri, N. (2017). Nonalcoholic fatty liver disease in adolescents and young adults: The next frontier in the epidemic. *Hepatology*. <https://doi.org/10.1002/hep.29068>

Etindele Sosso, F., & Matos, E. (2021). Socioeconomic disparities in obstructive sleep apnea: a systematic review of empirical research. In *Sleep and Breathing* (Vol. 25, Issue 4). <https://doi.org/10.1007/s11325-020-02274-z>

Floras, J. S. (2015). Hypertension and Sleep Apnea. *The Canadian Journal of Cardiology*, 31(7), 889–897. <https://doi.org/10.1016/J.CJCA.2015.05.003>

Fogel, R. B., Malhotra, A., Pillar, G., Pittman, S. D., Dunaif, A., & White, D. P. (2001). Increased Prevalence of Obstructive Sleep Apnea Syndrome in Obese Women with Polycystic Ovary Syndrome 1. *The Journal of Clinical Endocrinology & Metabolism*, 86(3), 1175–1180. <https://doi.org/10.1210/jcem.86.3.7316>

Forslund, M., Landin-Wilhelmsen, K., Trimpou, P., Schmidt, J., Brännström, M., & Dahlgren, E. (2020). Type 2 diabetes mellitus in women with polycystic ovary syndrome during a 24-year period: importance of obesity and abdominal fat distribution. *Human Reproduction Open*, 2020(1). <https://doi.org/10.1093/hropen/hoz042>

Gabbay, I. E., & Lavie, P. (2012). Age- and gender-related characteristics of obstructive sleep apnea. *Sleep and Breathing*, 16(2). <https://doi.org/10.1007/s11325-011-0523-z>

Glintborg, D., & Andersen, M. (2020). Medical treatment and comorbidity in polycystic ovary syndrome: An updated review. In *Current Opinion in Endocrine and Metabolic Research* (Vol. 12, pp. 33–40). Elsevier Ltd. <https://doi.org/10.1016/j.coemr.2020.02.014>

- Grosios, K., Gahan, P. B., & Burbidge, J. (2010). Overview of healthcare in the UK. In *EPMA Journal* (Vol. 1, Issue 4). <https://doi.org/10.1007/s13167-010-0050-1>
- Gunn, H. M. (2022). Proteomic and clinical insights into polycystic ovary syndrome in adolescents.
- Hanna, F., Wu, P., Heald, A., & Fryer, A. (2023). Diabetes detection in women with gestational diabetes and polycystic ovarian syndrome. *BMJ*, 382, e071675. <https://doi.org/10.1136/BMJ-2022-071675>
- Hazlehurst, J. M., Woods, C., Marjot, T., Cobbold, J. F., & Tomlinson, J. W. (2016). Non-alcoholic fatty liver disease and diabetes. *Metabolism: Clinical and Experimental*, 65(8), 1096–1108. <https://doi.org/10.1016/j.metabol.2016.01.001>
- Hillman, S. C., Bryce, C., Caleyachetty, R., & Dale, J. (2020). Women’s experiences of diagnosis and management of polycystic ovary syndrome: A mixed-methods study in general practice. *Cancer Prevention Research*, 13(5). <https://doi.org/10.3399/bjgp20X708881>
- Hippisley-Cox, J., & Coupland, C. (2017). Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. *BMJ (Clinical Research Ed.)*, 359. <https://doi.org/10.1136/bmj.j5019>
- Hirsch, S., & Gaultney, J. (2023). Sleep disturbances in individuals with down syndrome: An overview. *Journal of Intellectual Disabilities : JOID*. <https://doi.org/10.1177/17446295231173011>
- Hirst, V., & Cuthill, F. (2021). Benefits of GP care in outreach settings for people experiencing homelessness: A qualitative study. *British Journal of General Practice*, 71(709). <https://doi.org/10.3399/BJGP.2020.0749>
- Hodson, N., Ford, E., & Cooper, M. (2019). Adherence to guidelines on documentation required for registration to London GP practice websites: A mixed-methods cross-sectional

study. *British Journal of General Practice*, 69(687). <https://doi.org/10.3399/bjgp19X705581>

