National Institute for Health and Care Excellence

Polycystic ovary syndrome: assessment and management

Supplement A: Economic analysis report for AMH testing versus Transvaginal ultrasound

NICE guideline [NGXX]

Evidence underpinning recommendations [XX to XX]

July 2026

Draft

1

ISBN: [XXX]

1

Contents

2

1 Introduction

The topic of Transvaginal Ultrasound (TVUS) and Anti-Müllerian Hormone (AMH) for identifying polycystic ovary morphology (PCOM) was jointly identified by the committee and the guideline health economist as a high priority for new economic analysis. Although one existing health economic analysis was identified in this area, this was a cost-consequence analysis and was rated as being partially applicable with potentially serious limitations. The base case analysis also differed from how the committee thought AMH would be utilised in clinical practice. In addition, this topic was also an area where the International Guideline (IG) recommendations differed from current UK clinical practice.

In current practice, TVUS is offered to people to identify PCOM when there is clinical uncertainty regarding a PCOS diagnosis. PCOM identification is required if there is clinical uncertainty of a PCOS diagnosis after assessment of oligomenorrhea, clinical hyperandrogenism and biochemical hyperandrogenism. Coversley, the IG recommended that either a TVUS scan or an AMH test could be used for identifying PCOM. Therefore, due the differential between current UK practice and the IG's recommendations – and therefore the subsequent potential change in UK practice if the IG's recommendations were directly contextualised – the committee concluded that additional health economic evidence was required to supplement the existing evidence-base in order to aid their consideration of the most cost-effective use of NHS resources.

For further details on the included study mentioned above, and the committee's discussion of this evidence please see Adaptation Report 1. Details of the included study can be found in section 1.4.2 and Appendix C. The committee's discussion of the evidence can be found in section 1.4.4.

2 Health economic modelling methods

2.1 Model overview

A decision tree and Markov model was developed in Microsoft Excel to assess the relative cost effectiveness of an AMH test compared with a TVUS scan for diagnosing PCOS (when identification of PCOM is required.

The cost-effectiveness analysis was from the perspective of the NHS and Personal Social Services in England, in line with the NICE reference case for health interventions (see NICE's guidelines manual). Outcomes were measured in quality-adjusted life years (QALYs). Both costs and QALYs were discounted at an annual rate of 3.5%, as recommended by NICE's guidelines manual. The cost year of the analysis was 2024/25.

The model was run over a lifetime horizon, in accordance with the standard assumptions outlined in the NICE reference case for interventions with health outcomes in an NHS setting. A lifetime horizon was applied to capture all relevant differences in costs and outcomes between the strategies under comparison.

2.2 Model population and setting

The population for the model was women, trans men and non-binary people with suspected signs and symptoms of PCOS. The starting age of the model was assumed to be 20. The starting age of the model was decided with the committee and based on their expert opinion, alongside the clinical data utilised in the model. The committee acknowledged that in clinical practice, the average age of a PCOS diagnosis is likely greater than 20 years of age but were hopeful that this guideline would raise awareness of the condition and therefore lead to earlier diagnoses.

Typically, PCOS is initially investigated in primary care, and most diagnoses can also be determined by GPs. The model population was therefore women, trans men and non-binary people with suspected signs and symptoms of PCOS presenting to their GP for further investigation of their symptoms.

The model assumed that to obtain a diagnosis of PCOS, the presence of two of the following criteria listed below was required:

- oligomenorrhea
- hyperandrogenism (clinical, biochemical, or both)
- polycystic ovarian morphology (PCOM)

Of note, for the condition of hyperandrogenism to be met, this could be either clinical hyperandrogenism or biochemical hyperandrogenism alone or could be the existence of both clinical and biochemical hyperandrogenism.

The model assumed that a conclusive diagnosis of PCOS was never made prior to blood tests being conducted – even when oligomenorrhea and clinical hyperandrogenism were assessed to be present by the healthcare professional undertaking the initial evaluation for suspected PCOS. The committee acknowledged that a diagnosis of PCOS is technically possible when these two conditions have been met but noted that blood tests are required to rule out other potentially serious causes of symptom presentation. Therefore, within the framework of a PCOS diagnosis in the UK, blood tests are conducted to assess for biochemical hyperandrogenism and to also rule out other causes of symptom presentation.

The committee noted that the IG's recommendations have the potential to be interpreted differently. Whereby in the IG's diagnostic algorithm it is noted that other causes of symptom presentation should be excluded when diagnosing PCOS, additional detail on what other tests should be conducted to do so are not stated. The committee therefore emphasised the importance of other causes of symptom presentation being ruled out before a diagnosis of PCOS is made. For example, although rare, adrenal tumours can present with similar symptoms to clinical hyperandrogenism. In which case, blood tests measuring testosterone levels are required to rule out the need for referral to secondary care – whereby significantly elevated testosterone levels require further investigation. In general, the committee noted that a testosterone level above 4.8 nmol/L may be indicative that other non-PCOS conditions are elevating a women's testosterone levels and therefore further investigation may be warranted. It was also discussed that generally, testosterone levels for people with PCOS fall within the range of 2.8–5.2 nmol/L. Blood tests are also required to rule out hyperprolactinaemia, thyroid disease and hypogonadism for those presenting with irregular cycles.

Overall, the committee concluded that a diagnosis of PCOS can be made without the presence of biochemical hyperandrogenism (those people with oligomenorrhea and clinical hyperandrogenism). However, blood tests to rule out other causes need to be conducted prior to a diagnosis of PCOS being made.

The diagnostic criteria employed in this health economic model (prior to identifying PCOM) is reflective UK current practice and also reflective of the recommendations made in this guideline.

2.3 Model strategies

The following strategies were considered in the economic analysis:

- Anti-Müllerian Hormone (AMH) test
- Transvaginal ultrasound (TVUS) scan

TVUS is the gold standard for identifying and diagnosing PCOM as part of the PCOS diagnostic pathway. Previously AMH tests have been more commonly used for assessing ovarian reserve rather than identifying PCOM. A systematic literature review was conducted in the IG to ascertain the clinical effectiveness of using AMH for diagnosing both PCOS and PCOM (review question 1.5 in the IG). The clinical evidence identified in this review resulted in the IG recommending that either serum AMH testing or a TVUS scan may be used to define PCOM; however, both tests should not be performed. The guideline committee acknowledged that routinely conducting both tests would be an inefficient use of NHS resources as both tests are not required to identify PCOM. The IG did not recommend AMH for diagnosing PCOS.

As mentioned previously, the option of offering AMH to define PCOM would result in a change in UK current practice. The guideline committee therefore concluded that a cost-effectiveness analysis comparing AMH and TVUS was required to inform their recommendations and determine whether AMH is cost-effective use of NHS resources.

Of note, the NICE guideline committee acknowledged that in current practice when a TVUS cannot be performed, an abdominal ultrasound scan may be used to diagnose PCOM. Abdominal ultrasound scans were, however, not included as a comparator in the analysis due to lack of available clinical evidence in the IG. Anecdotally, the committee discussed that abdominal ultrasounds are less effective at detecting PCOM compared to TVUS scans, whilst the differences in costs are also small. In the

NHS national cost collection, a directly accessed TVUS scan costs £56 when conducted by diagnostic imaging. Whereas a regular ultrasound scan (20 mins or less without contrast) costs £81. The committee acknowledged that as regular ultrasound scans have significantly more applications compared to TVUS scans, the cost of an abdominal ultrasound for identifying PCOM may lower than the price quoted in the national cost collection, therefore the differences in costs may be closer in value. Overall, the committee concluded that TVUS scans are likely to be the more cost-effective scan of the two, and also have the potential to be dominant (less costly and more effective). The cost-effectiveness of abdominal ultrasounds and TVUS was, however, was not formally assessed. The committee also noted that for those people where TVUS scans are not appropriate (either due to preferences or religious reasons), abdominal ultrasound scans are a suitable alternative.

Further information on the clinical evidence utilised in the model for both AMH and TVUS can be found in section 2.5 of this report.

2.4 Model structure

The health economic model consisted of a one-year decision tree (Figure 1) and a subsequent life-time horizon Markov model (Figure 2). The structure of the decision tree was based on a previous published economic analysis (Garay 2025). Both this model and our model assumed that the population of the model was those with signs and symptoms of PCOS.

As a PCOS diagnosis is a diagnosis of exclusion, all people entering the model received an initial GP appointment and a set of blood tests. The aim of these appointments is to assess for oligomenorrhea, clinical hyperandrogenism and biochemical hyperandrogenism whilst simultaneously ruling out other causes. Within the model, a follow-up GP appointment is then assumed to take place after the results of the blood tests are available to discuss the results and determine an initial set of conclusions. If the person under evaluation has oligomenorrhea and hyperandrogenism, a diagnosis of PCOS is made. If, however, the person under evaluation only meets one of these criteria an AMH test or TVUS is undertaken to ascertain whether PCOM is present. If neither oligomenorrhea nor hyperandrogenism are present, it is assumed the person under evaluation does not have PCOS.

In other words, for those people were there is clinical uncertainty regarding a PCOS diagnosis – either when only oligomenorrhea or hyperandrogenism are present – two diagnostic strategies were compared (AMH & TVUS) to assess the cost-effectiveness of identifying PCOM when clinically indicated to require a PCOS diagnosis.

There was a lack of clinical evidence reporting on the sensitivity and specificity for diagnosing PCOS based on the presence, or absence, of oligomenorrhea and hyperandrogenism (for both clinical and biochemical hyperandrogenism). This confined the assumptions of the model, whereby it was not possible to model probability of an incorrect diagnosis for people with the criteria of having either both the presence, or absence, of oligomenorrhea and hyperandrogenism. Therefore, the associated decision tree probabilities that were used in Garay 2025 were also used in this model.

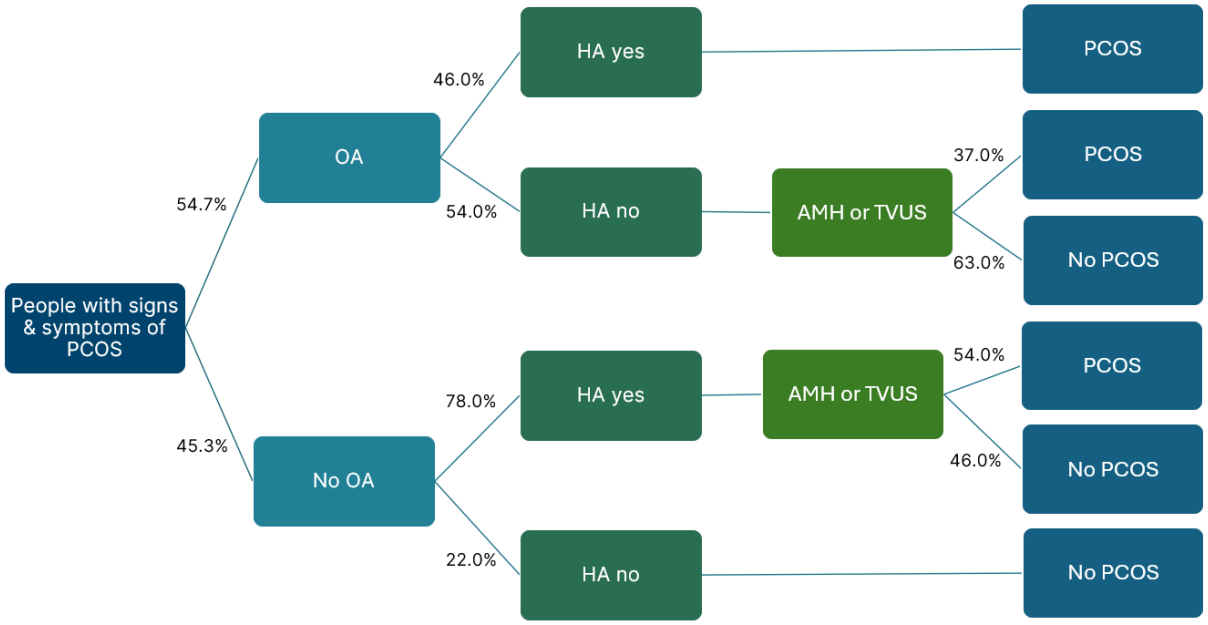
Further information on the data employed in the decision tree section of the model can be found in sections 2.5 and 2.7 of this report. Specifically, data on the diagnostic accuracy of AMH and TVUS can be found in section 2.5 and further information on the rest of the decision tree probabilities can be found in section 2.7.

