



HAL
open science

Development and validation of a model to identify polycystic ovary syndrome in the French national administrative health database

Eugénie Micolon, Sandrine Loubiere, Appoline Zimmermann, Julie Berbis, Pascal Auquier, Blandine Courbiere

► To cite this version:

Eugénie Micolon, Sandrine Loubiere, Appoline Zimmermann, Julie Berbis, Pascal Auquier, et al.. Development and validation of a model to identify polycystic ovary syndrome in the French national administrative health database. *BMC Medical Research Methodology*, 2025, 25 (1), pp.5. 10.1186/s12874-024-02447-4 . hal-04919311

HAL Id: hal-04919311

<https://amu.hal.science/hal-04919311v1>

Submitted on 3 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

RESEARCH

Open Access



Development and validation of a model to identify polycystic ovary syndrome in the French national administrative health database

Eugénie Micolon^{1*}, Sandrine Loubiere^{2,3}, Appoline Zimmermann¹, Julie Berbis^{2,3}, Pascal Auquier^{2,3} and Blandine Courbiere^{1,4}

Abstract

Background We aimed to develop and validate an algorithm for identifying women with polycystic ovary syndrome (PCOS) in the French national health data system.

Methods Using data from the French national health data system, we applied the International Classification of Diseases (ICD-10) related diagnoses E28.2 for PCOS among women aged 18 to 43 years in 2021. Then, we developed an algorithm to identify PCOS using combinations of clinical criteria related to specific drugs claims, biological exams, international classification of Diseases (ICD-10) related diagnoses during hospitalization, and/or registration for long-term conditions. The sensitivity, specificity and positive predictive value (PPV) of different combinations of algorithm criteria were estimated by reviewing the medical records of the Department of Reproductive Medicine at a university hospital for the year 2022, comparing potential women identified as experiencing PCOS by the algorithms with a list of clinically registered women with or without PCOS.

Results We identified 2,807 (0.01%) women aged 18 to 43 who received PCOS-related care in 2021 using the ICD-10 code for PCOS in the French National health database. By applying the PCOS algorithm to 349 women, the positive and negative predictive values were 0.90 (95%CI 83–95) and 0.93 (95%CI 0.90–0.96) respectively. The sensitivity of the PCOS algorithm was estimated at 0.85 (95%CI 0.77–0.91) and the specificity at 0.96 (95%CI 0.92–0.98).

Conclusion The validity of the PCOS diagnostic algorithm in women undergoing reproductive health care was acceptable. Our findings may be useful for future studies on PCOS using administrative data on a national scale, or even on an international scale given the similarity of coding in this field.

Keywords Polycystic ovary syndrome, Case-finding algorithm, Validation study, National administrative health data bases

*Correspondence:

Eugénie Micolon
e.micolon@hotmail.fr

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Polycystic ovary syndrome (PCOS) is a common endocrine disorder affecting women throughout their lifetime. Four sets of diagnostic criteria for PCOS have been proposed over the last 30 years: First in 1990, the National Institute Health (NIH) [1] suggesting a combination of ovulatory dysfunction and androgen excess. Then in 2003, the European Society for Human Reproduction and Embryology (ESHRE) and American Society for Reproductive Medicine (ASRM) proposed new diagnostic criteria (Rotterdam 2003), based on the presence of at least two criteria among hyperandrogenism, ovulatory dysfunction and polycystic ovary morphology [2]. In 2006, the Androgen Excess Society (AES) proposed a definition of PCOS making hyperandrogenism a mandatory criterion among ovulatory dysfunction and/or PCOM [3]. Excluding other diagnoses such as thyroid disorders, hyperprolactinemia, androgen-secreting tumors, Cushing syndrome and congenital adrenal hyperplasia is necessary before formally establishing the diagnosis of PCOS.

PCOS is associated with significant morbidity during reproductive age: thromboembolic diseases, infertility, increased obstetric morbidity (gestational diabetes, gestational hypertension, pre-eclampsia [4]), anxiety, depression [5], and cardiovascular morbidity [6]). Women with PCOS are also at risk of non-alcoholic fatty liver disease, obesity, metabolic syndrome, type 2 diabetes, hypertension and dyslipidemia [7–11].

The diagnostic delay and the morbidity of PCOS are associated with significant medical costs for the health-care system [12]. Ding et al. estimated at 237 million pounds the cost of only two comorbidities, type 2 diabetes and acne, among women with PCOS in the United Kingdom [13]. In the USA, a full economic evaluation, including healthcare costs of all the comorbidities associated with PCOS reached \$8 billion USD annually in 2020 [14]. Their estimate could be considered conservative because they used the lower PCOS prevalence based on the stricter NIH diagnostic criteria.

Indeed, the systematic review and meta-analysis by Chiaffarino and *al.* reported a PCOS prevalence of 6.2% in Europe and the USA according to NIH criteria, 19.5% according to ESHRE/ASRM criteria and 15% according to AES guidelines [15]. Within this meta-analysis, three US studies examined the prevalence of PCOS, using ICD 9 codes of PCOS (E256.4) from health claim databases between 2008 and 2013 [16–18]. Those systematic approaches to identify women with PCOS in health claim databases only identified a minority of PCOS patients (1.6%). Also, in 2016, the global burden of disease study estimated the prevalence of PCOS in Europe at only 276.4 cases per 100 000 women, which is much lower

than the 6 to 20% previously described [19]. To improve detection of PCOS patients in national health claim databases, we need to dispose of case-finding algorithm for finding PCOS care consuming women.