Hu, J., Cai, X., Zhu, Q., Heizhati, M., Wen, W., Hong, J., Dang, Y., Yang, W., Li, N., & Luo, Q. (2023). Relationship Between Plasma Aldosterone Concentrations and Non-Alcoholic Fatty Liver Disease Diagnosis in Patients with Hypertension: A Retrospective Cohort Study. *Diabetes, Metabolic Syndrome and Obesity*, 16, 1625–1636. <https://doi.org/10.2147/DMSO.S408722>

Ignatov, A., & Ortmann, O. (2020). Endocrine Risk Factors of Endometrial Cancer: Polycystic Ovary Syndrome, Oral Contraceptives, Infertility, Tamoxifen. *Cancers*, 12(7), 1–10. <https://doi.org/10.3390/CANCERS12071766>

Islam, H., Masud, J., Islam, Y. N., & Haque, F. K. M. (2022). An update on polycystic ovary syndrome: A review of the current state of knowledge in diagnosis, genetic etiology, and emerging treatment options. In *Women's Health* (Vol. 18). <https://doi.org/10.1177/17455057221117966>

Jaroenlapnopparat, A., Charoenngam, N., Ponvilawan, B., Mariano, M., Thongpiya, J., & Yingchoncharoen, P. (2023). Menopause is associated with increased prevalence of nonalcoholic fatty liver disease: a systematic review and meta-analysis. In *Menopause* (Vol. 30, Issue 3). <https://doi.org/10.1097/GME.0000000000002133>

Jaruvongvanich, V., Sanguankeo, A., Riangwiwat, T., & Upala, S. (2017). Testosterone, sex hormone-binding globulin and nonalcoholic fatty liver disease: A systematic review and meta-analysis. *Annals of Hepatology*, 16(3), 382–394. <https://doi.org/10.5604/16652681.1235481>

Jones, H., Sprung, V. S., Pugh, C. J. A., Daousi, C., Irwin, A., Aziz, N., Adams, V. L., Thomas, E. L., Bell, J. D., Kemp, G. J., & Cuthbertson, D. J. (2012). 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. *Journal of Clinical Endocrinology and Metabolism*, 97(10). <https://doi.org/10.1210/jc.2012-1382>

Jullian-Desayes, I., Trzepizur, W., Boursier, J., Joyeux-Faure, M., Bailly, S., Benmerad, M., Le Vaillant, M., Jaffre, S., Pigeanne, T., Bizieux-Thaminy, A., Humeau, M. P., Alizon, C., Goupil, F., Costentin, C., Gaucher, J., Tamisier, R., Gagnadoux, F., & Pépin, J. L. (2021). Obstructive sleep apnea, chronic obstructive pulmonary disease and NAFLD: an individual participant data meta-analysis. *Sleep Medicine*, 77. <https://doi.org/10.1016/j.sleep.2020.04.004>

Kahal, H., Kyrou, I., Tahrani, A. A., & Randeva, H. S. (2017). Obstructive sleep apnoea and polycystic ovary syndrome: A comprehensive review of clinical interactions and underlying pathophysiology. In *Clinical Endocrinology* (Vol. 87, Issue 4, pp. 313–319). Blackwell Publishing Ltd. <https://doi.org/10.1111/cen.13392>

Kahal, H., Kyrou, I., Uthman, O. A., Brown, A., Johnson, S., Wall, P. D. H., Metcalfe, A., Parr, D. G., Tahrani, A. A., & Randeva, H. S. (2020). The prevalence of obstructive sleep apnoea in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Sleep and Breathing*, 24(1), 339–350. <https://doi.org/10.1007/s11325-019-01835-1>

Kakoly, N. S., Earnest, A., Teede, H. J., Moran, L. J., & Joham, A. E. (2019). The impact of obesity on the incidence of type 2 diabetes among women with polycystic ovary syndrome. *Diabetes Care*, 42(4), 560–567. <https://doi.org/10.2337/dc18-1738>

Kakoly, N. S., Khomami, M. B., Joham, A. E., Cooray, S. D., Misso, M. L., Norman, R. J., Harrison, C. L., Ranasinha, S., Teede, H. J., & Moran, L. J. (2018). Ethnicity, obesity and the prevalence of impaired glucose tolerance and type 2 diabetes in PCOS: a systematic review and meta-regression. *Human Reproduction Update*, 24(4), 455–467. <https://doi.org/10.1093/humupd/dmy007>

