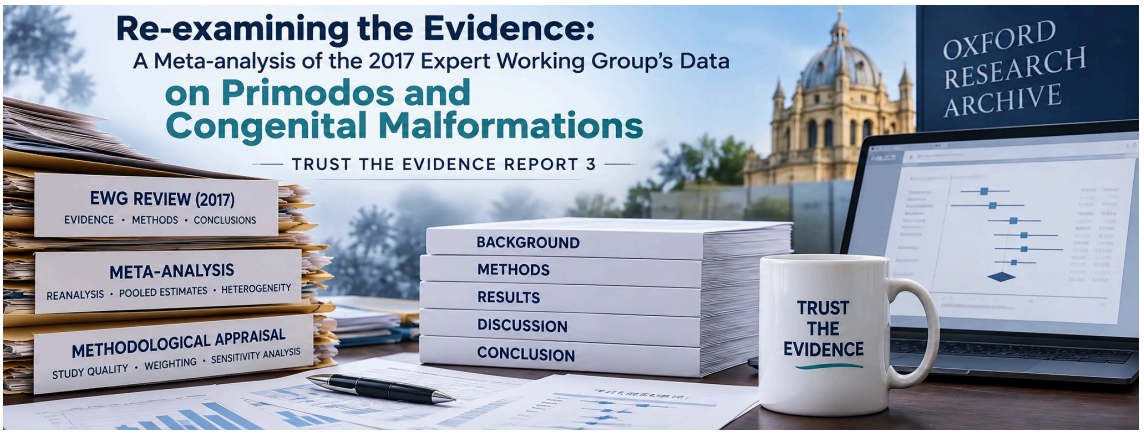


Re-examining the Evidence: A Meta-analysis of the 2017 Expert Working Group's Data on Primodos and Congenital Malformations

Carl Heneghan, [1] Tom Jefferson [1]

[1] Centre for Evidence-Based Medicine, University of Oxford, UK, and Trust the Evidence

Trust the Evidence Publications: Research Report 3: 9th November 2026

Abstract

Background: In 2017, the Commission on Human Medicines' Expert Working Group (EWG) concluded that the available epidemiological evidence did not, on balance, support a causal association between exposure to oral hormone pregnancy tests (HPTs), including Primodos, and congenital malformations. The EWG declined to undertake a formal meta-analysis, citing concerns about study quality, heterogeneity and statistical weighting. This report re-examines that decision and assesses whether quantitative synthesis of the EWG's underlying data provides additional evidence relevant to its conclusions.

Methods: Following the release of the EWG's extracted data in response to a Freedom of Information request, random-effects meta-analyses were conducted to estimate associations between HPT exposure and congenital heart defects, congenital malformations overall, urogenital defects, and other congenital abnormalities. The resulting estimates were compared with those reported in the systematic review and meta-analysis by Heneghan et al. The EWG's methodological objections to quantitative synthesis were also critically examined.

Results: Reanalysis of the EWG data identified statistically significant associations between HPT exposure and congenital heart defects (OR 1.91, 95% CI 1.36–2.68; I^2 = 22%) and any congenital malformation (OR 1.34, 95% CI 1.13–1.60; I^2 = 0%). These findings closely replicated the estimates reported by Heneghan et al. for congenital heart defects (OR 1.89, 95% CI 1.32–2.72) and all congenital malformations (OR 1.40, 95% CI 1.18–1.66). Urogenital malformations showed elevated but statistically inconclusive associations in both analyses. The close agreement between the two analyses supports the analytical reproducibility of the principal findings, although their overlapping evidence bases preclude interpretation as independent confirmation.

Conclusions: The reanalysis demonstrates that quantitative synthesis of the historical evidence assembled by the EWG was feasible and identified statistically significant associations between oral HPT exposure and congenital heart defects and congenital

malformations overall. The consistency with the previously published Heneghan et al. meta-analysis strengthens confidence in the reproducibility of these associations. The findings challenge the EWG's decision to reject quantitative synthesis without formally investigating the effects of study quality, heterogeneity and statistical weighting. A more transparent assessment incorporating meta-analysis, risk-of-bias evaluation and sensitivity analyses provides a more complete account of the available evidence and its uncertainties.

Keywords: Primodos; hormone pregnancy tests; congenital malformations; congenital heart defects; Expert Working Group; meta-analysis; reproducibility; MHRA.