Overall, the decision tree described the treatment pathway for people presenting to their GP with sing and symptoms of PCOS. The model assumed 100% accuracy for diagnoses where there was a combined presence, or absence, of oligomenorrhea and hyperandrogenism.

Sensitivity and specificity values were applied for AMH & TVUS tests to determine the number of, true positives, false positives, true negatives and false negatives, which were the starting health states of the long-term Markov model.

The structure of the decision tree is shown in Figure 1.

Figure 1. Schematic diagram of the decision tree structure



Abbreviations: AMH = Anti-Müllerian Hormone; HA = Hyperandrogenism; OA = Oligomenorrhea; PCOS = Polycystic ovarian syndrome; TVUS = Transvaginal ultrasound

To provide a summary of when costs and QALYs are accrued throughout the decision tree – prior to be people being categorised as having either oligomenorrhea or no oligomenorrhea, people will have received an initial GP appointment, a set of blood tests and a follow-up GP appointment to discuss the results of their blood tests. At this follow-up GP appointment, the healthcare professional evaluating the person with signs and symptoms of PCOS can ascertain whether the person under evaluation has hyperandrogenism (whilst simultaneously ruling out other causes of symptom presentation). If the person under evaluation has been assessed to have oligomenorrhea and hyperandrogenism, no further costs are accrued in the decision tree, with the same being true for those assessed as not having either oligomenorrhea or hyperandrogenism. For these two cohorts, the subsequent respective costs of having PCOS and not having PCOS are modelled in the long-term Markov model.

In the long-term Markov model for those people meeting the criteria of having oligomenorrhea and hyperandrogenism a utility value of a true positive diagnosis is assigned to those with PCOS, however, in the decision tree this utility value is assumed to be the same value as the false negative utility value as this represents

untreated PCOS. Once this cohort has started receiving in treatment, in the life-time horizon Markov model, a utility value of a true positive is applied. A utility value associated with a true negative diagnosis is assigned to those who do not have PCOS (no oligomenorrhea and no hyperandrogenism).

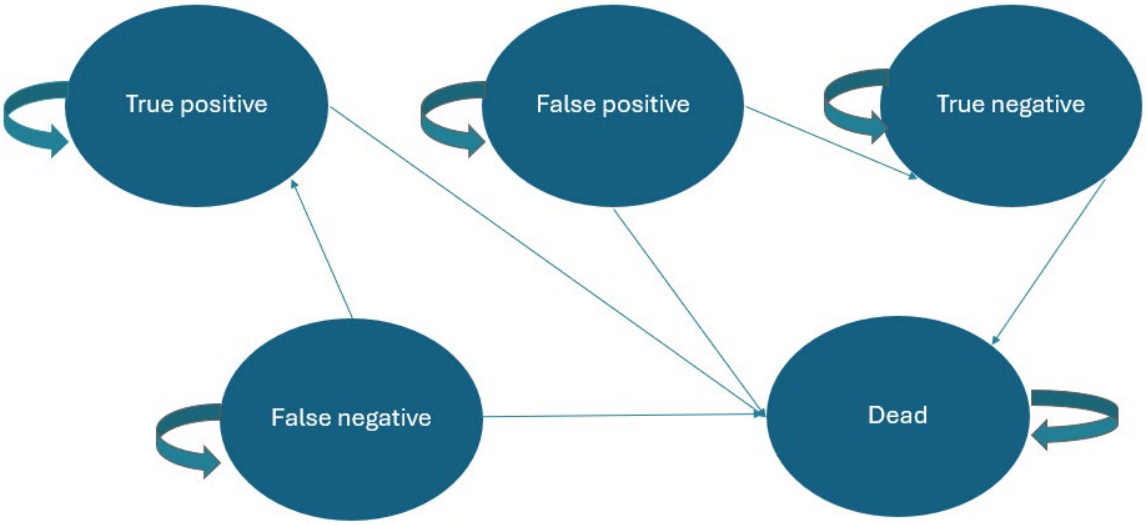
For those meeting only one criterion of oligomenorrhea or hyperandrogenism, after their second GP appointment, this cohort will go onto receive either an AMH test or a TVUS scan, incurring the cost associated with these tests. For AMH, people incur the cost of the test and the cost of a follow-up GP appointment to discuss the results of their AMH test. For TVUS it was assumed 80% of people would receive a directly accessed TVUS scan and the remaining 20% of people would receive an outpatient TVUS scan. All people who received a directly accessed TVUS had a follow-up GP appointment to discuss the results of their TVUS. For those people receiving a TVUS as an outpatient, only 70% had a follow-up GP appointment within the timeframe of the decision tree as it was assumed that additional information and short-term treatment strategies would be provided at the time of the TVUS scan outpatient appointment. A scenario analysis was conducted where 100% of people received a GP appointment after an outpatient TVUS scan. In terms of QALYs, as in line with the methodology detailed above, a utility value of a false negative is applied to people with a true positive diagnosis in the decision tree. In the long-term Markov model, those with a true positive diagnosis are applied a utility value that corresponds to this outcome to represent the fact that treatment for symptoms of PCOS has commenced and started to take effect.

The long-term outcomes of the model were evaluated using a cohort simulation. Based on the outcomes from the decision tree people were categorised as either having PCOS or not – entering the Markov model into one of the four health states. In other words, the diagnosis status at the end of the decision tree was either correct or incorrect, based on the accuracy data of the test. The long-term costs and effects of such were modelled in the life-time horizon Markov model. The Markov model consisted of annual cycles, and a half cycle correction was applied. A half cycle correction is an adjustment that accounts for the fact that health events and transitions between health states can occur at any point within a cycle, rather than exclusively at the beginning or end.

Time dependency was incorporated by referencing the cohort's age as a risk factor for mortality. Baseline utility was also modelled as time-dependent and conditional on the number of years since model entry. People could transition to the dead state at the end of each annual cycle, as determined by general population mortality transition probabilities.

Those people with an incorrect diagnosis (false positive or false negative) had an annual probability of obtaining a correct diagnosis. In the base case analysis this probability was assumed to be the same for both comparisons (AMH & TVUS) and both scenarios (transitioning from a false positive to a true negative and transitioning from a false negative to a true positive).

Figure 2: Schematic diagram of the life-time horizon Markov model



Costs were applied to each health state and were obtained from Berni 2024 who reported an annual cost of PCOS per person versus controls based on real-world evidence data. In the long-term Markov model, the annual cost per person for PCOS reported in Berni 2024 was applied to all people in the model apart from those with a true negative diagnosis (where the cost for controls was applied). Further information on the rational underpinning this assumption can be found in section 2.10. These costing assumptions were also tested in scenario analyses.

Utilities values were applied to the corresponding health states of the life-time horizon Markov model (true positive, false positive, true negative, false negative). As

in line with the methodology for costs, all utilities values were assumed to be the same apart from for those people who have a false negative. Further information on the rational underpinning this assumption can be found in section 2.9. Once again, this assumption was also tested in scenario analyses.

When comparing this original health economic analysis to the included health economic study conducted by Garay 2025, this analysis employed the same decision-tree structure as that employed by Garay 2025. This original cost-utility analysis also utilised the calculated data from Garay 2025 to populate the decision tree probabilities. The data employed for diagnostic accuracy differed slightly. For AMH, Garay 2025 used a single study for diagnostic accuracy data whereas this analysis employed the meta-analysed outcomes from the included studies in the IG (see section 2.5 for further information). For the diagnostic accuracy of TVUS, values were meta-analysed from the IG. This meta-analysed outcome also slightly differed from the value employed in Garay 2025. Further details can be found in section 2.5 of this report.

The two analyses also differed in how long-term outcomes were measured and the application of QALYs for this analysis. In Garay 2025, long-term outcomes were captured by assuming a risk for type two diabetes and cardiovascular disease for those with PCOS. This risk was lowered when a correct diagnosis was obtained through the prescription of an exercise intervention. Whereas this analysis estimated the long-term costs and effects by applying a cost and a utility value to the respective health states of the long-term Markov model (Figure 2). The probability, and the costs, associated with obtaining a correct diagnosis for those who initially receive an incorrect diagnosis was also incorporated into the model. In Garay 2025, it was assumed that all people who received a diagnosis of PCOS (true positive or false positive) received an exercise intervention. This was, however, not the case in this original economic analysis as the committee concluded that this does represent current practice or reflect the recommendations contextualised from the IG. Subsequently in Garay 2025 those people with a true positive diagnosis benefited from a risk reduction is type two diabetes and cardiovascular risk resulting from this exercise intervention. However, those with a false negative diagnosis (people with PCOS who have been incorrectly diagnosed) did not receive the exercise

intervention and were therefore at greater risk of developing type two diabetes and cardiovascular disease.

2.5 Diagnostic accuracy

Diagnostic accuracy data were derived from the relevant review questions in the International Guideline's systematic review and pairwise meta-analysis. The following outcomes were utilised in the model:

- The sensitivity of AMH
- The specificity of AMH
- The sensitivity of TVUS
- The specificity of TVUS

Anti- Müllerian hormone

The sensitivity and specificity of AMH testing for identifying PCOM was obtained from the IG's clinical review and therefore the evidence employed in this model sits within section 1.5 of the technical report produced by the IG. Further details on how NICE contextualised this evidence can be found in adaptation report 1 in section 1.5.

Review question 1.5 in the IG was split into two review questions (1.5.1 and 1.5.2). Review question 1.5.1 sought to determine whether AMH is effective for the diagnosis of PCOS and review question 1.5.2 sought to determine whether AMH is effective for the diagnosis of PCOM. Therefore, the clinical evidence employed within this model sits within evidence pertaining to review question 1.5.2 of the IG, which is specific to PCOM. The PICO criteria for this review question can be found in the technical report document in the IG.

The search strategy for the literature review, PRISMA flowchart and list of included and excluded studies can also be found in the IG technical report (2023).

Overall, seven studies were included that assessed the accuracy of AMH to diagnose PCOM. All studies were in an adult population, and these meta-analysed outcomes informed the sensitivity and specificity of AMH in the model. The resulting sensitivity and specificity employed in this model are 80% (95% CI 72%, 86%) and 84% (95% CI 79%, 88%) respectively. Details relating to the AMH assay used, the threshold for

identifying PCOM and the associated sensitivity and specificity can be found in Table 1. The forest plot for the meta-analysis can be found on page 459 of the IG technical report (2023). Further details on the study setting and population for the meta-analysed studies can be found in Appendix 1, alongside a screenshot of the forest plots from the IG.

The meta-analysed sensitivity & specificity values from the IG were applied to the estimated number of people with and without PCOS, for those who received an AMH test, to obtain the number of true positives, false positives, true negatives and false negatives. Please see section 2.7 for further information on how the estimated number of people with and without PCOS was calculated.