Kang, S. H., Lee, H. W., Yoo, J. J., Cho, Y., Kim, S. U., Lee, T. H., Jang, B. K., Kim, S. G., Ahn, S. B., Kim, H., Jun, D. W., Choi, J. Il, Song, D. S., Kim, W., Jeong, S. W., Kim, M. Y., Koh, H., Jeong, S., Lee, J. W., & Cho, Y. K. (2021). KASL clinical practice guidelines: Management of nonalcoholic fatty liver disease. *Clinical and Molecular Hepatology*, 27(3). <https://doi.org/10.3350/cmh.2021.0178>

Kara, O., Arsoy, H. A., & Keskin, M. (2023). Relationship between non-alcoholic fatty liver disease and hyperandrogenemia in adolescent with polycystic ovary syndrome. *Clinical and Experimental Pediatrics*. <https://doi.org/10.3345/CEP.2023.00353>

Katrina Loomis, A., Kabadi, S., Preiss, D., Hyde, C., Bonato, V., Louis, M. S., Desai, J., Gill, J. M. R., Welsh, P., Waterworth, D., & Sattar, N. (2016). Body mass index and risk of nonalcoholic fatty liver disease: Two electronic health record prospective studies. *Journal of Clinical Endocrinology and Metabolism*, 101(3). <https://doi.org/10.1210/jc.2015-3444>

Kazemi Jaliseh, H., Ramezani Tehrani, F., Behboudi-Gandevani, S., Hosseinpanah, F., Khalili, D., Cheraghi, L., & Azizi, F. (2017). Polycystic ovary syndrome is a risk factor for diabetes and prediabetes in middle-aged but not elderly women: a long-term population-based follow-up study. *Fertility and Sterility*, 108(6), 1078–1084. <https://doi.org/10.1016/j.fertnstert.2017.09.004>

Lavrentaki, A., Thomas, T., Subramanian, A., Valsamakis, G., Thomas, N., Toulis, K. A., Wang, J., Daly, B., Saravanan, P., Sumilo, D., Mastorakos, G., Tahrani, A. A., & Nirantharakumar, K. (2019). Increased risk of non-alcoholic fatty liver disease in women with gestational diabetes mellitus: A population-based cohort study, systematic review and meta-analysis. *Journal of Diabetes and Its Complications*, 33(10). <https://doi.org/10.1016/J.JDIACOMP.2019.06.006>

Lee, B., Jung, E. A., Yoo, J. J., Kim, S. G., Lee, C. B., Kim, Y. S., Jeong, S. W., Jang, J. Y., Lee, S. H., Kim, H. S., Jun, B. G., Kim, Y. D., & Cheon, G. J. (2020). Prevalence, incidence

and risk factors of tamoxifen-related non-alcoholic fatty liver disease: A systematic review and meta-analysis. *Liver International : Official Journal of the International Association for the Study of the Liver*, 40(6), 1344–1355. <https://doi.org/10.1111/LIV.14434>

Liao, W.-T., Huang, J.-Y., Lee, M.-T., Yang, Y.-C., & Wu, C.-C. (2022). Higher risk of type 2 diabetes in young women with polycystic ovary syndrome: A 10-year retrospective cohort study. *World Journal of Diabetes*, 13(3). <https://doi.org/10.4239/wjd.v13.i3.240>

Liao, X., & Cao, S. (2020). Association of the genetic polymorphisms rs6259 and rs727428 of the SHBG Gene with polycystic ovary syndrome risk: A meta-analysis. *Genetic Testing and Molecular Biomarkers*, 24(8). <https://doi.org/10.1089/gtmb.2019.0229>

Lin, T.-Y., Lin, P.-Y., Su, T.-P., Li, C.-T., Lin, W.-C., Chang, W.-H., Chen, T.-J., Bai, Y.-M., & Chen, M.-H. (2017). Risk of developing obstructive sleep apnea among women with polycystic ovarian syndrome: a nationwide longitudinal follow-up study. *Sleep Medicine*, 36, 165–169. <https://doi.org/10.1016/j.sleep.2016.12.029>