Cite as: Tom Jefferson, Carl Heneghan: Proposed working definition of an active case of influenza. Trust the Evidence Publications: Research Report 3: 9th November 2026. Re-examining the Evidence: A Meta-analysis of the 2017 Expert Working Group's Data on Primodos and Congenital Malformations. https://trusttheevidence.substack.com/p/trust-the-evidence-publications-research-f53

Background

The 2017 Expert Working Group Review

In 2014, the Medicines and Healthcare products Regulatory Agency (MHRA) initiated a review of the evidence concerning the possible association between oral hormone pregnancy tests (HPTs), including Primodos, and congenital malformations. [1]

The review was undertaken by an Expert Working Group (EWG) established by the UK Commission on Human Medicines (CHM). Its findings were published in 2017. The EWG concluded that the available evidence was insufficient to establish an association between exposure to oral HPTs during pregnancy and congenital malformations. It considered the evidence mixed and heterogeneous, with differences among the historical studies limiting the reliability of the conclusions.

Importantly, the EWG did not combine the historical study results into an overall pooled estimate through meta-analysis. The explanation subsequently provided to Parliament was that the EWG considered such an analysis potentially misleading.[2] This decision is central to the present report because it raises the question of whether the available evidence could nevertheless be subjected to appropriate statistical analysis and, if so, whether doing so would yield findings relevant to the EWG's conclusions.

The 2019 Scientific Review

On 18 March 2019, the findings of the systematic review by Heneghan et al.,[3] "*Oral hormone pregnancy tests and the risks of congenital malformations: a systematic review and meta-analysis*," were discussed at a meeting of an ad hoc expert group convened by the MHRA in London. [4]

The purpose of the meeting was to evaluate the suitability and robustness of the methods employed in the systematic review, including the selection and application of quality scores, and to consider any clinical implications arising from its findings. [5] The Heneghan et al. review was particularly relevant because it applied systematic review and meta-analytical methods to historical evidence concerning oral HPT exposure and congenital malformations.

Its findings provided an opportunity to examine whether quantitative synthesis could identify associations that the EWG's narrative assessment had not established.

Requests for the EWG's Underlying Data

Following discussions of the All-Party Parliamentary Group (APPG) on 21 January 2019, Marie Lyon submitted a Freedom of Information (FOI) request to the MHRA on 30 January 2019, seeking the release of the raw data used in the EWG report. A further request was made on 4 February 2019.

Marie Lyon raised concerns that the underlying data used to construct the EWG's forest plots had not been included in the published report, despite the EWG's commitment to transparency and public scrutiny. In her correspondence, she stated:

"*The APPG supporting the Association have put in an FOI request for the raw data used in the Forest plot conclusions. I was not aware this information was excluded in the EWG document. Would you please ask the MHRA/CHM why the raw data was not included and would you also please ask them to expedite the request for this information.*"

She further emphasised the EWG's pledge: "*The EWG pledged to allow 'full public scrutiny and to publish all evidence which had been gathered, together with the assessments of the data.' Exclusion of the raw data does not fulfil this pledge.*"

The correspondence demonstrates that access to the underlying numerical evidence had become an important issue in the independent examination of the EWG's findings.