Table 1: Breakdown by individual study of PCOM assays, thresholds and diagnostic accuracy

Author & year	Assay	Threshold	Diagnosis for PCOM	Sensitivity (95% CI)	Specificity (95% CI)
Bell 2021	Beckman Access 2	44pmol/L (6.16ng/mL)	AE-PCOS	0.81 (0.63, 0.93)	0.80 (0.72, 0.86)
Bozdag 2019	Elecsys	3.31ng/ml	Rott/NIH/AE	0.53 (0.45, 0.62)	0.93 (0.87, 0.97)
Dietz de loos 2021 (development cohort)	Elecsys	3.2ng/ml	Rotterdam	0.86 (0.83, 0.89)	0.86 (0.83, 0.89)
Dietz de loos 2021(validation cohort)	Elecsys	3.2ng/ml	Rotterdam	0.89 (0.85, 0.91)	0.85 (0.81, 0.88)
Eilertsen 2012	ELISA	20pmol/L (2.8ng/mL)	Rotterdam	0.80 (0.71, 0.87)	0.72 (0.65, 0.79)
Lauritsen 2014	ELISA	20pmol/L (2.8ng/mL)	Rotterdam	0.82 (0.77, 0.87)	0.84 (0.76, 0.91)
Indran 2018	Beckman	37pmol/L (5.18ng/mL)	Rotterdam	0.79 (0.73, 0.85)	0.83 (0.76, 0.88)

Scenario analyses were conducted varying the sensitivity and specificity values for AMH. These scenarios used single study data from the included studies in the meta-analysis, with one of the three scenarios using the data that was employed in the

existing health economic analysis by Garay 2025. For further information on these scenarios please see section 2.11 of this report.

Transvaginal ultrasound

The IG conducted a systematic literature review which sought to determine the most effective ultrasound criteria for diagnosing PCOS and to determine when ultrasound is indicated to diagnose PCOS (review question 1.4). The PICO criteria for this review question can be found in the technical report of the IG.

The literature review search strategy, PRISMA flowchart and list of included and excluded studies can also be found in the IG technical report (2023). Further details on how NICE contextualised this evidence can be found in adaptation report 1 in section 1.4.

In line with Garay 2025, the committee agreed that follicle number per ovary (FNPO) is the most appropriate outcome for determining PCOM. For this outcome, the IG included 16 studies and meta-analysed 15 of these. For the IG's meta-analysis no overall summary statistic was provided or presented alongside their forest plot. The included studies and forest plot for FNPO can be found on pages 161 – 163 in the IG technical report (2023) and can also be found in Appendix 2 of this report.

Garay 2025 intended to conduct their own meta-analysis and therefore conducted a systematic literature review on the diagnostic accuracy of TVUS for identifying PCOM. Garay 2025 noted that they identified 34 studies, but these studies were all excluded due to applicability concerns, primarily due to the majority of studies focusing on diagnosing PCOS rather than PCOM. No further information on the excluded studies was provided.

Garay 2025 therefore meta-analysed the included studies in the IG to obtain the sensitivity and specificity of TVUS. Garay 2025 acknowledged that the included studies in the IG were also predominantly based on PCOS but noted that PCOS without PCOM (those with phenotype B) is relatively limited. In addition, based on the model assumptions, those people with phenotype B would be diagnosed with PCOS after an initial GP appointment and set of blood tests. This is because those with phenotype B are categorised as having OA and HA (and no PCOM) as their clinical

features of PCOS. Further information on phenotype distributions can be found in section 2.7 of this report.

There was a discrepancy between the included studies included in the meta-analysis for Garay 2025 and the IG's meta-analysed studies presented in their forest plot. For reference, the included IG studies and forest plot can be found in Appendix 2. The discrepancies between the approach taken in IG and Garay 2025 are also detailed in Appendix 2 .

For this analysis, outcomes from the IG were re meta-analysed in line with the population of our model and the NICE reference case to obtain an overall summary statistic. This meta-analysis excluded the one study on adolescents (Villaorel 2015) as this did not reflect the population of the model or the existing recommendations made in the IG – that being there are no definitive criteria to define PCOM on ultrasound in adolescents, hence it is not recommended in adolescents. One other outcome was also excluded where two cut-off thresholds were reported for one study. Further information, and a list of the included studies and outcomes utilised in this model to obtain the sensitivity and specificity values for TVUS, can be found in Appendix 2.

The resulting sensitivity and specificity values were 84.6% (95% CI 81.5%, 87.3%) and 91.9% (95% CI 88.1%, 94.6%) respectively. These values where slightly lower than those employed in the analysis conducted by Garay 2025 (85.1% and 92.4%). The values that were employed in Garay 2025 were tested in a scenario analysis.

2.6 Model parameters: identification and selection

When parameters related to quality of life, resource use and cost were sought, searches were done in specific databases designed for this purpose, such as the Cost–Effectiveness Analysis (CEA) Registry and the NHS Economic Evaluation Database (NHS EED).

The committee was asked to identify relevant papers. Sources of parameters used in the published CUAs identified in the systematic review were reviewed, and relevant articles that did not meet the formal inclusion criteria were also retrieved and examined if they appeared to report data pertinent to the model. Reference lists of

articles identified through any of these approaches were also examined to identify additional publications of interest.

In cases where published literature lacked values essential to parameterising key aspects of the model, data were obtained from unpublished sources or the committee's expert opinion.

To select model parameter estimates from the retrieved literature, the following selection criteria were used:

- The selected studies should report outcomes that correspond as closely as possible to the health states and events simulated in the model.
- The selected studies should report a population that closely matches the UK population (ideally, the population should be from the UK).
- All other factors being equal, more powerful studies (based on sample size or number of events) were preferred.
- When no clear reason existed to discriminate between multiple potential sources for a given parameter, the committee's expert opinion was used to select the most appropriate data source, or the quality of the available sources was considered.

2.7 Baseline event rates and natural history

The model used the data inputs employed in Garay 2025 for the proportion of people presenting with the clinical features associated with PCOS for the population of the model (oligomenorrhea, hyperandrogenism and PCOM).

Garay 2025 used data from Gabrielli 2012 to obtain the proportion of people with signs and symptoms of PCOS that have oligomenorrhea. Gabrielli 2012 calculated that 54.7% of people presenting with signs and symptoms of PCOS have oligomenorrhea. Data from Lizneva 2016, which reported PCOS phenotype distributions, was used alongside this value of 54.7% to calibrate their decision tree probabilities.

Calibration was conducted through an iterative process assuming the proportion of people with OA is fixed at 54.7%. The number of people with HA for both those with and without OA was then estimated with the objective of aligning calculated data inputs with the PCOS phenotype distributions from Lizneva 2016. This iterative

process was then conducted for PCOM to estimate the proportion of people with PCOM for people and without HA, and with and without OA. Data inputs used in the calibration with brief details pertaining to these studies can be found in Table 2.

Further details on this can be found in the supplementary data for Garay 2025.

Table 2: Data inputs used to estimate decision tree probabilities

Data input	Value	Study details
Proportion of people with OA	54.7%	Gabrielli 2012 is a cross-sectional two-phase study based on primary healthcare units for cervical cancer screening in Brazil
Phenotype distribution		
A: HA, OA and PCOM	19.0%	Lizneva 2016 is a systematic literature review and meta-analysis comparing the prevalence of PCOS phenotypes detected in referral versus unselected populations. 33 studies were identified in a referral population, and 9 studies were in an unselected population. Values employed in this model are those meta-analysed outcomes from an unselected population. Further details on the meta-analysed studies can be found in Appendix 1.
B: HA, OA	25.0%	
C: HA, PCOM	34.0%	
D: OA, PCOM	19.0%	

Of note, the phenotype distributions in the table above do not sum to 100% but sum to 97%. This is because Lizneva 2016 conducted separate pooled analyses for each phenotype (A, B, C, D), rather than forcing them to sum to 100%. This discrepancy can be explained by between study heterogeneity. In addition, some of the studies may have had incomplete data or slightly different classifications.

The resulting data inputs that were calculated by Garay 2025 and utilised in this model are presented below in Table 3.

Table 3: Decision tree probabilities

Description	Value
Proportion of people with OA	54.7%
Proportion of people with OA & HA	46.0%
Proportion of people with OA & PCOM that do not have HA	37.0%
Proportion of people with HA that do not have OA	78.0%
Proportion of people with HA & PCOM that do not have OA	54.0%

These probabilities were applied in the decision tree (Figure 1) to calculate the number of people in the model with and without PCOS.

In Garay 2025, the probability of having PCOM, HA and OA were also included in the analysis as the base case analysis for this study assumed that AMH testing was done at the time of the initial set of blood tests. The committee, however, agreed that this would unlikely represent a cost-effective use of NHS resources, especially given the fact that the model assumes 35% of people will not require additional testing with either AMH or TVUS.

2.8 Mortality

General population mortality rates were applied to the model and obtained from the Office for National Statistics (ONS) lifetables for the years 2022–2024. Disease-specific mortality related to PCOS was not included in the model. This was partly due to a lack of data, but the committee also concluded that the population under evaluation would unlikely observe significant differences in mortality.

The committee noted that people with appropriately managed PCOS are unlikely to have a greater risk of mortality compared to the general population – whilst acknowledging that people with PCOS do have a slightly higher risk of diabetes, cardiovascular disease and endometrial cancer. The committee therefore concluded that if PCOS goes undetected people could be at a slightly higher risk of mortality compared to the general population. This model, however, assumes that 88% to 93% of people are diagnosed correctly within a year (for AMH and TVUS respectively). All people within the model are correctly diagnosed within 17 years but there is a 25% annual probability of obtaining a correct diagnosis (when an incorrect diagnosis was initially obtained) and therefore 96% and 98% of people, respectively, have received a correct diagnosis after four years. In terms of the risk for those with undetected PCOS, the committee noted that people with PCOS are unlikely to develop diabetes or cardiovascular disease before the age of 40. Because the starting age of this model is 20 years of age, all people have been correctly diagnosed by the age of 40.

For endometrial hyperplasia the committee discussed that this would need to go undetected for a long period of time to increase the risk of endometrial cancer, with

symptoms of endometrial hyperplasia typically being troublesome for patients and therefore the overall risk of missing a diagnosis of endometrial hyperplasia is low.

The committee also acknowledged that the hypothetical cohort of people in this model is those presenting with signs and symptoms of PCOS, as opposed to a healthy general population. In other words, the committee noted that people that do not have PCOS, but have signs and symptoms of PCOS, will likely have a higher probability of living with a condition that could affect mortality compared to the general population. The committee, therefore noted that compared to the general population, the differential relative risk of mortality for the population of the model is likely lower.

2.9 Utility data

To express outcomes in the form of QALYs, the health states of the economic model (True positive, False positive, True negative and False negative) must be linked to relevant utility values, which represent the health-related quality of life (HRQoL) associated with specific health states on a scale from 0 (representing death) to 1 (representing perfect health). Utility values are typically estimated using preference-based measures that capture individuals' valuations for the HRQoL experienced in the health states under consideration. QALYs are estimated by multiplying the time spent in a specific health state by the health state utility value representing the HRQoL in that state.

According to NICE's guidance on the selection of utility values for use in cost–utility analysis, the measurement of changes in HRQoL should be reported directly by people with the condition being examined or, if this is not possible, by their carers. The valuation of health states should be based on public preferences elicited using a choice-based method, such as time trade-off (TTO) or standard gamble (SG), from a representative sample of the UK population. NICE currently recommends the EQ-5D-3L (Brooks, 1996; Dolan, 1997) as the preferred measure of HRQoL in adults for use in cost–utility analysis.

Utility values employed in the model were obtained from Hahn 2006, who reported SF-36 for a control population and those with PCOS receiving treatment with metformin at 0 months, 1 month and 6 months. The SF-36 values were presented in

a graph (Hahn 2006, Figure 1) and therefore numerical values from the graph were obtained using WebPlotDigitizer (automeris.io). These SF-36 values can be found in Appendix 3.