Liu, D., Gao, X., Pan, X. F., Zhou, T., Zhu, C., Li, F., Fan, J. G., Targher, G., & Zhao, J. (2023). The hepato-ovarian axis: genetic evidence for a causal association between non-alcoholic fatty liver disease and polycystic ovary syndrome. *BMC Medicine*, 21(1). <https://doi.org/10.1186/S12916-023-02775-0>

Ludwig, U., Holzner, D., Denzer, C., Greinert, A., Haenle, M. M., Oeztuerk, S., Koenig, W., Boehm, B. O., Mason, R. A., Kratzer, W., & Graeter, T. (2015). Subclinical and clinical hypothyroidism and non-alcoholic fatty liver disease: A cross-sectional study of a random population sample aged 18 to 65 years. *BMC Endocrine Disorders*, 15(1). <https://doi.org/10.1186/s12902-015-0030-5>

Madan, A., Manek, N., & Gregory, S. (2017). General practice: the heart of the NHS. *The British Journal of General Practice*, 67(657), 150. <https://doi.org/10.3399/BJGP17X689965>

Magadi, J. P., & Magadi, M. A. (2022). Ethnic inequalities in patient satisfaction with primary health care in England: Evidence from recent General Practitioner Patient Surveys (GPPS). *PLoS ONE*, 17(12 December). <https://doi.org/10.1371/journal.pone.0270775>

Mantovani, A., Petracca, G., Beatrice, G., Tilg, H., Byrne, C. D., & Targher, G. (2021). Non-alcoholic fatty liver disease and risk of incident diabetes mellitus: an updated meta-analysis of 501 022 adult individuals. *Gut*, 70(5), 962–969. <https://doi.org/10.1136/GUTJNL-2020-322572>

Manzano-Nunez, R., Santana-Dominguez, M., Rivera-Esteban, J., Sabiote, C., Sena, E., Bañares, J., Tacke, F., & Pericàs, J. M. (2023). Non-Alcoholic Fatty Liver Disease in Patients with Polycystic Ovary Syndrome: A Systematic Review, Meta-Analysis, and Meta-Regression. In *Journal of Clinical Medicine* (Vol. 12, Issue 3). MDPI. <https://doi.org/10.3390/jcm12030856>

March, W. A., Moore, V. M., Willson, K. J., Phillips, D. I. W., Norman, R. J., & Davies, M. J. (2010). The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. *Human Reproduction*, 25(2). <https://doi.org/10.1093/humrep/dep399>

McDonnell, T., Cussen, L., McIlroy, M., & O'Reilly, M. W. (2022). Characterizing skeletal muscle dysfunction in women with polycystic ovary syndrome. In *Therapeutic Advances in Endocrinology and Metabolism* (Vol. 13). <https://doi.org/10.1177/20420188221113140>

Morrison, A. E., Zaccardi, F., Khunti, K., & Davies, M. J. (2019). Causality between non-alcoholic fatty liver disease and risk of cardiovascular disease and type 2 diabetes: A meta-analysis with bias analysis. *Liver International*, 39(3). <https://doi.org/10.1111/liv.13994>

Muthiah, M. D., Cheng Han, N., & Sanyal, A. J. (2022). A clinical overview of non-alcoholic fatty liver disease: A guide to diagnosis, the clinical features, and complications—What the non-specialist needs to know. *Diabetes, Obesity and Metabolism*, 24(S2). 335

<https://doi.org/10.1111/dom.14521>

Ng, N. Y. H., Wu, H., Lau, E. S. H., Zhang, X., Yang, A., Tsang, A. Y. T., Yau, T. T. L., Kong, A. P. S., Ng, K., Chung, J. P. W., Chow, E. Y. K., Chan, J. C. N., Cheung, L. P., Luk, A. O. Y., & Ma, R. C. W. (2023). Young-onset diabetes in women with Polycystic Ovary Syndrome: A territory-wide retrospective analysis in Hong Kong. *Diabetes Research and Clinical Practice*, 199. <https://doi.org/10.1016/j.diabres.2023.110640>

NICE. (2017). Type 2 diabetes: prevention in people at high risk. NICE National Institute of Health and Care Excellence, Public health guideline.