We launched the PCOS-ECO study in our research department to address the lack of real-world data about PCOS, by identifying women with PCOS between 2013 and 2022 in the French national health data system, and to allow an economic burden study of this population.

The objective of this present study was to conduct and discuss the first stage of the PCOS-ECO study: First to identify the prevalence for the year 2021 of women with PCOS using ICD10 code of PCOS (E28.2) in the French National Health Data system, then to develop a case-finding algorithm for women with PCOS, and to validate it through a retrospective cohort validation study.

Methods

Preliminary study: identifying women with PCOS using ICD-10 code E28.2 in the French National Health Data System

The French National Health Data System “Système National des données de Santé” (SNDS) is a national database that contains a wide range of health-related information. SNDS database covers nearly the entire French population (98.8%), which accounted for more than 66 million people in 2021 [20]. The composition of the dataset integrates nationwide health insurance data associated with outpatient care (Système national d’Information Inter-régime de l’Assurance Maladie (SNIIRAM), hospitalisations (Programme de Médicalisation des Systèmes d’Informations (PMSI), mortality records (Centre d’Epidémiologie des Causes de Décès (CépiDc)) and registration for long-term conditions. As a result of this integration, using a unique pseudonymized identifier, the SNDS database offers a comprehensive overview of all reimbursed medical and paramedical encounters. We conducted a non-interventional retrospective study to identify women with PCOS in SNDS database using ICD-10 code for PCOS E28.2 over the year 2021. All women who met the following criteria were included: (1) aged ≥ 18 and ≤ 43 years-old, (2) affiliated with reimbursement scheme.

Definition of PCOS algorithm

To build a case-finding algorithm to better identifying women with PCOS, we had to create an algorithm grouping together the consumption of medical care linked to PCOS, using ICD-10 codes for diseases, Identification code for drugs (CIP: Code Identifiant de Présentation), French Common Classification of Medical Acts (CCAM) and National Table for biology (TNB) for biological exam. TNB codes only indicate that the tests were carried

out, and does not identify positive test results. Indeed, to avoid overestimating the PCOS population, we had to group together TNB codes with other SNDS codes in algorithms. A complete list of codes listed from criterion (a) to criterion (l) is given in Table 1. Each individual criterion (from criterion (a) to criterion (i)) allows the identification of women with PCOS.

Two criteria are related to the clinical definition of PCOS: the ICD 10 code for PCOS, and a composite criterion of ICD 10 codes related to PCOS (oligomenorrhea and/or hirsutism and/or hyperandrogenism). Two criteria are related to the treatment of infertile women with PCOS: history of ovarian drilling and the use of clomiphene citrate. Letrozole is not authorized for PCOS induction ovulation in France and wasn't included in algorithms. We also did not include myo-inositol or anti-androgen oral contraceptive pills because they are not reimbursed, and therefore not listed in SNDS database.

Women with PCOS representing 72.1% of non-tumoral hirsutism [21], three algorithm criteria are related to hirsutism treatment: the use of cyproterone acetate, described as treatment for severe hirsutism, the composite criterion of a biological blood test of hyperandrogenism, either testosterone, or 17-OH-Progesterone, with the use of spironolactone, prescribed to treat moderate hirsutism, and the prescription of laser treatment, in women with no history of skin grafting or hidradenitis

suppurativa, which are pathologies that allow laser treatment to be reimbursed. Regarding the use of androgen tests in algorithms, we used the ESHRE 2018 guidelines, which recommended prescribing total testosterone and free testosterone first in the diagnosis of PCOS. We did not use androstenedione and dehydroepiandrosterone sulfate (DHEAS), as they are only prescribed second-line if testosterone levels are normal, and therefore would not have been selective. In line with the ESHRE 2018 guidelines, we also used 17-OH-Progesterone as its prescription is recommended for all PCOS women to exclude non-classical congenital hyperplasia.

Knowing that the prevalence of infertility in women with PCOS varies between 70 and 80% [22], and that, depending on body mass index, 75% to 95% of these women have insulin resistance [23], we developed a composite criterion combining a biological search for hyperandrogenism (either testosterone or 17-OH-Progesterone) with an infertility and a biological blood test for insulin resistance using the HOMA-IR ratio.

In the modelling study of Ding et al., 2017 [13], using individual patient data from a UK primary care database, the estimated prevalence of diabetes in the PCOS UK population aged 15 to 69 years-old was 26.5%. With such a high prevalence of type two diabetes in women with PCOS, we assumed that premenopausal women aged 18 to 43 years, treated with metformin or metformin

Table 1 Combination of international Classification of diseases, 10th Revision, CIP codes for drugs, CCAM codes for procedures and TNB codes for biological exam used to identify polycystic ovary syndrome (PCOS) care consuming women in the French National Health Data System "Système National des données de Santé" (SNDS)