SF-36 values were mapped to EQ-5D values using Model 1 from Ara & Brazier (Ara 2008) and are presented below in Table 4.

Table 4: Mapped SF-36 values from Hahn 2006

Health state	Value	95% CI lower value	95% CI upper value	SE
PCOS 0	0.782	0.724	0.840	0.030
PCOS 1	0.783	0.723	0.844	0.031
PCOS 6	0.812	0.755	0.868	0.029
Controls	0.905	0.883	0.938	0.014

Abbreviations: CI: confidence interval; SE: standard error

The Hahn 2006 study was conducted in a German population and all PCOS women (n=64) received monotherapy with metformin. Women with a body weight (BW) ≤60 kg were treated with 500 mg metformin, women with a BW > 60 kg and ≤100 kg were treated with 850 mg and women with a BW >100 kg were treated with 1000 mg of metformin. At baseline (PCOS 0), the mean age was 29.31 (± 6.3 years) and 68.7% of people (n=44) were oligomenorrheic. In terms of the clinical signs of hyperandrogenism, 34.4% of people presented with acne, 20.3% presented with mild alopecia (Ludwig Score 40 kg/m^2), and insulin resistance (Homeostatic Model Assessment for Insulin Resistance [HOMA-IR] > 2.5) was diagnosed in 68.7% of people (median HOMA-IR: 2.9). In terms of fertility, ten women had children and 62.5% of those women who did not have children said that they wished to conceive a child. Overall, 71.9% of the population said they felt afraid of not being able to conceive.

Hahn 2006 noted that the sample of people with PCOS used was comparable (regarding the clinical, endocrine and metabolic characteristics) to a larger sample of German PCOS patients (described in Hahn et al., 2005b).

For the control utility values, the German version of the SF-36 (Bullinger and Kirchberger, 1998) was standardized on a representative population-based sample

(n = 2914) and an age-matched subgroup of female subjects aged 21–30 years (n = 263) was used as comparison group for the primary SF-36 scales. The authors noted that for the Sum scales, no age- or gender-specific reference data exist; therefore, the normative published data which includes both men and women aged 18–65 years was utilised.

The committee concluded that the population evaluated in Hahn 2006 was reflective of UK practice. The committee discussed the limitations of this study, specifically that treatment is with metformin. However, this study was the only study that provided utility values for people with PCOS and controls. To overcome some of the limitations with the study, utilities were applied in a conservative manner. For example, in the base case analysis, all utility values were assumed to be the same for the long-term Markov model apart from those people who receive a false negative diagnosis. In other words, those people with PCOS that is not being treated experience a lower utility value, but all other people have the same utility. In the decision tree people with a true positive diagnosis also have a lower utility value of 0.782 (false negative) to represent untreated PCOS, but in the long-term Markov model their utility increased to 0.812 once they have commenced treatment. The utility values employed in the base case analysis of model can be found in Table 5.

Table 5: Health state utility values employed in the model

Health state	Value	95% CI lower value	95% CI upper value	SE
False negative & True positive in the decision tree (PCOS 0)	0.782	0.724	0.840	0.030
True positive, false positive, true negative (PCOS 6)	0.812	0.755	0.868	0.029

Abbreviations: CI: confidence interval; SE: standard error

A number of assumptions were tested utilising the utility values from Hahn 2006 (Table 6) in scenario analyses, details of which can be found in section 2.11.

Utilities for the general population have been used to age-adjust health state utilities. This was done by calculating an age-adjusted multiplier and multiplying this by the utility value associated with the health state a person is residing in. The age-adjusted multiplier was calculated using the general population utility values for the respective ages of the population throughout the model and dividing this utility value by the utility value for the starting age of the model.

2.10 Resource use and costs

Intervention costs

Intervention costs were estimated by combining the resource use associated with each intervention with respective unit costs.

The NHS England National Cost Collection 2024/2025 (previously known as NHS Reference Costs) were used as the source of unit costs for inpatient and outpatient procedures.

The annual report on Unit Costs for Health and Social Care by the Personal Social Services Research Unit (PSSRU 2024/25) was used to specify costs for community-based and hospital-based staff.

Intervention costs are shown in Table 6.

Table 6. Intervention costs (2024 prices)

Intervention	Resource use details and unit costs	Total intervention cost per person
AMH test	AMH test • £40[1] GP appointment • £45[2] x3 = £135 Blood tests • £15[3]	£190 – for those requiring an AMH test

TVUS scan	TVUS scan
	• 80% of people received a directly accessed TVUS and 20% of people received an outpatient scan
	  ○ Directly accessed - £56[4]
	  ○ Outpatient gynaecology - £220[4]
		£236 – for those requiring a TVUS scan

	• *Total average cost per person - £88.80*
	<u>GP appointment</u>
	• All people receiving a directly accessed TVUS have a follow-up GP appointment after the scan
	  ○ £45[(2)] x3 = £135
	  ○ *Total average cost per person - £108*
	• 70% of people receiving an outpatient gynaecology TVUS scan have a follow-up GP appointment after the scan
	  ○ £45[(2)] x3 = £135
	  ○ *Total average cost per person - £18.90*
	• 30% of people receiving an outpatient gynaecology TVUS scan do not have a follow-up GP appointment after the scan
	  ○ £45[(2)] x2 = £90
	  ○ *Total average cost per person - £5.40*
	<u>Blood tests</u>
	• £15[(3)]	

Sources of unit costs:

1. *Garay 2025*
2. *PSSRU 2024*
3. *Committee estimate based on costs presented from the national cost collection for blood testing. Estimate informed for the average number of blood tests and type of blood tests people would likely require to*
4. *NHS National Cost Collection data 2024/25*

For those people who do not require an AMH test or a TVUS scan, the cost of evaluation for people with signs and symptoms of PCOS is £105 (two GP appointments and one set of blood tests).

Costs associated with model health states

Unit costs for the respective health states in the model were obtained from Berni 2024. This study used the Clinical Practice Research Datalink (CPRD) primary care (Aurum) and linked Hospital Episode Statistics (HES) datasets to estimate the total healthcare costs for a person with and without PCOS. This study provided an estimated annual cost per person for PCOS and controls. The total annual cost for PCOS (per person) was estimated to be £1,546 and the total annual cost for controls was estimated to be £940. The model assumed that all people, apart from those with a true negative diagnosis incurred the cost of £1,546. People with a true negative diagnosis incurred an annual cost of £940. The higher cost (PCOS cost) assigned to those people with a false negative and false positive accounts for the higher costs

associated with an incorrect diagnosis. In the model, once all people have been correctly diagnosed, those with PCOS incur an annual cost of £1,546 a year and those without PCOS incur a cost of £940 a year.

Further information on the methodology underpinning these cost estimates be found in Berni 2024 but to provide a summary, health care contacts were costed and analysed from 2019-2020. The study focused on primary and secondary care setting costs, including general practitioner contacts, prescriptions, inpatient admissions and outpatient attendances. Contacts and costs relating to fertility assessments and treatment were not included due to the restrictions around coding of sensitive data.

To provide a summary of how costs were calculated. Primary care contacts were classified according to staff role and consultation type with costs being derived from the Unit Costs of Health and Social Care 2020 based on mapping tables derived by the authors for previous CPRD studies. Primary care contacts for PCOS were defined by those contacts for which a SNOMED code for PCOS was recorded on the same consultation.

Inpatient admission costs were calculated from the HES admitted patient care dataset (described by number, length of stay, and cost). Healthcare resource groups (HRGs) were assigned to each admission and processed using HRG grouper software. The allocated HRG was then costed by linking to the 2020 National Tariff and adjusting for the nature of the admission and excess length of stay. Admissions were classified as an admission relating to PCOS where a diagnosis code relating to PCOS was recorded as the first diagnosis.

As for inpatient admissions, outpatient appointments were identified where a first diagnosis code was recorded as a PCOS event. Outpatient appointment costs were calculated by collating data from the HES outpatient dataset and processed using HRG 4 grouper software. The allocated HRG was then linked to the 2019 National Tariff.

2.11 Handling uncertainty

Probabilistic analysis

Probabilistic analysis was employed as the base-case analysis. Model input parameters were assigned probability distributions rather than expressed as point estimates (which is the approach adopted in a deterministic analysis). This allowed for a more comprehensive consideration of the uncertainty surrounding the input parameters and accounted for the non-linearity inherent in the economic model structure. As a result, more informative estimates were produced than those generated by deterministic analysis, which relies solely on the mean value of each input parameter and disregards the associated uncertainty (Briggs et al., 2006).

The type of distribution was determined with reference to the properties of data of that type (see Table 7). When possible, each distribution was parameterised using dispersion data from the source from which the value was obtained. When no such data were available, plausible ranges were applied based on committee advice and the usual properties of similar data.

A beta distribution was assigned to proportions, probabilities and sensitivity and specificity values as their values are bound between 0 and 1. Costs were assigned a gamma distribution as they cannot be negative and typically follow a skewed distribution. No distribution was assumed for costs obtained from the NHS national cost collection as they are not characterised by uncertainty. Utility values were made probabilistic by applying a gamma distribution to utility decrements. See Table 7 for further information.

Table 7. Description of the types and properties of distributions used in probabilistic analysis

Parameter	Type of distribution	Properties of distribution
Proportions, probabilities, sensitivity and specificity of AMH & TVUS	Beta	Bounded between 0 and 1. Derived from mean and its standard error, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = $mean^2 \times [(1-mean)/SE^2] - mean$ Beta = $alpha \times [(1-mean)/mean]$

Utility decrements Costs	Gamma	Bounded at 0, positively skewed. Derived from mean and its standard error. Alpha and beta values were calculated as follows: Alpha = $(mean/SE)^2$ Beta = $SE^2/Mean$

Abbreviations: 95% CI = 95% confidence interval; SE = standard error

Summary of model inputs

Table 8 presents the point estimates of all input parameters used in the economic model (employed in deterministic sensitivity analysis) and provides details on the types of distributions assigned to each input parameter and the methods used to define their ranges.

Table 8. Model inputs (deterministic values and associated probability distributions)

Diagnostic accuracy (Base case analysis)	Point estimate	Probability distribution	Sources and notes
Sensitivity of AMH	0.800 0.840	Beta, α = 99.548; β = 24.887	International guideline meta-analysed value for the sensitivity of AMH for identifying PCOM
Specificity of AMH	0.840	Beta, α = 213.326; β = 40.633	International guideline meta-analysed value for the specificity of AMH for identifying PCOM
Sensitivity of TVUS	0.846	Beta α = 502.610; β = 91.492	Original meta-analysis from studies in the international guideline for the outcome follicle number per ovary. Studies included in the meta-analysis are those that align to the model population and NICE's methods. For further information please see section 2.5.
Specificity of TVUS	0.919	Beta α = 77.159; β = 6.346	Original meta-analysis from studies in the international guideline for the outcome follicle number per ovary. Studies included in the meta-analysis are those that align to the model population and NICE's methods. For further information please see section 2.5.

Diagnostic accuracy (Values employed in scenario analyses)	Point estimate	Probability distribution	Sources and notes
Sensitivity of AMH – Dietz de Loos 2021	0.886	NA	Dietz de Loos 2021. One of the studies included in the internation guideline meta-analysis. Also, the study used in Garay 2025.