O'Reilly, M. W., Taylor, A. E., Crabtree, N. J., Hughes, B. A., Capper, F., Crowley, R. K., Stewart, P. M., Tomlinson, J. W., & Arlt, W. (2014). Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *The Journal of Clinical Endocrinology and Metabolism*, 99(3), 1027–1036. <https://doi.org/10.1210/jc.2013-3399>

Patel-Sanchez, N., Perito, E., Tsai, P., Raymond-Flesch, M., Lodish, M., & Sarkar, M. (2023). Prevalence of nonalcoholic fatty liver disease increased with type 2 diabetes mellitus in overweight/obese youth with polycystic ovary syndrome. *Journal of Pediatric Endocrinology and Metabolism*. <https://doi.org/10.1515/jpem-2022-0527>

Peykari, N., Mehdipour, P., Larijani, B., Sepanlou, S. G., Djalalinia, S., Kasaeian, A., Parsaeian, M., Ahmadvand, A., Khosravi, A., Malekzadeh, R., & Farzadfar, F. (2023). The levels and trends of diabetes prevalence at national and sub-national levels in Iran (1990 - 2016). *Journal of Diabetes and Metabolic Disorders*, 22(1), 743–752. <https://doi.org/10.1007/S40200-023-01197-2>

Potter, L. C., Horwood, J., & Feder, G. (2022). Access to healthcare for street sex workers in the UK: perspectives and best practice guidance from a national cross-sectional survey of frontline workers. *BMC Health Services Research*, 22(1). <https://doi.org/10.1186/s12913->

Ramezani Tehrani, F., Behboudi-Gandevani, S., Bidhendi Yarandi, R., Saei Ghare Naz, M., & Carmina, E. (2021). Prevalence of acne vulgaris among women with polycystic ovary syndrome: a systemic review and meta-analysis. *Gynecological Endocrinology*, 37(5). <https://doi.org/10.1080/09513590.2020.1859474>

Reilly, J., Hassanally, K., Budd, J., & Mercer, S. (2020). Accident and emergency department attendance rates of people experiencing homelessness by GP registration: a retrospective analysis. *BJGP Open*, 4(5). <https://doi.org/10.3399/bjgpopen20X101089>

Rezayat, A. A., Moghadam, M. D., Nour, M. G., Shirazinia, M., Ghodsi, H., Zahmatkesh, M. R. R., Noghabi, M. T., Hoseini, B., & Rezayat, K. A. (2018). Association between smoking and non-alcoholic fatty liver disease: A systematic review and meta-analysis. *SAGE Open Medicine*, 6. <https://doi.org/10.1177/2050312117745223>

Rocha, A. L. L., Faria, L. C., Guimarães, T. C. M., Moreira, G. V., Cândido, A. L., Couto, C. A., & Reis, F. M. (2017). Non-alcoholic fatty liver disease in women with polycystic ovary syndrome: systematic review and meta-analysis. *Journal of Endocrinological Investigation*, 40(12), 1279–1288. <https://doi.org/10.1007/s40618-017-0708-9>

Rong, F., Dai, H., Wu, Y., Li, J., Liu, G., Chen, H., & Zhang, X. (2021). Association between thyroid dysfunction and type 2 diabetes: a meta-analysis of prospective observational studies. *BMC Medicine*, 19(1). <https://doi.org/10.1186/s12916-021-02121-2>

Roy, S., Abudu, A., Salinas, I., Sinha, N., Cline-Fedewa, H., Yaw, A. M., Qi, W., Lydic, T. A., Takahashi, D. L., Hennebold, J. D., Hoffmann, H. M., Wang, J., & Sen, A. (2022). Androgen-mediated Perturbation of the Hepatic Circadian System Through Epigenetic Modulation Promotes NAFLD in PCOS Mice. *Endocrinology*, 163(10). <https://doi.org/10.1210/endocr/bqac127>

Sam, S., & Ehrmann, D. A. (2019). Pathogenesis and Consequences of Disordered Sleep in

PCOS. *Clinical Medicine Insights: Reproductive Health*, 13.