Inclusion criteria

- (a) ICD10 code of PCOS (E28.2)
- (b) Composite criteria of ICD 10 code oligomenorrhea (N91.5) and/or hirsutism (L68.0) and/or hyperandrogenism (E28.1)
- (c) CCAM code of ovarian drilling (JJCC003)
- (d) CIP code of clomiphene citrate (3,400,932,623,389)
- (e) CIP code of cyproterone acetate (3,400,932,351,008)
- (f) Composite criteria of TNB code of Testosterone (1136 or 0107) or TNB code of 17-OH-Progesterone (1135) with at least one of these criteria:
 - CIP code of spironolactone (3,400,937,847,469))
 - TNB code for HOMA-IR ratio (7422 + 0552) and Infertility
 - CIP code for Metformin or Metformin embonate (295 or 319,275-0 or 3,400,931,927,501 or 319,276-7 or 3,400,931,927,679)
- (g) CCAM code for Laser (qznp029 or qznp030) without CCAM code of skin graft (Z945) or CIM10 code of hidradenitis suppurativa (L732)
- (h) Composite criteria of Infertility with (CIP code for Metformin or Metformin embonate (295 or 319,275-0 or 3,400,931,927,501 or 319,276-7 or 3,400,931,927,679)) or type 2 diabetes
- (i) CIP code for dydrogesterone (321,929-4 or 340,093,219,294) and CIP code for Metformin or metformin embonate (295 or 319,275-0 or 3,400,931,927,501 or 319,276-7 or 3,400,931,927,679)

Exclusion Criteria

- (j) ICD-10 code for congenital genito-adrenal anomalies linked to enzymatic deficiency (E25.0)
- (k) ICD-10 code for Cushing syndrome (E24.8)
- (l) ICD-10 code for Acromegaly (E22.0)

HOMA-IR ratio Homeostatic Model Assessment of Insulin Resistance

CIP Code Identifiant de Présentation (Identification code for drugs)

TNB Table Nationale de Biologie (French National Table for biology tests)

ICD-10 International Classification of Diseases

embonate (currently used to treat insulin resistance or type two diabetes in women with PCOS) were likely to have PCOS if they were also treated with dydrogesterone (as diagnosis test or as treatment for anovulation), if they had infertility or if they underwent a biological blood test for hyperandrogenism (either testosterone or 17-OH-Progesterone). Based on the same hypothesis, we assumed that infertile women with type 2 diabetes could be targeted as PCOS in the algorithms.

The algorithm included three exclusion criteria: ICD 10 codes for congenital genito-adrenal anomalies linked to enzymatic deficiency, Cushing syndrome and acromegaly.

Internal validation of the PCOS algorithm

Study design

We conducted a retrospective cohort study at our Department of Gynecology, Obstetric and Reproductive Medicine, in a university teaching hospital.

Data source and study population

We extracted the medical consultations of all women aged between 18 to 43 who had undergone a medically assisted reproductive technique within the year 2022. Our age limit was conditioned by the fact that in France, medically assisted reproduction is only authorized and reimbursed up to the age of 43. Among 736 patients, 112 women were identified as having probable PCOS. To estimate the diagnostic performance, we selected a sample of control cases based on the following sample size calculation (PASS23, version 23.0.3): A single-group diagnostic test design was used to obtain two-sided 95% confidence intervals for sensitivity and specificity. The Exact formula (Clopper-Pearson) was used to calculate the limits of the confidence intervals. Sample sensitivity was assumed to be 0.85, sample specificity 0.95, and prevalence 0.152. To produce a sensitivity confidence interval with a width of no more than 0.21, 349 subjects were needed. To produce a specificity confidence interval with a width of no more than 0.21, 32 subjects were needed. The sample size required, so that both confidence intervals have widths of no more than 0.21, was 349, the larger of the two sample sizes. Subsequently, we decided to include all positive PCOS cases in order to consider all criteria potentially used by algorithms, even those with a low frequency of occurrence (such as Ovarian drilling or Cyproterone acetate). A total of 349 charts were therefore used for the validation step: 112 probable PCOS cases and 237 probable non PCOS cases, selected according to the date of their first consultation. Given a prevalence of 0.32, the sample size of 349 produced a confidence interval for sensitivity and specificity with a width no more than 0.14,

for a sample sensitivity and specificity assumed to be 0.85 and 0.95 respectively.

Confirmation of PCOS diagnosed women

To confirm the diagnoses of PCOS in the cohort of women having consulted our reproductive medicine unit in 2021 with or without a diagnosis of PCOS ticked in their medical record, hospital discharge summaries and medical records were reviewed by a gynecologist specialized in reproductive medicine (EM). The diagnosis of PCOS was established according to the ESHRE guideline [2], based on the presence of at least two criteria among hyperandrogenism, ovulatory dysfunction and polycystic ovary morphology. All cases of PCOS had undergone a blood test with TSH and Prolactinemia level, routinely performed in our department, to exclude diagnoses that could mimic PCOS. A second senior physician expert in reproductive endocrinology (AZ) reviewed all medical charts whose assessment was debatable. Final categorization was reached by consensus.

Internal Validation of algorithm criteria

Medical records corresponding to all selected women were reviewed by a physician (EM), aware of the PCOS status of the patients. As a result, PCOS algorithm was considered filled in if any of criteria (c) to (i), as listed in Table 1, and described in "Definition of PCOS algorithm" section were reported. If a patient had a history of acromegaly, cushing syndrome or congenital genito-adrenal anomalies linked to enzymatic deficiency, the PCOS algorithm was considered unfulfilled.