Specificity of AMH – Dietz de Loos 2021	0.846	NA	Dietz de Loos 2021. One of the studies included in the internation guideline meta-analysis. Also, the study used in Garay 2025.
Sensitivity of AMH – Eliertsen 2021	0.800	NA	Eliertsen 2021. One of the studies included in the internation guideline meta-analysis.
Specificity of AMH – Eliertsen 2021	0.720	NA	Eliertsen 2021. One of the studies included in the internation guideline meta-analysis.
Sensitivity of AMH – Bell 2021	0.810	NA	Bell 2021. One of the studies included in the internation guideline meta-analysis.
Specificity of AMH – Bell 2021	0.800	NA	Bell 2021. One of the studies included in the internation guideline meta-analysis.
Sensitivity of TVUS – Garay 2025	0.851	NA	Meta-analysed summary statistic from Garay 2025, based on studies from the international guideline. For further information on how this value differs from the base case values see section 2.5.
Specificity of TVUS – Garay 2025	0.924	NA	Meta-analysed summary statistic from Garay 2025, based on studies from the international guideline. For further information on how this value differs from the base case values see section 2.5.

1

Baseline probabilities	**Point estimate**	**Probability distribution**	**Sources and Notes**
Probability of oligomenorrhea	0.547	NA	Garay 2025. For further information on how these values were calculated in Garay 2025 please see section 2.7.

Probability of hyperandrogenism with oligomenorrhea	0.460	NA	Garay 2025. For further information on how these values were calculated in Garay 2025 please see section 2.7.
Probability of PCOM if with hyperandrogenism and oligomenorrhea	0.370	NA	Garay 2025. For further information on how these values were calculated in Garay 2025 please see section 2.7.
Probability of hyperandrogenism without oligomenorrhea	0.780	NA	Garay 2025. For further information on how these values were calculated in Garay 2025 please see section 2.7.
Probability of PCOM with hyperandrogenism and no oligomenorrhea	0.540	NA	Garay 2025. For further information on how these values were calculated in Garay 2025 please see section 2.7.

1

Additional probabilities	**Point estimate**	**Probability distribution**	**Sources and Notes**
Proportion of people receiving a directly accessed TVUS scan	0.800	Beta: $\alpha = 4.200$; $\beta = 1.050$	Committee estimate
Proportion of people receiving a follow-up GP appointment if they receive an outpatient TVUS scan	0.700	Beta: $\alpha = 6.800$; $\beta = 2.914$	Committee estimate
Annual probability of obtaining a correct diagnosis	0.250	Beta: $\alpha = 18.500$; $\beta = 55.500$	Committee estimate

2

<table>
<tr><th>Utility data</th><th>Point estimate</th><th>Probability distribution</th><th>Sources and Notes</th></tr>
<tr><td colspan="4">Base case analysis</td></tr>
<tr><td>True positive</td><td>0.812</td><td colspan="2" rowspan="4">Utility values were made probabilistic by calculating the utility decrement between health states.
<table><tr><th>Health state</th><th>Mean</th></tr><tr><td>Full health (U0)</td><td>1.000</td></tr><tr><td>Control (U1)</td><td>0.905</td></tr><tr><td>PCOS 6 (U2)</td><td>0.812</td></tr><tr><td>PCOS 0 (U3)</td><td>0.782</td></tr></table>A gamma distribution was applied to the utility decrements to main ranking of utility scores in every iteration.
<table><tr><th>Utility decrement</th><th>Mean</th><th>Alpha</th><th>Beta</th></tr><tr><td>UD1</td><td>0.095</td><td>α =69.034</td><td>β =0.001</td></tr><tr><td>UD2</td><td>0.093</td><td>α =9.117</td><td>β =0.010</td></tr><tr><td>UD3</td><td>0.030</td><td>α =0.518</td><td>β =0.057</td></tr></table></td></tr>
<tr><td>False positive</td><td>0.812</td></tr>
<tr><td>True negative</td><td>0.812</td></tr>
<tr><td>False negative</td><td>0.782</td></tr>
<tr><td colspan="4">Scenario analyses</td></tr>
<tr><td colspan="4"><u>Scenario 1</u></td></tr>
<tr><td>True negative value changed</td><td>0.858</td><td>NA</td><td>For the true negative health state, it is assumed 50% have control utility & 50% have TP utility.</td></tr>
<tr><td colspan="4"><u>Scenario 2</u></td></tr>
</table>

True negative value changed	0.830	NA	For true negatives 20% have control utility & 80% have true positive utility
False positive value changed	0.782	NA	False positive assumed to have the same utility as a false negative

1

Cost data	**Point estimate**	**Probability distribution**	**Sources and Notes**
AMH cost – base case analysis	£40	Gamma: α =25.0; β =1.600	Garay 2025
AMH cost – scenario 1	£36	NA	Cost provided by committee member. Of note, this is the cost of the test as does not include staff time.
AMH cost – scenario 2	£20	NA	Cost provided by committee member. Of note, this is the cost of the test as does not include staff time.
Directly accessed TVUS	£56	NA	NHS National Cost Collection data 2024/25
Outpatient TVUS (gynaecology)	£220	NA	NHS National Cost Collection data 2024/25
GP appointment	£45	NA	Unit Costs for Health and Social Care by the Personal Social Services Research Unit 2024
Blood tests	£15	Gamma: α =25.0; β =0.600	Committee estimate based on costs from the NHS National Cost Collection data 2024/25
Annual cost of a true positive diagnosis	£1546	Gamma: α =25.0; β =61.840	Berni 2024
Annual cost of a false positive diagnosis	£1546	Gamma: α =25.0; β =61.840	Berni 2024

Annual cost of a true negative diagnosis	£940	Gamma: α =25.0; β =37.600	Berni 2024
Annual cost of a false negative diagnosis	£1546	Gamma: α =25.0; β =61.840	Berni 2024

Sensitivity analysis

Various sensitivity analyses were done to test the robustness of model assumptions. In these analyses, one or more inputs were changed at the same time, and the analysis was rerun to evaluate the impact on the results and whether the conclusions would change.

The following one-way, two-way and scenario analyses were explored:

- Three scenario analyses were conducted that jointly varied the sensitivity and specificity of AMH. These analyses were conducted to assess the impact of using different assays used to identify PCOM.
  - Sensitivity and specificity from Dietz de Loos 2021 (sensitivity & specificity employed in Garay 2025)
  - Sensitivity and specificity from Eliertsen 2021
  - Sensitivity and specificity from Bell 2021
- One scenario analysis was conducted that jointly varied the sensitivity and specificity of TVUS. The values employed in this scenario were the ones that were also employed in Garay 2025.
- Two scenario analyses were conducted that tested the utility value assumptions
  - The first scenario was the same as the base case analysis except it was assumed that those with true negative diagnosis have a higher utility value. This higher utility was calculated by assuming 50% have the control utility value and 50% have a true positive utility value. The corresponding utility values for the health states true positive, false positive, true negative and false negative were 0.812, 0.812, 0.858 and 0.782 respectively (further information on utilities can be found in section 2.9). The proportions employed in this analysis was based on committee opinion.
  - The second scenario assumed the same utility value for an incorrect diagnosis, whilst also altering the value of a true negative whereby 20% of people have the control utility value and 80% of people have the true positive utility value. The corresponding utility values for the health states true positive, false positive, true negative and false negative were 0.812, 0.782, 0.830 and 0.782 respectively. The proportions employed in this analysis was based on committee opinion.

- Two scenarios were conducted where the cost of AMH was lowered from the value of £40 in the base case analysis. These costs were obtained from committee members.
  - In one scenario the cost of an AMH test was assumed to be £36
  - In another scenario the cost of an AMH test was assumed to be £20
- In one scenario analysis the cost of a true negative was changed to £1546 (from £940). Therefore, in this scenario the cost of all health states in the life-time horizon Markov model were the same.
- One scenario analysis was conducted where it was assumed the proportion of people receiving a directly accessed TVUS scan was 50% (compared to 80% in the base case analysis). The remaining 50% of people receive an outpatient TVUS scan.
- Two scenario analyses were conducted where the proportion of people receiving a GP appointment after an outpatient TVUS scan was altered. One analysis assumed 20% of people received a GP appointment after an outpatient TVUS scan and one analysis assumed 100% of people received a GP appointment after an outpatient TVUS scan (compared to 70% in the base case analysis).
- Four scenario analyses were conducted where it was assumed that people had a probability of receiving the other test (AMH or TVUS) after their initial test. These were as follows:
  - Probability of receiving a TVUS scan after receiving an AMH test – 20%
  - Probability of receiving a TVUS scan after receiving an AMH test – 75%
  - Probability of receiving an AMH test after receiving a TVUS scan – 20%
  - Probability of receiving an AMH test after receiving a TVUS scan – 75%
  - In these analyses, the cost of receiving the additional was applied. In addition, a combined sensitivity and specificity for receiving both tests was calculated. Details of these calculations can be found in Appendix 4.

All scenario analyses were run deterministically.

2.12 Cost effectiveness estimation and interpretation

To assess cost effectiveness, the incremental cost–effectiveness ratio (ICER) was estimated. The ICER represents the additional cost per additional unit of effectiveness (QALY) associated with one strategy relative to its comparator. The

ICER was calculated by dividing the difference in total costs between 2 alternative strategies by the difference in their QALYs. If the ICER falls below a specified cost per QALY threshold, the assessed strategy is then considered cost effective relative to its comparator. A threshold of £20,000 per QALY has typically been used in analyses done for NICE guidelines (see NICE's guidelines manual).

Figure 3. ICER calculation

$$ICER = \frac{Costs(B) - Costs(A)}{QALYs(B) - QALYs(A)}$$

Where B is the assessed strategy and A is the comparator. Strategy B is cost effective if the ICER is less than the threshold.

Above a most plausible ICER of £20,000 per QALY gained, decisions about the acceptability of the technology as an effective use of NHS resources specifically consider the following factors (see NICE's health technology evaluations manual):

- the degree of certainty and uncertainty around the ICER
- aspects that relate to uncaptured benefits and non-health factors
- aspects that relate to health inequalities.

A strategy is considered cost effective if it results in higher QALYs and lower costs than its comparator. In this case, the strategy is dominant, and the estimation of the ICER is not needed (and it is not meaningful).

Net health benefit (NHB) was also estimated for each strategy, which represents the impact on population health associated with the strategy. NHB was calculated as the difference between the total QALYs gained and the health expected to be forgone elsewhere, with the latter estimated by dividing the total cost associated with the strategy by the £20,000 per QALY gained threshold (see Figure 4).

The strategy with the highest NHB is the most cost-effective option at the specified threshold, that is, the option that provides the highest number of QALYs at an acceptable cost.

Figure 4. NHB calculation

$$Net\ Health\ Benefit\,(X) = \left(QALYs(X)\right) - Costs(X) / \lambda$$

Where X is the assessed strategy and lambda (λ) is the threshold (£20,000 per QALY gained). A strategy is cost-effective if the NHB is the highest net benefit.

Results are also presented graphically in the form of cost–effectiveness planes, which show the incremental costs and QALYs of AMH compared with TVUS.

The probability of the best strategy being the most cost-effective option at the NICE threshold of £20,000 per QALY is also provided, calculated as the proportion of iterations (out of the 10,000 iterations run) in which the best strategy has had the highest NHB among all strategies considered in the analysis.

2.13 Model validation

The economic model was developed by the guideline health economist in consultation with a health economics subgroup formed by members of the committee. The conceptual model, final model structure, and methods for identifying and selecting input parameters were presented to, and agreed by the committee, who also provided clinical validation. All inputs and model formulas were systematically checked. The model was tested for logical consistency by setting input parameters to null and extreme values to examine whether the results were plausible given the inputs and changed in the expected direction. The base-case results and results of sensitivity analysis were discussed with the committee to confirm their plausibility and support interpretation. The model was peer reviewed by a second, independent health economist with relevant experience, this included a systematic check of the model calculations.