Chronology of the Freedom of Information Request

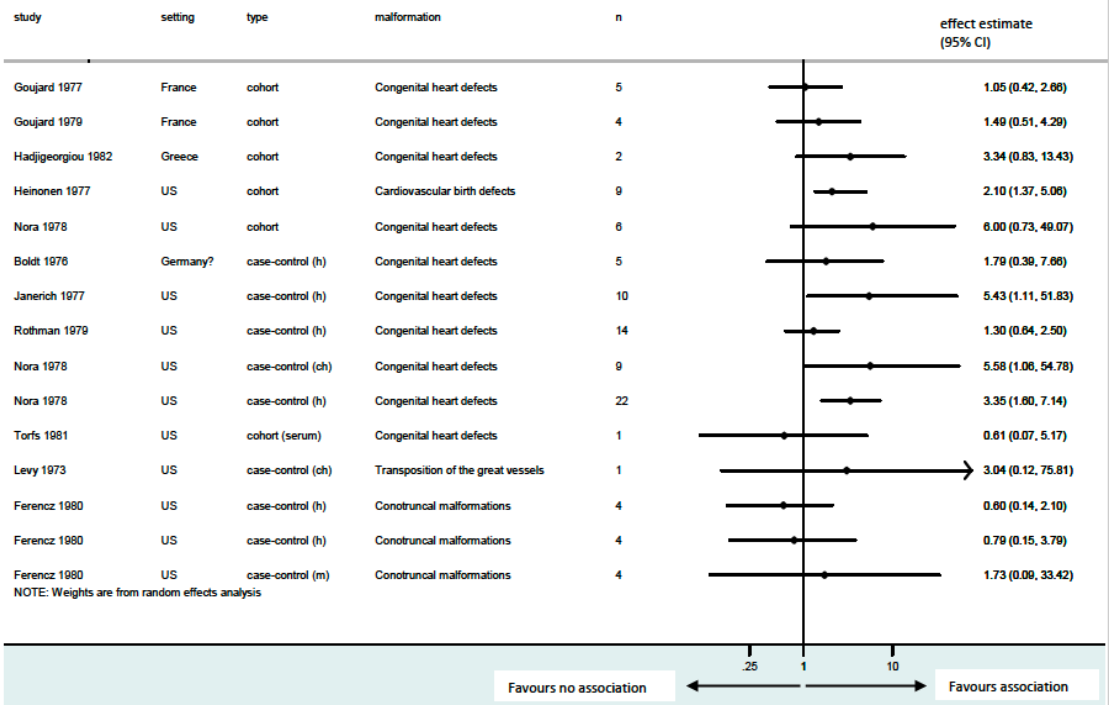
On 6 February 2019, Marie Lyon received a response indicating that her request would be forwarded to the appropriate department. A copy of the FOI request was sent to the Independent Medicines and Medical Devices Safety Review (IMMDS) on 20 February 2019. On 28 February 2019, Lord Alton raised the question of the availability of the raw data in the House of Lords. [6] On 5 March 2019, the IMMDS contacted Marie Lyon to establish whether the MHRA had responded to her request. She confirmed that the requested data had not yet been received.

On 6 March 2019, she sent a further enquiry: "*Would you please let me know if the raw data I requested has been actioned yet?*"

The MHRA responded on 8 March 2019, but the response did not contain the attachment holding the requested raw data. Marie Lyon requested the missing attachment that same day. The underlying data were finally received on 11 March 2019, approximately six weeks after the original FOI request. Their release made it possible to undertake an independent quantitative analysis of the data extracted by the EWG.

The Statistical Question Raised by the EWG's Forest Plots

Figure 2. Forest plot and quality assessment of epidemiological studies of HPTs and heart defects



A particular concern arose from Figure 2 of the original EWG report, which presented a forest plot relating to congenital heart defects. [7] A footnote accompanying the figure stated: "*Weights are from random effects analysis.*" However, the published report did not present the corresponding study weights or pooled estimates from a random-effects meta-analysis.

This distinction is important. A forest plot can display the effect estimates and confidence intervals from individual studies without providing a combined estimate of the overall association. A random-effects meta-analysis, by contrast, provides a pooled estimate while allowing for variation in the underlying effects between studies. [8]

The reference to random-effects weighting therefore raised a methodological question: why were random-effects methods referenced in the EWG's forest plot when the report did not present pooled results? It also raised a broader question about whether the EWG's decision not to pool the evidence was justified by the statistical characteristics of the available studies.

Purpose of the Subsequent Meta-analysis

Following the release of the underlying data on 11 March 2019, we undertook a random-effects meta-analysis using the data extracted by the EWG. The purpose was to examine the associations between exposure to oral hormone pregnancy tests and congenital malformations using the evidence assembled for the EWG's own review.

The analysis considered congenital heart defects, congenital malformations overall, urogenital defects and other reported congenital abnormalities. This approach enabled the EWG's conclusions to be considered alongside pooled statistical estimates derived from its underlying evidence.

The central question for this report is whether the EWG's decision not to combine the historical evidence materially affected the interpretation of the association between oral HPT exposure and congenital malformations. The following sections examine the results of that

meta-analysis, the extent of statistical heterogeneity between studies, and the implications for interpreting the EWG's 2017 conclusions. [9]

Methods

Study Design and Data Sources

We undertook a secondary analysis of the epidemiological data extracted by the Commission on Human Medicines' Expert Working Group (EWG) for its 2017 review of oral hormone pregnancy tests (HPTs), including Primodos, and congenital malformations. The underlying numerical data were obtained from the Medicines and Healthcare products Regulatory Agency (MHRA) following a Freedom of Information request submitted in January 2019. The data were received on 11 March 2019 and provided the basis for calculating pooled associations from the study data extracted by the EWG.

The analysis used EWG-extracted study data rather than conducting a new systematic literature search.

Four categories of congenital abnormalities were examined:

- Congenital heart defects.
- Any congenital malformation.
- Urogenital malformations.
- Other congenital defects.