Data analysis of our validation study

Sensitivity was calculated as the proportion of true positive cases among the positive cases from the chart review (PCOS cases correctly identified by the algorithm). Specificity was calculated as the proportion of true negative cases with algorithm among the negative cases from the chart review (non-PCOS cases correctly identified). Positive predictive value was calculated by dividing the number of true positive cases by the the number of positive cases according to the algorithm. Negative predictive value was calculated by dividing the true negative cases by the algorithm's negative cases. The 95% confidence intervals (CIs) were calculated using the exact Clopper-Pearson interval method. The statistical analysis was performed using R version 4.3.1. This study was approved by the local Review Committee (PADS23-178).

Results

Preliminary study

Among the 20,339,672 women aged 18 to 43 registered in the SNDS in 2021, we identified 2,807 (0.01%) women with PCOS using ICD-10 code E28.2.

Internal validation study

After reviewing the medical records of the 112 probable PCOS cases, 107 were classified as PCOS, and 5 as

non PCOS. After reviewing the medical records of the 237 probable women without PCOS, 2 were classified as PCOS and 235 as non PCOS. So the algorithm was applied on 109 PCOS cases and 240 non PCOS cases, as described in the flow diagram, in Fig. 1.

The characteristics of the patients in the validation study are presented in Table 2. The population of women with and without PCOS shared the same age (mean age 35.5/-+ 4.6 vs 35.1/- 5.2, $p=0.460$) and

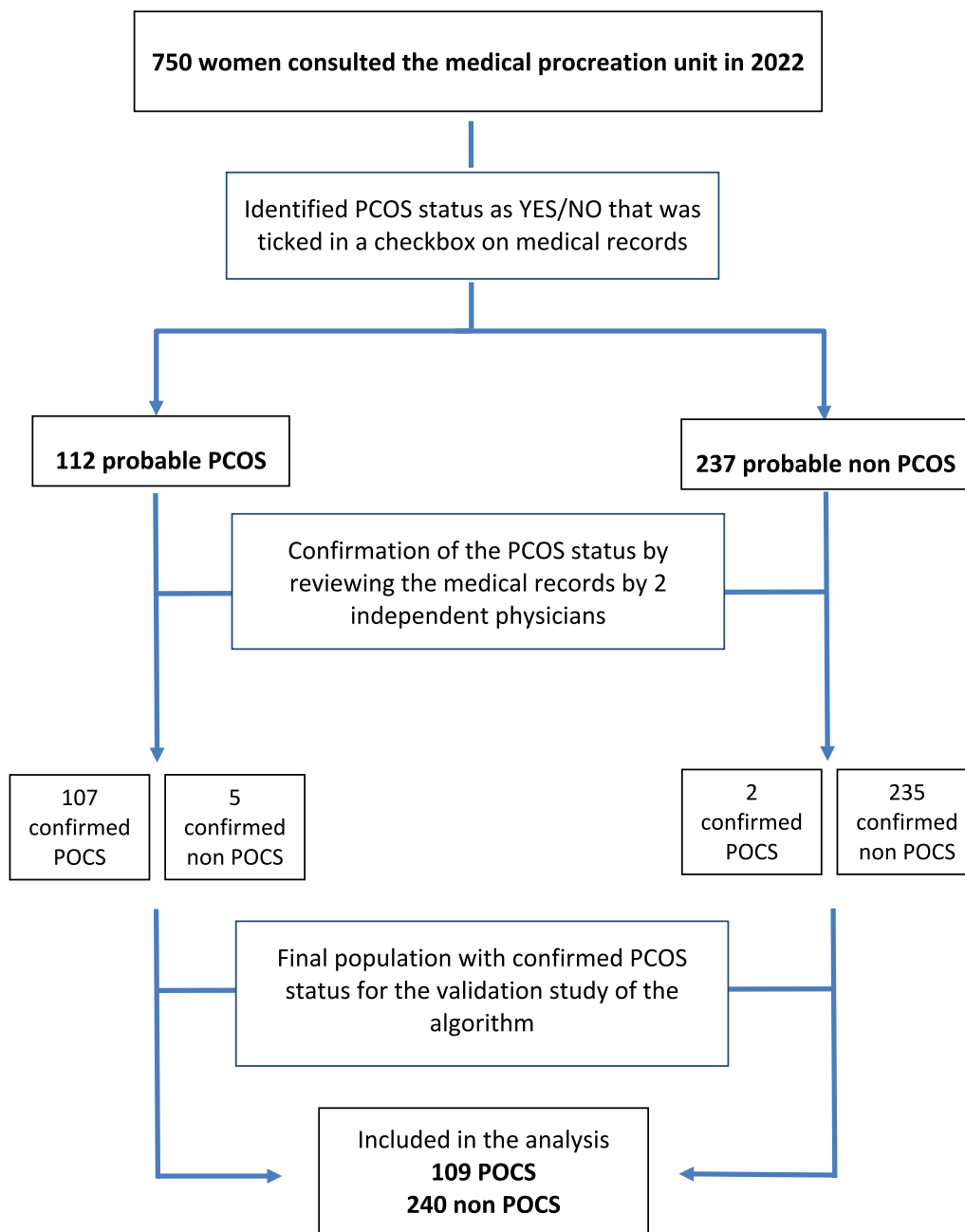


Fig. 1 Flow chart

Table 2 Patient characteristics of reference population

Characteristics	Women with PCOS (n = 109)	Women without PCOS (n = 240)	P value
Age, n (%)			0.136
18–29	10 (9.2)	40 (16.7)	
30–35	40 (36.7)	72 (30.0)	
36–43	59 (54.1)	128 (53.3)	
mean (sd)	35.5 (4.6)	35.1 (5.2)	0.460
BMI, mean (sd)	25.8 (5.3)	25.3 (5.2)	0.241
AMH, ng/ml, mean (sd)	6.7 (4.3)	2.5 (1.6)	<0.001