3 Results

3.1 Base-case analysis

The results of the model indicated that TVUS was the dominant strategy (cheaper and more effective) when compared to AMH. Over a life-time horizon, TVUS was £27 cheaper and resulted in an additional 0.001 QALY gains. The results also indicated a

very small benefit of TVUS as measured by NHB. The results of the base-case probabilistic analysis are presented in Table 9.

Table 9. Results from the base-case probabilistic analysis

Strategy	Mean cost per person	Mean QALYs per person	Incremental cost	Incremental QALYs	Incremental cost per QALY	NHB (95% CI) at £20,000/QALY
TVUS	£32,172	20.185	-	-	-	18.576 (16.747 to 20.045)
AMH	£32,199	20.183	£27	-0.001	Dominated	18.573 (16.746 to 20.045)

Figure 5 depicts the scatter plot for the sampled costs and QALYs for both AMH and TVUS.

Figure 5. Scatter plot of expected costs and QALYs

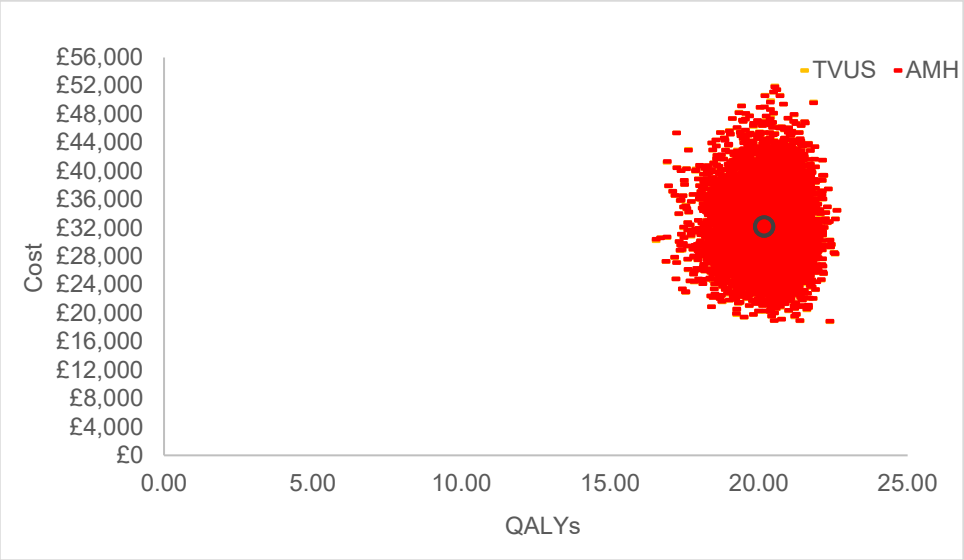


Figure 6 depicts the scatter plot of incremental costs and QALYs for AMH and TVUS. The yellow circle represents the 95% confidence ellipse.

Figure 6. Scatter plot of incremental expected costs and QALYs

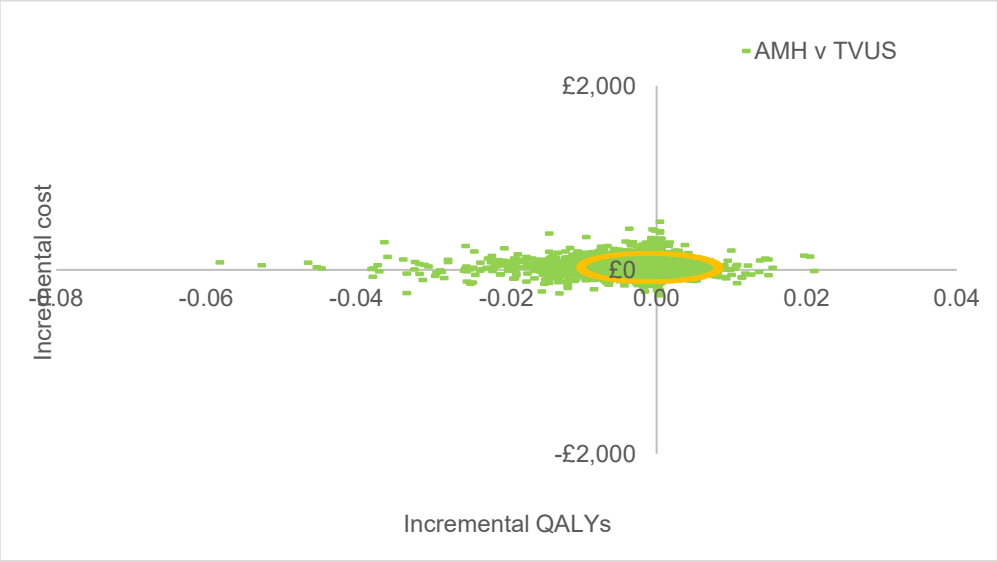
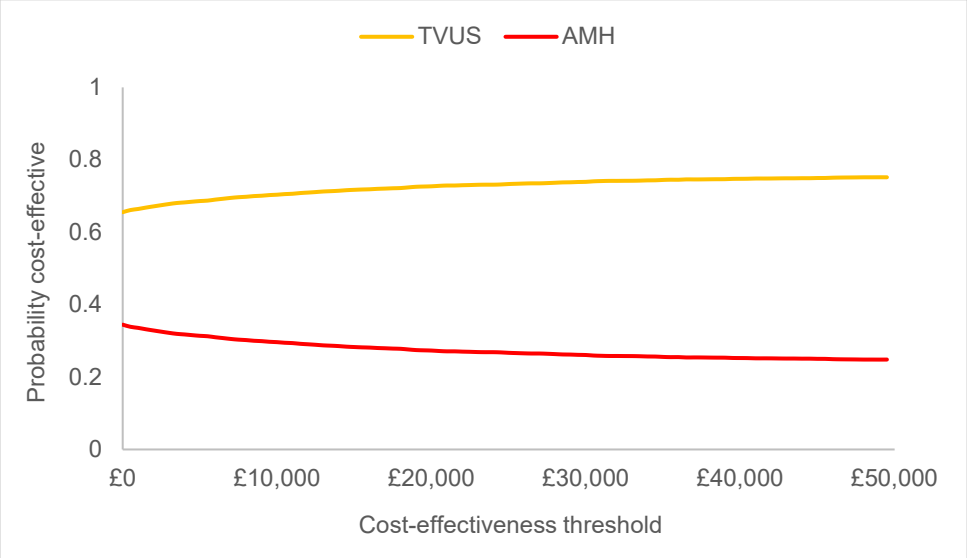


Figure 7 depicts the cost-effectiveness acceptability curve for AMH and TVUS.

Figure 7. Cost–effectiveness acceptability curve



3.3 Scenario analyses

Scenario analyses were run deterministically. The results of the scenario analyses can be found in Table 10.

The deterministic base case results are presented at the top of this table.

Table 10: Scenario analysis results

Strategy	Mean cost per person	Mean QALYs per person	ICER v TVUS	NHB (95% CI) at £20,000/QALY
Deterministic base case results				
TVUS	£32,120	20.171		18.565
AMH	£32,145	20.170	Dominated	18.563

AMH sensitivity and specificity from Dietz de Loos 2021				
TVUS	£32,120	20.171		18.565
AMH	£32,139	20.172	£17,090	18.565
AMH sensitivity and specificity from Eliertsen 2021				
TVUS	£32,120	20.171		18.565
AMH	£32,227	20.170	Dominated	18.558
AMH sensitivity and specificity from Bell 2021				
TVUS	£32,120	20.171		18.565
AMH	£32,172	20.170	Dominated	18.562
TVUS sensitivity and specificity values employed in Garay 2025				
TVUS	£32,116	20.171		18.565
AMH	£32,145	20.170	Dominated	18.563
Utility – true negative diagnoses have a higher utility value (50% control utility and 50% true positive utility)				
TVUS	£32,120	20.686	-	19.080
AMH	£32,145	20.680	Dominated	19.073
Utility – same utility for an incorrect diagnosis and true				

negative utility is: 20% control utility and 80% true positive utility					
TVUS	£32,120		20.374		18.768
AMH	£32,145		20.367	Dominated	18.760
Cost AMH - £36					
TVUS	£32,120		20.171		18.565
AMH	£32,142		20.170	Dominated	18.563
Cost AMH - £20					
TVUS	£32,120		20.171		18.565
AMH	£32,132		20.170	Dominated	18.563
Cost of a true negative - £1,546					
TVUS	£38,861		20.171		18.228
AMH	£38,835		20.170	SW: £20,797	18.228
Proportion of people receiving a directly accessed TVUS scan = 50%					
TVUS	£32,149		20.171		18.564
AMH	£32,145		20.170	SW: £3,333	18.563
Proportion of people receiving a					

GP appointment after an outpatient TVUS scan was altered – 20%						
TVUS	£32,117			20.171		18.565
AMH	£32,145			20.170	Dominated	18.563
Proportion of people receiving a GP appointment after an outpatient TVUS scan was altered – 100%						
TVUS	£32,122			20.171		18.565
AMH	£32,145			20.170	Dominated	18.563
Probability of receiving a TVUS scan after receiving an AMH test – 20%						
TVUS	£32,120			20.171		18.565
AMH	£32,171			20.170	Dominated	18.562
Probability of receiving a TVUS scan after receiving an AMH test – 75%						

TVUS	£32,120			20.171		18.565
AMH	£32,242			20.172	£128,954	18.560
Probability of receiving an AMH test after receiving a TVUS scan – 20%						
TVUS	£32,140			20.171		18.564
AMH	£32,145			20.170	Dominated	18.563
Probability of receiving an AMH test after receiving a TVUS scan – 75%						
TVUS	£32,196			20.172		18.563
AMH	£32,145			20.170	SW: £20,173	18.563

In 12 out of the 17 scenario analyses the results remained unchanged, whereby AMH was dominated. In one scenario analysis, when the proportion of people receiving a directly accessed TVUS scan was 50% (compared to 80% in the base case), AMH was the most cost-effective strategy. However, differences in overall costs and QALYs were marginal (see Table 10). AMH was also marginally cheaper and resulted in slightly higher QALY gains when the sensitivity and specificity values form Dietz de Loos 2021 were employed (£17,090 per QALY gained). In the remaining three scenarios the cost per QALY for AMH was above NICE's £20,000 threshold. As in the base case analysis, all scenario analyses resulted in small cost and QALY differences over a life-time horizon.

4 Discussion

The results of the model indicated that TVUS was the dominant strategy (cheaper and more effective) when compared to AMH. Over a life-time horizon, TVUS was £27 cheaper and resulted in an additional 0.001 QALY gains. The base-case deterministic results yielded very similar results to the probabilistic base case analysis, whereby TVUS was £25 cheaper over a life-time horizon and QALY differences were the same. In addition, in the majority of scenario analyses the overall conclusions of the analysis did not differ. The committee, however, acknowledged that in all analyses, small differences in costs and QALYs were driving these results and conclusions.

For diagnostic accuracy data inputs, parameters were derived from meta-analyses. Therefore, the quality and limitations of the studies included in the meta-analyses unavoidably influenced the quality of the clinical input parameters used in the economic model and, subsequently, the economic analysis results. For example, the economic results may have been affected by reporting and publication bias.

The committee also acknowledged that a key limitation of the analysis is that the model assumes 100% diagnostic accuracy when oligomenorrhea and hyperandrogenism are either both present or both absent. Although this does not reflect clinical practice, the committee considered it unlikely to change the overall conclusions because the assumption would affect both diagnostic alternatives in the model equally.