<https://doi.org/10.1177/1179558119871269>

Sarkar, M., Terrault, N., Chan, W., Cedars, M. I., Huddleston, H. G., Duwaerts, C. C., Balitzer, D., & Gill, R. M. (2020). Polycystic ovary syndrome (PCOS) is associated with NASH severity and advanced fibrosis. *Liver International*, 40(2). <https://doi.org/10.1111/liv.14279>

Sarkar, M., Wellons, M., Cedars, M. I., VanWagner, L., Gunderson, E. P., Ajmera, V., Torchen, L., Siscovick, D., Carr, J. J., Terry, J. G., Rinella, M., Lewis, C. E., & Terrault, N. (2017). Testosterone Levels in Pre-Menopausal Women are Associated With Nonalcoholic Fatty Liver Disease in Midlife. *American Journal of Gastroenterology*, 112(5). <https://doi.org/10.1038/ajg.2017.44>

Schwimmer, J. B., Khorram, O., Chiu, V., & Schwimmer, W. B. (2005). Abnormal aminotransferase activity in women with polycystic ovary syndrome. *Fertility and Sterility*, 83(2). <https://doi.org/10.1016/j.fertnstert.2004.08.020>

Setia, M. (2016). Methodology Series Module 2: Case-control Studies. *Indian Journal of Dermatology*, 61(2), 146. <https://doi.org/10.4103/0019-5154.177773>

Shahbaz, M., Almatooq, H., Foucambert, P., Esbrand, F. D., Zafar, S., Panthangi, V., Kurupp, A. R. C., Raju, A., Luthra, G., & Khan, S. (2022). A Systematic Review of the Risk of Non-alcoholic Fatty Liver Disease in Women With Polycystic Ovary Syndrome. <https://doi.org/10.7759/cureus.29928>

Shaib, F., & Attarian, H. (2023). Sex and gender differences in sleep disorders: An overview. *Principles of Gender-Specific Medicine*, 661–679. <https://doi.org/10.1016/B978-0-323-88534-8.00036-5>

Sharma, A., Walker, D. M., & Atiomo, W. (2010). National survey on management of weight

reduction in PCOS women in the United Kingdom. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 152(2). <https://doi.org/10.1016/j.ejogrb.2010.07.015>

Shengir, M., Chen, T., Guadagno, E., Ramanakumar, A. V., Ghali, P., Deschenes, M., Wong, P., Krishnamurthy, S., & Sebastiani, G. (2021). Non-alcoholic fatty liver disease in premenopausal women with polycystic ovary syndrome: A systematic review and meta-analysis. *JGH Open*, 5(4). <https://doi.org/10.1002/jgh3.12512>

Simon, S. L., McWhirter, L., Diniz Behn, C., Bubar, K. M., Kaar, J. L., Pyle, L., Rahat, H., Garcia-Reyes, Y., Carreau, A. M., Wright, K. P., Nadeau, K. J., & Cree-Green, M. (2019). Morning Circadian Misalignment Is Associated With Insulin Resistance in Girls With Obesity and Polycystic Ovarian Syndrome. *The Journal of Clinical Endocrinology and Metabolism*, 104(8), 3525–3534. <https://doi.org/10.1210/JC.2018-02385>

Singh, A., Gainder, S., Banerjee, P., Goel, A., Kumar, P., Mondal, B., Banik, S. P., & Bagchi, D. (2022). Efficacy of a Proprietary Fenugreek Seed Extract (*Trigonella foenum-graecum*, Furocyst®) in Women with Polycystic Ovary Syndrome (PCOS): a Randomized, Double-Blind, Placebo-Controlled Study. *Journal of the American Nutrition Association*. <https://doi.org/10.1080/27697061.2022.2126410>

Subramanian, A., Adderley, N. J., Tracy, A., Taverner, T., Hanif, W., Toulis, K. A., Neil Thomas, G., Tahrani, A. A., & Nirantharakumar, K. (2019). Risk of incident obstructive sleep apnea among patients with type 2 diabetes. *Diabetes Care*, 42(5). <https://doi.org/10.2337/dc18-2004>

Subramanian, A., Anand, A., Adderley, N. J., Okoth, K., Toulis, K. A., Gokhale, K., Sainsbury, C., O'Reilly, M. W., Arlt, W., & Nirantharakumar, K. (2021). Increased COVID-19 infections in women with polycystic ovary syndrome: A population-based study. *European Journal of Endocrinology*, 184(5), 637–645. <https://doi.org/10.1530/EJE-20-1163>