AMH Anti-müllerian hormone

BMI Body mass index

Table 3 Identification of polycystic ovary syndrome (PCOS) care consuming women: Positive Predictive value (PPV), Negative predictive value (NPV), Sensitivity (Se) and Specificity (Sp) of the PCOS algorithm

	Women with PCOS	Women without PCOS	Total	
Algorithm +	93	10	103	PPV = 0.90% (95%CI 0.83%—0.95%)
Algorithm -	16	230	246	NPV = 0.93% (95%CI 0.90%—0.96%)
Total	109	240	349	
	Se = 0.85% (95%CI 0.77%—0.91%)	Sp = 0.96% (95%CI 0.92%—0.98%)		

PCOS Polycystic ovary syndrome

BMI (25.8 +/- 5.3 kg/m² vs 25.3 +/- 5.2 kg/m² $p = 0.241$). Women with PCOS had higher levels of anti-müllerian hormone (6.7 +/- 4.3 vs. 2.5 +/- 1.6 ng/ml) than women without PCOS. After applying the PCOS algorithm to the 109 confirmed PCOS cases and 240 confirmed non-PCOS cases, the respective positive (PPV) and negative (NPV) predictive values were 0.90 (95%CI 0.83—0.95) and 0.93 (95%CI 0.90—0.96). The sensitivity of the algorithm was estimated at 0.85 (95%CI 0.77 – 0.91) and the specificity at 0.96 (95%CI 0.92 – 0.98) as presented in Table 3.

Regarding the performance of the algorithm criteria as presented in Table 4. It should be noted that a history of ovarian drilling, laser treatment or ciproterone acetate therapy led to the identification of a limited number of women with PCOS ($n = 3$, $n = 1$ and $n = 3$, respectively). Four criteria that combine care consumed by a larger part of our study population, such as Clomiphene citrate, dydrogesterone with metformin or metformin embonate treatment, and our two composite criteria achieved a VPP ranked between 0.91 (95%CI 0.79–0.97) and 0.94 (95%CI 0.86–0.98). The combination of those criteria, presented in Table 5, identified 92 women with PCOS, representing 99% of women with PCOS identified by the algorithm as PCOS, of whom 86% ($n = 80$) were identified by the combination of the two composite criteria.

Regarding our exclusion criteria, one woman was diagnosed with cushing syndrome, and no woman was struggling with congenital genito-adrenal anomalies linked to enzymatic deficiency, skin graft, hidradenitis suppurativa or acromegaly.

Discussion

The preliminary study that identified only 0.01% of women with PCOS aged 18 to 43 using ICD10 code E28.2 underlines the necessity of creating an algorithm to detect women with PCOS in SNDS. We therefore carried out an internal validation study of our algorithm,

Table 4 Performances of the algorithm criteria to identify PCOS care consuming women

Criteria of algorithm	Number of women with PCOS with algorithm	Number of true positive cases	VPP (95%CI)
Ovarian drilling	3	3	1 (0.29–1)
Clomiphene citrate	50	46	0.92 (0.81 – 0.98)
Cyproterone acetate	1	1	1 (0.02–1)
(Testosterone or 17-hydroxyprogesterone) and (Spironolactone) or (HOMA and infertility) or (Metformin or Metformin embonate)	72	68	0.94 (0.86–0.98)
Laser	3	3	1 (0.29–1)
Infertility and (metformin or metformin embonate) or (Type 2 diabetes)	49	45	0.91 (0.79–0.97)
Dydrogesterone and (Metformin or Metformin embonate)	46	42	0.91 (0.79–0.97)

HOMA Homeostatic Model Assessment

Table 5 Combination of highest performing criteria of PCOS algorithm

	PCOS women N(%) n = 109	Non-PCOS women N(%) n = 240	VPP (95%CI)	PCOS women correctly identified by algorithm (%) n = 93
Testosterone or 17-OH-Progesterone with spironolactone And/ Or (HOMA-IR ratio and Infertility) And/or Metformin or Metformin embonate	68	4	0.91 (0.81–0.98)	63
Composite criteria ^a	80	6	0.93 (0.90–0.95)	86
Composite criteria + Clomiphene citrate	92	10	0.90 (0.86–0.93)	99
Composite criteria + Clomifene citrate + Dydrogesterone and (metformin or Metformin embonate)	92	10	0.90 (0.86–0.93)	99