Another potential limitation of the model is that mortality was assumed to be the same in people with and without PCOS because relevant clinical data were unavailable. The committee acknowledged that the overall effect of PCOS on mortality remains uncertain, particularly because of coexisting conditions associated with PCOS. However, it was noted that an earlier diagnosis of PCOS would mean earlier treatment and could therefore may reduce mortality risk if such risks are prevalent. It was also noted that earlier diagnoses may result in earlier accruing of management costs. Therefore, for the purposes of this analysis, if there was evidence that PCOS resulted in an overall greater risk of mortality, the results of the analysis would likely favour TVUS as a higher proportion of people are diagnosed

sooner due to the higher sensitivity and specificity of TVUS. In terms of earlier management costs, the base case analysis conservatively assumed the same costs for all health states apart from those with a true negative diagnosis. If, however, the model would have assumed greater differences in costs for those with and without PCOS, in the short-term this may favour AMH. Although in general the committee discussed that an earlier diagnosis is always preferable in terms of QALY gains. The committee also acknowledged that an earlier diagnosis is also likely to result in long-term downstream cost savings

For health state costs, contacts and costs relating to fertility assessments and treatment were not included in Berni 2024 due to the restrictions around the coding of sensitive data. However, if these costs would have been captured and included in Berni 2024 this would have likely increased the cost differential between those with PCOS versus controls, once again favouring TVUS due to more people being correctly diagnosed with TVUS scans.

The base case analysis for Garay 2025 assumed that AMH testing was done with the initial set of blood tests. As mentioned previously, the committee, agreed that this would unlikely represent a cost-effective use of NHS resources, as the model assumes 35% of people will not require additional testing with either AMH or TVUS. In general, the committee also agreed that for certain symptom presentations a TVUS would be required over AMH testing to rule out other causes. The committee therefore made recommendations in line with this and noted that this further strengthened the rationale for not conducting AMH testing at the time of initial blood testing as both AMH testing and TVUS are not required to identify PCOM.

When comparing the results of this analysis to Garay 2025, the committee noted that although the scenario analysis that best reflected this model resulted in a different conclusion to the base case results in the analysis undertaken for this guideline. However, in both cases the cost differences between AMH and TVUS were very small.

The committee noted that AMH testing and TVUS scans can be used for other diagnostic purposes. Therefore, there may be instances where both tests are conducted for investigations over and above for identifying PCOM. Scenario

analyses were conducted to assess the overall impact of cost-effectiveness. These were one-way scenario analyses, and the results of these analyses can be found in Table 10. However, in general the committee concluded that there would likely be a greater need for a TVUS scan after an AMH test versus an AMH test after a TVUS scan for diagnostic purposes, thus favouring TVUS as a strategy. It was, however noted that those presenting with fertility problems would likely receive an AMH test irrespective of whether an AMH test or TVUS test was conducted as part of the person's diagnostic pathway for PCOS.

In terms of implementation, the committee noted that there are currently backlogs for accessing TVUS scans and therefore for those where AMH testing is indicated, this could help alleviate this backlog. The committee also noted the need for training on AMH testing for GPs as this is not currently conducted in primary care. Concerns around the potential needed to address any fertility related questions off the back of AMH testing were also raised. The committee noted that further research would be helpful on AMH testing and specific thresholds for identifying PCOM for various populations such as different ethnicities, different ages and body mass indices.

5 Conclusion

Overall, the analysis indicated that TVUS was the dominant strategy. However, small differences in costs and QALYs were driving this result. Taking into account the strengths and weaknesses of both this analysis and the published analysis by Garay 2025, the committee recommended that either TVUS or AMH could be used to identify PCOM but also stipulated the in instances when symptoms or signs could suggest another condition, or co-existing conditions (such as, pelvic pain or abnormal uterine bleeding) TVUS should be the strategy of choice. The committee noted than in these instances a TVUS is required to rule out other more series causes of symptom presentation which is not possible with AMH testing alone.

6 References

Bell R, Islam R, Skiba M, Herbert D, Garcia A, Davis S (2021) Substituting serum anti-Müllerian hormone for polycystic ovary morphology increases the number of

women diagnosed with polycystic ovary syndrome: a community-based cross-sectional study, Human reproduction Volume 37 Issue 1

Berni TR, Morgan CL, Rees DA (2025) Rising Incidence, Health Resource Utilization, and Costs of Polycystic Ovary Syndrome in the United Kingdom. J Clin Endocrinol Metab 110(5):e1580–e1589.

Briggs A, Sculpher M, Claxton K (2006) Decision Modelling for Health Economic Evaluation. New York, NY: Oxford University Press

Brooks R (1996) EuroQol: the current state of play. Health Policy 37(1), 53-72

Dietz de Loos A, Hund M, Buck K, Meun C, Sillman J, Laven J (2021) Antimullerian hormone to determine polycystic ovarian morphology, Fertility and Sterility Vol. 116, No. 4, October 2021 0015-0282

Dolan P (1997) Modelling valuations for EuroQol health states. Medical Care 35(11), 1095-108.

Eilertsen T, Vanky E, Carlsen S (2012) Anti-Mullerian hormone in the diagnosis of polycystic ovary syndrome: can morphologic description be replaced?, Human reproduction Volume 27 Issue 8

Garay OU, Olziersky A-M, Laven J, Mawson R, Piltonen T, Franks S, et al (2025) Economic impact of the Elecsys anti-Müllerian hormone Plus immunoassay for anti-Müllerian hormone testing as part of polycystic ovary syndrome assessment in the United Kingdom, PLoS One 20(6): e0326162

Hahn S, Bneson S, Elsenbruch S, Pleger K, Tan S, Mann K, Schedlowski M et al (2006) Metformin treatment of polycystic ovary syndrome improves health-related quality-of-life, emotional distress and sexuality Volume 21 Issue 7

NHS England National Cost Collection 2024/2025 [online; accessed April 2026]

Office for National Statistics [online; accessed February 2026]

PSSRU 2024/25 [online; accessed January 2026]

1

Appendices

Appendix 1

Study author and year	Study country and setting	Total number of people in the study and total number of people with PCOS (%)
Gabrielli 2012	**Brazil:** Primary healthcare units for cervical cancer screening	Total: 859 PCOS: 73 (8.5%)
Lauritsen 2014	**Denmark:** Female employees of Copenhagen University Hospital	Total:447 PCOS: 86 (19.2%)
Li 2013	**China:** Female residents from 10 provinces and municipalities in China	Total: 15924 PCOS: 833 (5.2%)
Ma 2010	**China:** Female residents in the Beijing	Total: 2111 PCOS: 129 (6.1%)
March 2010	**Australia:** Female residents born in the Queen Elizabeth Hospital, were evaluated when they were around 30	Total: 728 PCOS: 130 (17.9%)
Moran 2010	**Mexico:** Female employees of an Obstetrics and Gynecology Hospital of the IMSS	Total: 150 PCOS: 10 (6.7%)
Tehrani 2011	**Iran:** reproductive aged women living in urban areas of four randomly selected provinces: Ghazvin, Kermanshah, Golestan and Hormozgan	Total: 929 PCOS: 136 (14.6%)
Tehrani 2014	**Iran:** Urban areas of three cities of Khouzestan province	Total: 602 PCOS: 85 (14.1%)
Yildiz 2012	**Turkey:** Female employees of General Directorate of Mineral Research	Total: 392 PCOS: 78 (19.9%)

Figure 8: Forest plot for AMH sensitivity

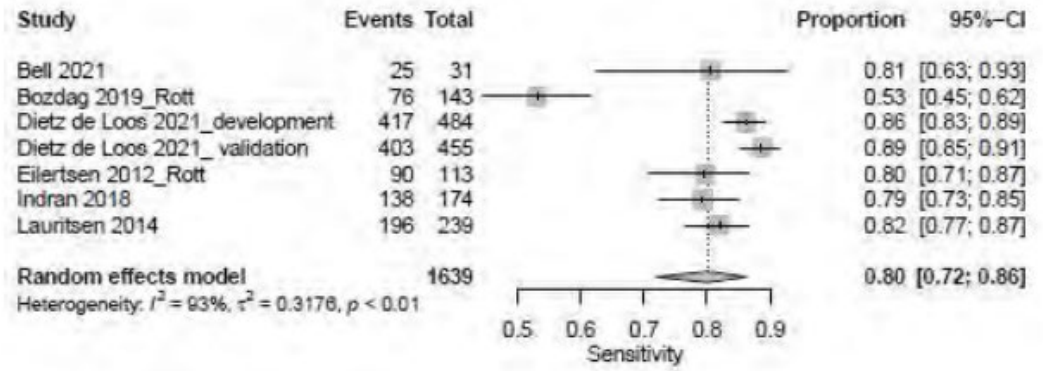
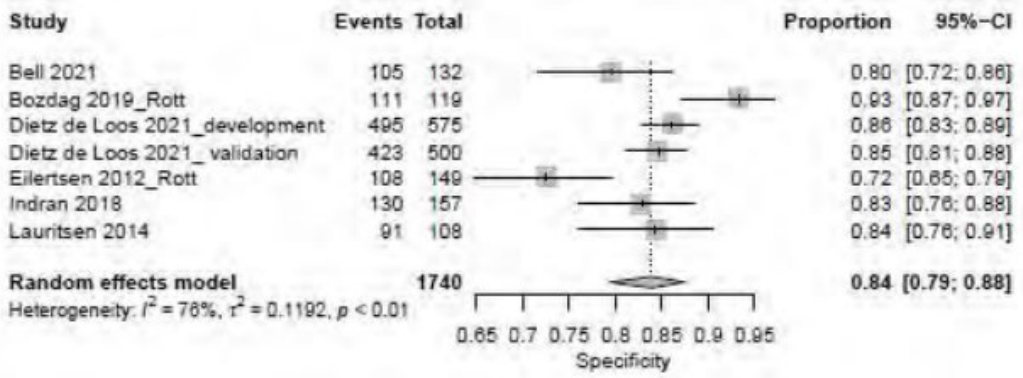


Figure 9: Forest plot for AMH specificity



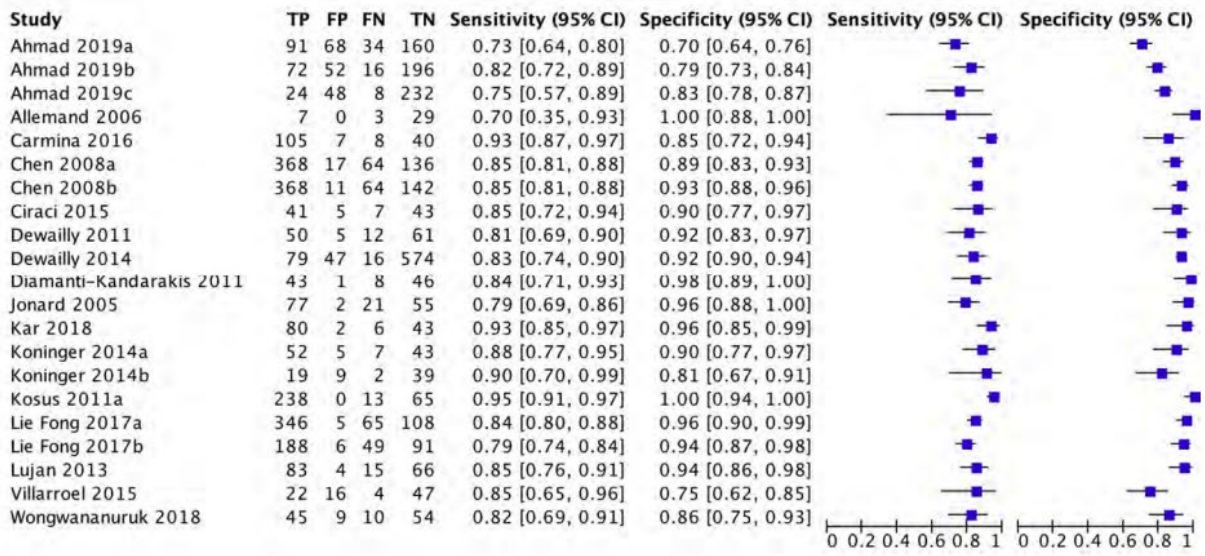
Appendix 2

Table 11: Included studies in the IG for TVUS for the outcome Follicle Number Per Ovary