Subramanian, A., Lee, S. I., Phillips, K., Toulis, K. A., Kempegowda, P., O'reilly, M. W., Adderley, N. J., Thangaratinam, S., Arlt, W., & Nirantharakumar, K. (2022). Polycystic ovary syndrome and risk of adverse obstetric outcomes: a retrospective population-based matched cohort study in England. *BMC Medicine*, 20(1). <https://doi.org/10.1186/s12916-022-02473-3>

Sun, D., Zhou, T., Heianza, Y., Li, X., Fan, M., Fonseca, V. A., & Qi, L. (2019). Type 2 Diabetes and Hypertension: A Study on Bidirectional Causality. *Circulation Research*, 124(6), 930–937. <https://doi.org/10.1161/CIRCRESAHA.118.314487>

Talari, K., & Goyal, M. (2020). Retrospective studies - Utility and caveats. *Journal of the Royal College of Physicians of Edinburgh*, 50(4). <https://doi.org/10.4997/JRCPE.2020.409>

Targher, G., Marra, F., & Marchesini, G. (2008). Increased risk of cardiovascular disease in non-alcoholic fatty liver disease: Causal effect or epiphenomenon? In *Diabetologia* (Vol. 51, Issue 11). <https://doi.org/10.1007/s00125-008-1135-4>

Tate, A. R., Dungey, S., Glew, S., Beloff, N., Williams, R., & Williams, T. (2017). Quality of recording of diabetes in the UK: How does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database. *BMJ Open*, 7(1). <https://doi.org/10.1136/bmjopen-2016-012905>

Teede, H., Boivin, J., Thondan, M., & Thien Tay, C. (2023). A systematic review of models of care for polycystic ovary syndrome.

Teede, H. J., Misso, M. L., Costello, M. F., Dokras, A., Laven, J., Moran, L., Piltonen, T., Norman, R. J., & International PCOS Network. (2018). Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Human Reproduction (Oxford, England)*, 33(9), 1602–1618. <https://doi.org/10.1093/humrep/dey256>

Thavaraputta, S., Dennis, J. A., Laoveeravat, P., Nugent, K., & Rivas, A. M. (2019).

Hypothyroidism and Its Association with Sleep Apnea among Adults in the United States: NHANES 2007-2008. *Journal of Clinical Endocrinology and Metabolism*, 104(11). <https://doi.org/10.1210/jc.2019-01132>

Todd, A., Copeland, A., Husband, A., Kasim, A., & Bambra, C. (2015). Access all areas? An area-level analysis of accessibility to general practice and community pharmacy services in England by urbanity and social deprivation. *BMJ Open*, 5(5). <https://doi.org/10.1136/bmjopen-2014-007328>

Turner, G. M. (2016). Missed opportunities for primary prevention of stroke and transient ischaemic attack (TIA) and residual impairments after TIA.

Vgontzas, A. N., Legro, R. S., Bixler, E. O., Grayev, A., Kales, A., & Chrousos, G. P. (2001). Polycystic Ovary Syndrome Is Associated with Obstructive Sleep Apnea and Daytime Sleepiness: Role of Insulin Resistance 1. *The Journal of Clinical Endocrinology & Metabolism*, 86(2), 517–520. <https://doi.org/10.1210/jcem.86.2.7185>

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*, 370(9596), 1453–1457. <http://www.ncbi.nlm.nih.gov/pubmed/25046131>

Westfall, E., Jeske, R., & Bader, A. R. (2020). Nonalcoholic Fatty Liver Disease: Common Questions and Answers on Diagnosis and Management. *American Family Physician*, 102(10).