The analyses included cohort and case-control studies for which the necessary numerical data were available. Studies reporting effect estimates without sufficient underlying event data for the planned analyses were excluded from the pooled calculations. These included Lammer (1986), Sainz (1987) and Tümmler (2014).

Statistical Analysis

Random-effects meta-analyses were undertaken to estimate the association between exposure to oral HPTs and each congenital malformation outcome. Associations were expressed as pooled odds ratios (ORs) with 95% confidence intervals (CIs). Statistical heterogeneity was assessed using the I^2 statistic. Where relevant, cohort and case-control studies were examined separately, with overall pooled estimates calculated for the principal outcomes.

The interpretation considered the magnitude and precision of the estimated associations, statistical heterogeneity and the clinical and methodological comparability of the included studies.

Assessment of Methodological Limitations

The EWG's original concerns about study quality, heterogeneity and statistical weighting were critically examined. Particular attention was given to the reliability of exposure classification, selection of comparison groups, adjustment for confounding, differences in study design and outcome definitions.

Statistical heterogeneity was considered alongside clinical and methodological heterogeneity, recognising that low I^2 values do not necessarily establish that studies are comparable or free from bias. The implications of statistical weighting were also considered, including the

distinction between the precision of an effect estimate and the methodological validity of the underlying study.

Comparison with the Published Meta-analysis

The pooled estimates derived from EWG-extracted data were compared with those reported in the systematic review and meta-analysis by Heneghan et al. (2018). Comparisons examined the direction and magnitude of associations, 95% confidence intervals and measures of statistical heterogeneity for congenital heart defects, congenital malformations overall and genital or urogenital abnormalities.

The comparison was intended to assess the reproducibility and consistency of the principal findings. As the two analyses drew substantially on overlapping epidemiological evidence, agreement between their estimates was interpreted as analytical reproducibility rather than independent confirmation.

Assessment of the EWG's Methodological Decisions

The EWG's stated reasons for declining formal meta-analysis were examined against established systematic review methodology, including approaches described in the Cochrane Handbook. The assessment considered whether concerns about study quality, heterogeneity and statistical weighting could be investigated through quantitative synthesis and appropriate methodological appraisal. The EWG's published forest plots, quality assessments and accompanying explanations were also examined to assess the transparency and reproducibility of its analytical approach.

Results

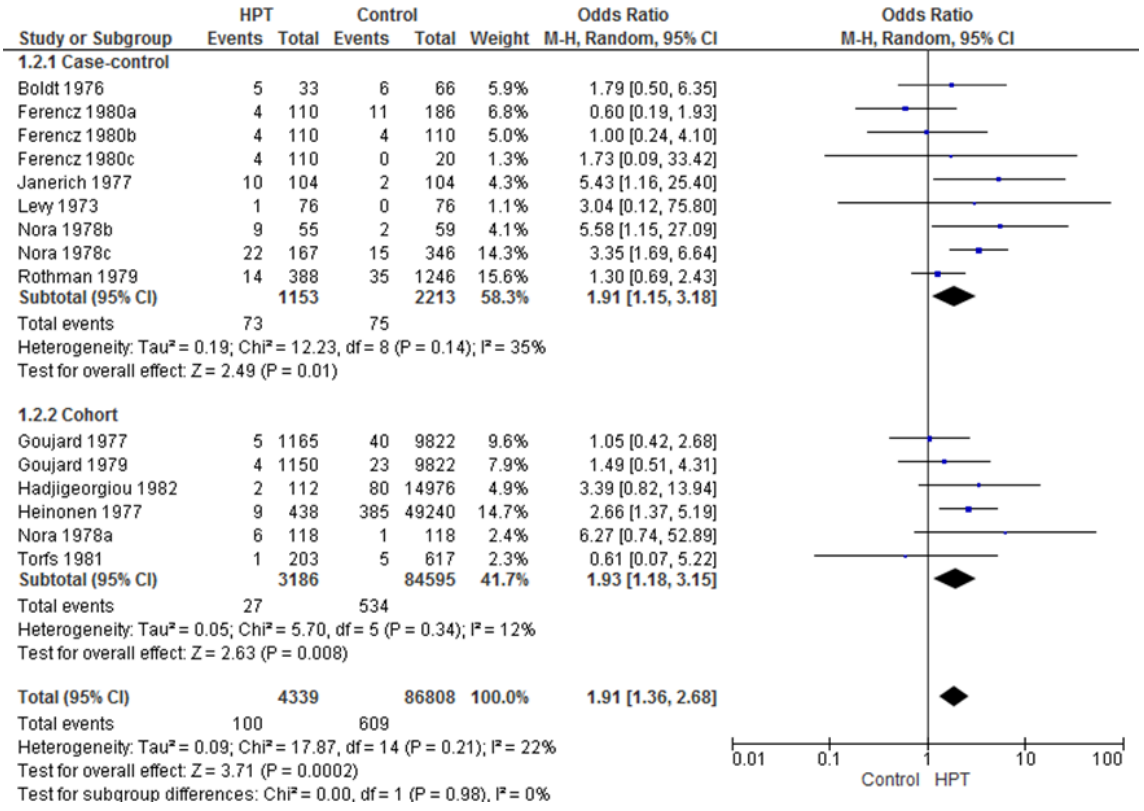
Summary Table 1. Random-effects meta-analysis of data extracted by the EWG