^a Composite criteria: (Testosterone or 17-OH-Progesterone with spironolactone or (HOMA-IR ratio and Infertility) or Metformin or Metformin embonate) + (Infertility with Metformin or Metformin embonate or type 2 diabetes)

which suggests a good reliability on a retrospective cohort of women with an irrefutable diagnosis of PCOS. The characteristics of women with PCOS combines hyperandrogenic, infertile and metabolic phenotypes, the management of which defines the algorithm. Given the heterogeneity of the syndrome, the algorithm combines criteria that are for the most part non-exclusive for PCOS patients, but which combined together specifically target the PCOS population. Our best-performing criterion appears to be the composite criterion combining hyperandrogenic phenotypes (blood test for androgens and spironolactone treatment) with metabolic phenotype (HOMA ratio and metformin or metformin embonate treatment) and infertile phenotype. It thus identifies 68 women, representing 63% of women with PCOS correctly identified by our algorithm. The criterion combining dysovulation treatment (via dydrogesterone) and metabolic treatment (via metformin) does not appear to add value to the criterion targeting infertility with clomiphene citrate, since it does not identify any new PCOS patients, when combined with the other three best-performing criteria of the algorithm, as shown in Table 5. Our algorithm did not identify all women with PCOS, but only those with PCOS complications, depending on their phenotypes, which are often crossed.

We chose to focus on a population of reproductive aged woman because the excess burden of comorbidities associated with PCOS seems to be attenuated or even abolished after the perimenopausal period. Indeed, recent data suggested that type two diabetes, hypertension and cardiovascular diseases were increased only in reproductive aged women with PCOS [24, 25]. The large prospective, population-based cohort study of Ollila and *al* reported that the risk of major adverse cardiovascular events, such as stroke and myocardial infarction, begins as early as age 35 [6]. This highlights the importance of a study population encompassing women with early PCOS through perimenopausal age to assess PCOS comorbidities.

In the French National Health Data System, whereas some diseases can easily be identified using ICD-10 codes, women with PCOS were truly difficult to identify. The small prevalence based on using PCOS ICD10 code is far from the 6 to 19% prevalence described for women with PCOS in the meta-analysis by Chiaffarino and *al*. [15]. The study by Atiomo et al. [26] estimated that the prevalence in Electronic Health Records (EHR) of PCOS using only ICD 9 and ICD 10 codes for PCOS ranks between 0.27% and 1.6%, based on a scoping review of the literature, which is low but nevertheless well above our prevalence in the SNDS. Such a low quotation of ICD codes of PCOS in France can be explained by the fact that ICD codes are only used in France to code the diagnosis of a reason for hospitalization. PCOS being a primary care pathology, patients with PCOS are rarely hospitalized for this indication. Furthermore, in view of this practice, the ICD-10 codes for acne, alopecia and polycystic ovaries did not seem relevant for use in the algorithm to discriminate between patients. In other countries, ICD codes are probably used for medical consultations, which may explain this higher prevalence. While many studies have sought to identify PCOS in EHRs [27], which are clinical databases, with local, regional or even national coverage, no study to our knowledge has identified PCOS in a comprehensive, anonymized national healthcare database, such as the SNDS. Thus, applied to the SNDS database over a 4-year period (2018 to 2021), this algorithm identified 232,225 women with PCOS out of 14,191,879 women aged 18 to 43, representing a PCOS management prevalence of 1.63%. This suggests that our algorithm is effective in identifying women consuming PCOS-related care in SNDS database. Databases comparable to the SNDS exist in most developed countries, and are essential tools for researching and improving healthcare systems. They enable in-depth, large-scale analysis of health data, as well as precise, large-scale medico-economic studies. Nonetheless, as

each country has its own health insurance and social protection system and protocols, some of the criteria we have used could be modified and improved in such context. For example, our approach was limited by the absence of data relating to hospital care and non-reimbursed care, which might be present in health insurance databases of other countries.

The main limitation of our validation study is the retrospective monocentric approach to the evaluation of the algorithm. This study focused on infertile women consulting a French reproductive medicine hospital service and may not reflect the algorithm's performance at national level and among fertile women. In addition, the prevalence of PCOS in our cohort was probably higher than in the general population, and may overestimate the calculation of the algorithm's predictive values when performed in national health data systems. However, this approach enables us to validate the algorithm on a population of women with an irrefutable diagnosis of PCOS.

Our cohort was probably too small to properly evaluate some of the criteria in our algorithm relating to registration of type two diabetes and under-prescribed therapies. Indeed, with only one patient treated with cyproterone acetate, one with spironolactone and three with laser, the criteria of our algorithm relating to hirsutism did not achieve significant results. Given that idiopathic hirsutism accounts for around 7% of women [28], our study validation population was probably too small to assess the proportion of women without PCOS with severe idiopathic hirsutism who could be identified by our algorithm. Furthermore, we were unable to assess the misuse of cyproterone acetate, which could be prescribed off-label by gynecologists to treat acne or metrorrhagia in women without PCOS.