Wijarnpreecha, K., Thongprayoon, C., Panjawanatan, P., & Ungprasert, P. (2017). Insomnia and risk of nonalcoholic fatty liver disease: A systematic review and meta-analysis. *Journal of Postgraduate Medicine*, 63(4), 226–231. https://doi.org/10.4103/JPGM.JPGM_140_17

Wijeyaratne, C. N., Seneviratne, R. d. A., Dahanayake, S., Kumarapeli, V., Palipane, E.,

- Kuruppu, N., Yapa, C., Seneviratne, R. d. A., & Balen, A. H. (2011). Phenotype and metabolic profile of South Asian women with polycystic ovary syndrome (PCOS): results of a large database from a specialist Endocrine Clinic. *Human Reproduction*, 26(1), 202–213. <https://doi.org/10.1093/humrep/deq310>
- Witchel, S. F., Teede, H. J., & Peña, A. S. (2020). Curtailing PCOS. In *Pediatric Research* (Vol. 87, Issue 2, pp. 353–361). Springer Nature. <https://doi.org/10.1038/s41390-019-0615-1>
- Won, Y. Bin, Seo, S. K., Yun, B. H., Cho, S. H., Choi, Y. S., & Lee, B. S. (2021). Non-alcoholic fatty liver disease in polycystic ovary syndrome women. *Scientific Reports 2021* 11:1, 11(1), 1–11. <https://doi.org/10.1038/s41598-021-86697-y>
- Wu, J., Yao, X. Y., Shi, R. X., Liu, S. F., & Wang, X. Y. (2018). A potential link between polycystic ovary syndrome and non-alcoholic fatty liver disease: an update meta-analysis. *Reproductive Health*, 15(1). <https://doi.org/10.1186/S12978-018-0519-2>
- Yang, J., Zhang, K., Xi, Z., Ma, Y., Shao, C., Wang, W., & Tang, Y. Da. (2022). Short sleep duration and the risk of nonalcoholic fatty liver disease/metabolic associated fatty liver disease: a systematic review and meta-analysis. *Sleep and Breathing*. <https://doi.org/10.1007/s11325-022-02767-z>
- Yang, W., Lu, J., Weng, J., Jia, W., Ji, L., Xiao, J., Shan, Z., Liu, J., Tian, H., Ji, Q., Zhu, D., Ge, J., Lin, L., Chen, L., Guo, X., Zhao, Z., Li, Q., Zhou, Z., Shan, G., & He, J. (2010). Prevalence of Diabetes among Men and Women in China. *New England Journal of Medicine*, 362(12). <https://doi.org/10.1056/nejmoa0908292>
- Yuan, S., Xue, H. L., Yu, H. J., Huang, Y., Tang, B. W., Yang, X. H., Li, Q. X., & He, Q. Q. (2019). Cigarette smoking as a risk factor for type 2 diabetes in women compared with men: A systematic review and meta-analysis of prospective cohort studies. In *Journal of Public Health (United Kingdom)* (Vol. 41, Issue 2). <https://doi.org/10.1093/pubmed/fdy106>

- Zaman, C. F., Sultana, J., Dey, P., Dutta, J., Mustarin, S., Tamanna, N., Roy, A., Bhowmick, N., Khanam, M., Sultana, S., Chowdhury, S., Khanam, F., Sakibuzzaman, M., & Dutta, P. (2022). A Multidisciplinary Approach and Current Perspective of Nonalcoholic Fatty Liver Disease: A Systematic Review. *Cureus*. <https://doi.org/10.7759/cureus.29657>
- Zeng, X., Ren, Y., Wu, K., Yang, Q., Zhang, S., Wang, D., Luo, Y., & Zhang, N. (2023). Association Between Smoking Behavior and Obstructive Sleep Apnea: A Systematic Review and Meta-Analysis. *Nicotine & Tobacco Research : Official Journal of the Society for Research on Nicotine and Tobacco*, 25(3), 364–371. <https://doi.org/10.1093/NTR/NTAC126>
- Zhang, J., Ye, J., Tao, X., Lu, W., Chen, X., & Liu, C. (2022). Sleep disturbances, sleep quality, and cardiovascular risk factors in women with polycystic ovary syndrome: Systematic review and meta-analysis. *Frontiers in Endocrinology*, 13, 971604. <https://doi.org/10.3389/FENDO.2022.971604/BIBTEX>
- Zhang, M., Zhang, W., Tan, J., Zhao, M., Zhang, Q., & Lei, P. (2016). Role of hypothyroidism in obstructive sleep apnea: A meta-analysis. *Current Medical Research and Opinion*, 32(6). <https://doi.org/10.1185/03007995.2016.1157461>