INDEX TEST: Follicle Number Per Ovary (FNPO)				OUTCOME TYPE: continuous								
COMPARISON (if applicable): N/A												
Author, year	Unit of outcome	Method of measurement	N	Threshold cut-off	True Pos	False Pos	True Neg	False Neg	Sensitivity (95% CI)	Specificity (95%CI)	AUC	Precision
Ahmad et al. 2019 (25 to <30yo)	Count	TVUS (2D Real-Time)	353	15 (Mean)	91	68	160	34	0.730	0.700	N/A	N/A
Ahmad et al. 2019 (30 to <35yo)	Count	TVUS (2D Real-Time)	336	14 (Mean)	72	52	196	16	0.820	0.790	N/A	N/A
Ahmad et al. 2019 (35 to <40yo)	Count	TVUS (2D Real-Time)	312	12 (Mean)	24	48	232	8	0.750	0.830	N/A	N/A
Allemand et al. 2006	Count	TVUS (3D-Offline)	39	20.1 (Mean)	7	0	29	3	0.700	1.00	0.987	N/A
Carmina et al. 2016	Count	TVUS (2D Real-Time)	160	22 (Mean)	105	7	40	8	0.930	0.850	N/A	N/A
Chen et al. 2008	Count	TVUS (2D Real-Time)	585	10 (Mean)	368	17	136	64	0.852	0.888	0.909	N/A
				12 (Max)	368	11	142	64	0.852	0.926	0.911	N/A
Christ et al. 2014	Count	TVUS (2D Offline with Grid Overlay)	142	28 (Mean)	70	1	59	12	0.850	0.980	0.971	0.948-0.993
Çiraci et al. 2015	Count	TVUS (2D Real-Time)	96	12 (Mean)	41	5	43	7	0.854	0.895	0.959	N/A
Dewailly et al. 2011	Count	TVUS (2D Real-Time)	128	19 (Mean)	50	5	61	12	0.810	0.920	0.950	0.915-0.982
Dewailly et al. 2014	Count	TVUS (2D Real-Time)	716	12 (Mean)	79	47	574	14	0.832	0.925	0.940	0.909-0.971
Diamanti-Kandarakis et al. 2011	Count	TVUS (2D Real-Time)	97	19.5 (Mean)	43	1	46	8	0.850	0.980	0.940	0.890-0.990
ionard et al. 2005	Count	TVUS (2D Real-Time)	155	12 (Mean)	77	2	55	21	0.790	0.970	0.960	N/A
Kar and Swoyam 2018	Count	TVUS (2D Real-Time)	131	12 (Mean)	80	2	43	6	0.935	0.952	0.973	N/A
		TVUS (3D Real-Time)		10 (Max)	76	1	44	10	0.880	0.977	N/A	N/A
Köninger et al. 2014 Severe PCOS)	Count	TVUS (2D Real-Time)	107	9.5 (Max)	52	5	43	7	0.881	0.896	0.940	0.880-0.980
Köninger et al. 2014 Mild PCOS)	Count	TVUS (2D Real-Time)	69	8.5 (Max)	19	9	39	2	0.905	0.812	0.920	0.820-0.980
Köşüş et al 2011b	Count	TVUS (2D Real-Time)	316	8 (Mean)	238	0	65	13	0.950	1.00	0.998	0.992-1.004
Lie Fong et al. 2017 Young)	Count	TVUS (2D Real-Time)	524	12.25 (Median)	346	5	108	65	0.842	0.956	0.915	0.891-0.940
Lie Fong et al. 2017 Old)	Count	TVUS (2D Real-Time)	334	10.75 (Median)	188	6	91	49	0.795	0.935	0.874	0.836-0.912
Lujan et al. 2013	Count	TVUS (2D Offline with Grid Overlay)	168	26	83	4	66	15	0.850	0.940	0.969	0.948-0.990
Villaroel et al. 2015 (adolescent)	Count	TAUS	89	12 (Max)	22	16	47	4	0.846	0.746	0.877	P<0.0001
Wongwananuruk et al. 2018	Count	TVUS/TRUS (2D Real-Time)	118	15 (Max)	45	9	54	10	0.818	0.857	0.918	0.866-0.970

The IG conducted a meta-analysis, but no summary statistic was provided. This forest plot for the IG's meta-analysis can be found in Figure 9.

Figure 10: Forest plot for meta-analysed studies in the IG for TVUS for the outcome Follicle Number Per Ovary



Of the included studies from the IG (Table 11), the following studies were not included in their meta-analysis (Figure 10).

- Christ 2014
- Only one outcome for Kar 2018 was included (TVUS 2D real-time outcomes were included and TVUS 3D real-time outcomes were excluded)

When comparing the studies that were included in Garay 2025's meta-analysis (Table 12) for them to obtain an overall summary statistic for the sensitivity and specificity of TVUS.

- Christ 2014 was included in their meta-analysis
    - This was excluded from the IG's meta-analysis
- Only one outcome reported for Chen was included
    - Both were included in the IG's meta-analysis

- Both outcomes for Kar 2018 were included (TVUS 2D real-time and TVUS 3D real-time outcomes)
    - Only TVUS 2D real-time outcomes were included in the IG's meta-analysis

No rationale was provided for in Garay 2025 as to why their approach differed from the IG.

Table 12: Studies from the IG that were meta-analysed in Garay 2025

Study	Year	Sens	Spec	N	PCOS	Controls
Ahmad et al. 2019 (25–<30yo)	2019	0.73	0.7	353	125	228
Ahmad et al. 2019 (30–<35yo)	2019	0.82	0.79	336	88	248
Ahmad et al. 2019 (35–<40yo)	2019	0.75	0.83	312	32	280
Allemand et al. 2006	2006	0.7	1	39	10	29
Carmina et al. 2016	2016	0.93	0.85	160	113	47
Chen et al. 2008 (MaxFN)	2008	0.852	0.926	585	432	153
Christ et al. 2014	2014	0.85	0.98	142	82	60
Çiraci et al. 2015	2015	0.854	0.895	96	48	48
Dewailly et al. 2011	2011	0.81	0.92	128	62	66
Dewailly et al. 2014	2014	0.832	0.925	716	95	621
Diamanti-Kandarakis et al. 2011	2011	0.85	0.98	97	50	47
Jonard et al. 2005	2005	0.79	0.97	211	154	57
Kar 2018 (2D)	2018	0.935	0.952	131	86	45

Kar 2018 (3D)	2018	0.88	0.977	131	86	45
Köninger et al. 2014 (severe PCOS)	2014	0.881	0.896	107	59	48
Köninger et al. 2014 (Mild PCOS)	2014	0.905	0.812	69	21	48
Köşüş et al. 2011	2011	0.95	1	310	210	100
Lie Fong et al. 2017 (young)	2017	0.842	0.956	—	28	118
Lie Fong et al. 2017 (old)	2017	0.795	0.935	—	51	100
Lujan et al. 2013	2013	0.85	0.94	168	98	70
Villaroel et al. 2015 (adolescent)	2015	0.846	0.746	89	26	63
Wongwananuruk et al. 2018	2018	0.818	0.857	118	55	63

For the purpose of NICE's cost-utility analysis, outcomes from the IG meta-analysis were meta-analysed to provide a summary statistic for the sensitivity and specificity of TVUS.

- Villaroel was excluded from this analysis due to this study having an adolescent population and this reflect the population of the model (starting age 20 years of age)

- Only one outcome was included for Chen 2008. This was the max threshold cut-off value and aligned with the approach taken by Garay 2025 for this study

A list of the included studies that were employed to obtain a summary static for the sensitivity and specificity for TVUS can be found in Table 13.

Table 13: Studies included in the meta-analysis to obtain the sensitivity and specificity of TVUS

Author	Year	True positive	False negative	False positive	True negative
Ahmad	2019a	91	34	68	160
Ahmad	2019b	72	16	52	196
Ahmad	2019c	24	8	48	232
Allemand	2006	7	3	0	29
Carmina	2016	105	8	7	40
Chen	2008	368	64	11	142
Ciraci	2015	41	7	5	43
Dewailly	2011	50	12	5	61
Dewailly	2014	79	16	47	574
Diamanti-Kandarakis	2011	43	8	1	46
Jonard	2005	77	21	2	55
Kar	2018	80	6	2	43
Koninger	2014a	52	7	5	43
Koninger	2014b	19	2	9	39
Kosus	2011a	238	13	0	65
Lie Fong	2017a	346	65	5	108
Lie Fong	2017b	188	49	6	91
Lujan	2013	83	15	4	66
Wongwananuruk	2018	45	10	9	54

Appendix 3

Table 14: Mean SF-36-dimension scores from Hahn 2006

	PF	SF	RP	RE	MH	VT	BP	GH
PCOS 0	78.322	69.819	81.827	65.672	61.015	46.414	78.720	65.029
PCOS 1	76.482	73.123	84.750	72.038	63.810	52.525	78.996	64.512
PCOS 6	84.695	76.428	85.026	79.424	67.114	54.044	73.188	67.434
Control	93.966	88.915	87.684	91.147	71.694	62.195	84.576	74.854

Abbreviations: PF: physical functioning; SF: social functioning; RP: role-physical; RE: role-emotional; MH: mental health; VT: vitality; BP: bodily pain; GH: general health

Appendix 4

The combined sensitivity and specificity of AMH and TVUS was calculated by firstly selecting a dependence strength between the two sensitivity and specificity values. For the purpose of this analysis the dependence strength was assumed to be 0.5 which indicates a moderately strong positive correlation between the two tests.

A gamma sensitivity and specificity value was calculated based on this dependence strength and the base case sensitivity and specificity for AMH and TVUS. These calculated values were the used to calculate the combined sensitivity and specificity values. See Table 15 for further details.

Table 15: Calculating combined sensitivity and specificity values

	Value	Source and notes
Sensitivity of AMH	0.800	Value from the international guideline employed in the model
Specificity of AMH	0.840	
Sensitivity of TVUS	0.846	Calculated value based on included studies in the international guideline employed in the model
Specificity of TVUS	0.919	
Dependence strength	0.500	Assumption

Gamma sensitivity	0.0616	Calculated as: Dependence strength * the minimum value of either ○ Sensitivity of AMH * (1 – Sensitivity of TVUS) ○ Sensitivity of TVUS * (1 – Sensitivity of AMH)
Gamma specificity	0.0340	Calculated as: Dependence strength * the minimum value of either ○ Specificity of AMH * (1 – Specificity of TVUS) ○ Specificity of TVUS * (1 – Specificity of AMH)
Sensitivity combined – OR methodology	0.908	Calculated as: Sensitivity of AMH + Sensitivity of TVUS – (Sensitivity of AMH * Sensitivity of TVUS + gamma sensitivity)
Specificity combined – OR methodology	0.806	Calculated as: Specificity of AMH * Specificity of TVUS + gamma specificity
Sensitivity combined – AND methodology	0.738	Calculated as: Sensitivity of AMH * Sensitivity of TVUS + gamma sensitivity
Specificity combined – AND methodology	0.953	Calculated as: 1 – ((1 - Specificity of AMH) * (1 - Specificity of TVUS) + gamma specificity

The committee concluded that the OR methodology for combining sensitivity and specificity was the most appropriate for the topic under evaluation. This is because this methodology is underpinned by the logic that a positive result from either test can detect the condition under evaluation. The OR methodology therefore increases overall sensitivity. Conversely, the AND methodology assumes that a diagnosis is only possible if both tests agree that the condition is present, therefore significantly lowering the combined sensitivity.