The ideal internal validation study would have involved the use of a primary care database, with electronic medical records registered by general practitioners or gynecologists, but we were not aware, locally, or nationally, of such a database grouping enough women with PCOS with specific data concerning their diagnoses and management. Indeed, we are aware of only one example of French primary care database linked with SNDS: in 2014, Perlberg et al. [29] matched a cohort of 29 000 patients from 30 GP users of a specific software to the SNIIRAM database. This underlines the need, in France, to develop electronic health records based on general medical data. Data matching with the SNDS based on the National Health Identification number (NIR) has developed in recent years, enabling the collection of data on claims, registers and electronic medical records, which considerably enrich the data available in the electronic medical

records reconstituted in the SNDS database [30]. Before conducting a medico-economic study of PCOS in France using this algorithm, an external validation step should be required, and will be carried out by data matching with SNDS.

Conclusion

Overall, we have developed an original algorithm that identify reproductive aged women with PCOS in a French National Health database, and our internal validation study suggests good reliability of our algorithm on a retrospective cohort. Before using this algorithm to assess the economic burden of French women with PCOS, external validation is required, and should be performed by data matching with SNDS. A precise estimate of the economic impact of PCOS would be valuable to enable better prioritization of this condition as a public health interest.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-024-02447-4>.

Supplementary Material 1.

Acknowledgements

We would like to thank Professor Thomas Cuny for his assistance in the editing of the manuscript.

Authors' contributions

Eugénie Micolon: conceptualization, methodology, formal analysis, software, writing – original draft, writing – review & editing. Sandrine Loubiere: conceptualization, methodology, data interpretation, writing – review & editing. Apolline Zimmermann: data interpretation, writing – review & editing. Julie Berbis: conceptualization, methodology, data interpretation, writing – review & editing. Pascal Auquier: conceptualization, writing – review & editing. Blandine Courbiere: conceptualization, investigation, methodology, data interpretation, writing – review & editing.

Funding

This research received funding from the STARTER Project 2022 of AP-HM for PCOS-Eco study. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript and include this statement in the Funding section of your manuscript.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the APHM Institutional Review Board and the Database Use Advisory Committee. All patients were automatically informed of the potential use of their data for retrospective research purposes. In the absence of any opposition, their consent was considered to have been obtained.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Gynecology–Obstetric and Reproductive Medicine, AP-HM, La Conception University teaching Hospital, 147 Boulevard Baille, Marseille 13005, France. ²Department of Clinical Research and Innovation, Support Unit for Clinical Research and Economic Evaluation, Assistance Publique – Hôpitaux de Marseille, Marseille, France. ³Research Unit UR 3279, CEReSS–Health Service Research and Quality of Life Center, Aix-Marseille University, Marseille, France. ⁴IMBE, Aix Marseille Univ, Avignon Univ, CNRS, IRD, Marseille, France.

Received: 9 January 2024 Accepted: 16 December 2024

Published online: 10 January 2025

References

- Lujan ME, Chizen DR, Pierson RA. Diagnostic criteria for polycystic ovary syndrome: pitfalls and controversies. *J Obstet Gynaecol Can.* 2008;30:671–9. [https://doi.org/10.1016/S1701-2163\(16\)32915-2](https://doi.org/10.1016/S1701-2163(16)32915-2).
- Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod.* 2018;33:1602–18. <https://doi.org/10.1093/humrep/dey256>.
- Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, et al. Criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an androgen excess society guideline. *J Clin Endocrinol Metab.* 2006;91:4237–45. <https://doi.org/10.1210/jc.2006-0178>.
- Naver KV, Grinsted J, Larsen SO, Hedley PL, Jørgensen FS, Christiansen M, et al. Increased risk of preterm delivery and pre-eclampsia in women with polycystic ovary syndrome and hyperandrogenaemia. *BJOG.* 2014;121:575–81. <https://doi.org/10.1111/1471-0528.12558>.
- Berni TR, Morgan CL, Berni ER, Rees DA. Polycystic ovary syndrome is associated with adverse mental health and neurodevelopmental outcomes. *J Clin Endocrinol Metab.* 2018;103:2116–25. <https://doi.org/10.1210/jc.2017-02667>.
- Ollila M-M, Arffman RK, Korhonen E, Morin-Papunen L, Franks S, Junttila J, et al. Women with PCOS have an increased risk for cardiovascular disease regardless of diagnostic criteria—a prospective population-based cohort study. *Eur J Endocrinol.* 2023;189:96–105. <https://doi.org/10.1093/ejendo/lvad077>.
- 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.* 2012;18:618–37. <https://doi.org/10.1093/humupd/dms030>.
- Barber TM, Golding SJ, Alvey C, Wass JAH, Karpe F, Franks S, et al. Global adiposity rather than abnormal regional fat distribution characterizes women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2008;93:999–1004. <https://doi.org/10.1210/jc.2007-2117>.
- Kakoly NS, Khomami MB, Joham AE, Cooray SD, Misso ML, Norman RJ, 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;24:455–67. <https://doi.org/10.1093/humupd/dmy007>.
- Sampson M, Kong C, Patel A, Unwin R, Jacobs HS. Ambulatory blood pressure profiles and plasminogen activator inhibitor (PAI-1) activity in lean women with and without the polycystic ovary syndrome. *Clin Endocrinol.* 1996;45:623–9. <https://doi.org/10.1046/j.1365-2265.1996.00863.x>.
- Coronary Heart Disease Risk Factors in Women With Polycystic Ovary Syndrome n.d. <https://doi.org/10.1161/01.ATV.15.7.821>.
- Gibson-Helm M, Teede H, Dunaif A, Dokras A. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. *The Journal of Clinical Endocrinology & Metabolism* 2016;96:2963–2963. <https://doi.org/10.1210/jc.2016-2963>.
- Ding T, Hardiman PJ, Petersen I, Baio G. Incidence and prevalence of diabetes and cost of illness analysis of polycystic ovary syndrome: a Bayesian modelling study. *Hum Reprod.* 2018;33:1299–306. <https://doi.org/10.1093/humrep/dey093>.
- 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 & Metabolism | Oxford Academic n.d. <https://academic.oup.com/jcem/article/107/2/575/6371160?login=false>. Accessed 27 Aug 2023.
- Chiapparino F, Cipriani S, Dalmartello M, Ricci E, Esposito G, Fedele F, et al. Prevalence of polycystic ovary syndrome in European countries and USA: A systematic review and meta-analysis. *European J Obstet Gynecol Reprod Biol.* 2022;279:159–70. <https://doi.org/10.1016/j.ejogrb.2022.10.020>.
- Okoroh EM, Hooper WC, Atrash HK, Yusuf HR, Boulet SL. Prevalence of polycystic ovary syndrome among the privately insured, United States, 2003–2008. *Am J Obstet Gynecol.* 2012;207:299.e1–299.e7. <https://doi.org/10.1016/j.ajog.2012.07.023>.
- Sirmans S, Pate K. Epidemiology, diagnosis, and management of polycystic ovary syndrome. *CLEP.* 2013;1:1. <https://doi.org/10.2147/CLEP.S37559>.
- Lo JC, Feigenbaum SL, Yang J, Pressman AR, Selby JV, Go AS. Epidemiology and adverse cardiovascular risk profile of diagnosed polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2006;91:1357–63. <https://doi.org/10.1210/jc.2005-2430>.
- Miazgowski T, Martopullo I, Widecka J, Miazgowski B, Brodowska A. 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. *Arch Med Sci.* 2021;17:343–51. <https://doi.org/10.5114/aoms.2019.87112>.
- Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Revue d'Épidémiologie et de Santé Publique.* 2017;65:S149–67. <https://doi.org/10.1016/j.respe.2017.05.004>.
- Carmina E, Rosato F, Janni A, Rizzo M, Longo RA. Relative prevalence of different androgen excess disorders in 950 women referred because of clinical hyperandrogenism. *J Clin Endocrinol Metab.* 2006;91:2–6. <https://doi.org/10.1210/jc.2005-1457>.
- Melo AS, Ferriani RA, Navarro PA. Treatment of infertility in women with polycystic ovary syndrome: approach to clinical practice. *Clinics (Sao Paulo).* 2015;70:765–9. [https://doi.org/10.6061/clinics/2015\(11\)09](https://doi.org/10.6061/clinics/2015(11)09).
- Stepito NK, Cassar S, Joham AE, Hutchison SK, Harrison CL, Goldstein RF, et al. Women with polycystic ovary syndrome have intrinsic insulin resistance on euglycaemic–hyperinsulinaemic clamp. *Hum Reprod.* 2013;28:777–84. <https://doi.org/10.1093/humrep/des463>.
- Jaliseh HK, Tehrani FR, Behboudi-Gandevani S, Hosseiniapanah F, Khalili D, Cheraghi L, et al. 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. *Fertil Steril.* 2017;108:1078–84. <https://doi.org/10.1016/j.fertnstert.2017.09.004>.
- Amiri M, Ramezani Tehrani F, Behboudi-Gandevani S, Bidhendi-Yarandi R, Carmina E. Risk of hypertension in women with polycystic ovary syndrome: a systematic review, meta-analysis and meta-regression. *Reprod Biol Endocrinol.* 2020;18:23. <https://doi.org/10.1186/s12958-020-00576-1>.
- Atiomo W, Rizwan MNH, Bajwa MH, Furniturewala HJ, Hazari KS, Harab D, et al. Prevalence and diagnosis of PCOS using electronic health records: a scoping review and a database analysis. *Int J Environ Res Public Health.* 2024;21: 354. <https://doi.org/10.3390/ijerph21030354>.
- Barrera FJ, Brown EDL, Rojo A, Obeso J, Plata H, Lincango EP, et al. Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review. *Front Endocrinol.* 2023;14:14. <https://doi.org/10.3389/fendo.2023.1106625>.
- Amiri M, Fallahzadeh A, Sheidaei A, Mahboobifard F, Ramezani TF. Prevalence of idiopathic hirsutism: a systematic review and meta-analysis. *J of Cosmetic Dermatology.* 2022;21:1419–27. <https://doi.org/10.1111/jocd.14313>.
- Perlbarg J, Allonier C, Boissault P, Daniel F, Le Fur P, Szidon P, et al. Faisabilité et intérêt de l'appariement de données individuelles en médecine générale et de données de remboursement appliqué au diabète et à

l'hypertension artérielle. *Santé Publique*. 2014;26:355–63. <https://doi.org/10.3917/spub.139.0355>.

30. Scailteux L-M, Droitcourt C, Balusson F, Nowak E, Kerbrat S, Dupuy A, et al. French administrative health care database (SNDS): The value of its enrichment. *Therapies*. 2019;74:215–23. <https://doi.org/10.1016/j.therap.2018.09.072>.

Publisher's Note

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