Congenital abnormality	Odds ratio (95% CI)	I^2	Statistical significance
Congenital heart defects	1.91 (1.36–2.68)	22%	P = 0.0002
Any congenital malformation	1.34 (1.13–1.60)	0%	P = 0.0008
Urogenital defects	2.22 (0.82–6.02)	0%	P = 0.12
Other congenital defects	3.62 (1.11–11.82)	76%	P = 0.03

The pooled analyses identified statistically significant associations for congenital heart defects, any congenital malformation and the category of other congenital defects (see Table 1). The association with urogenital defects was not statistically significant. However, substantial heterogeneity was observed in the category of other congenital defects, limiting the interpretability of that pooled estimate.

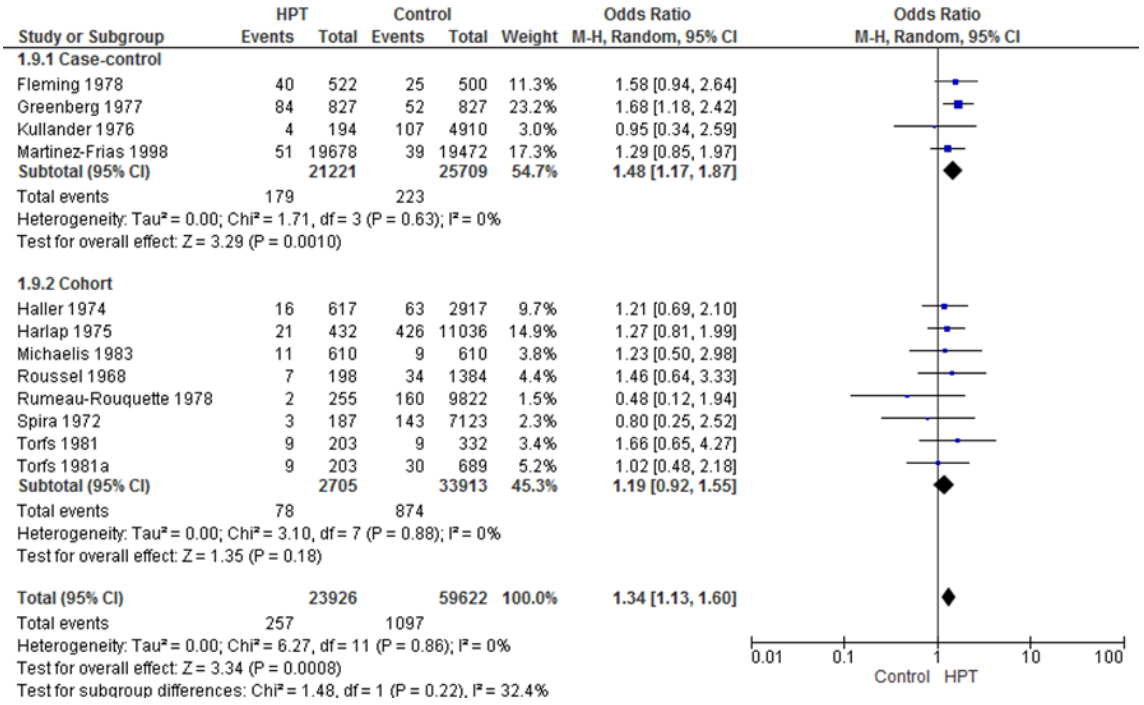
Congenital Heart Defects

The meta-analysis included data from nine case-control studies and six cohort studies examining congenital heart defects. The pooled analysis identified a statistically significant association between exposure to oral HPTs and congenital heart defects: OR = 1.91 (95% CI 1.36–2.68; P = 0.0002). Statistical heterogeneity was relatively low (I^2 = 22%), suggesting that the variation between study results was limited.



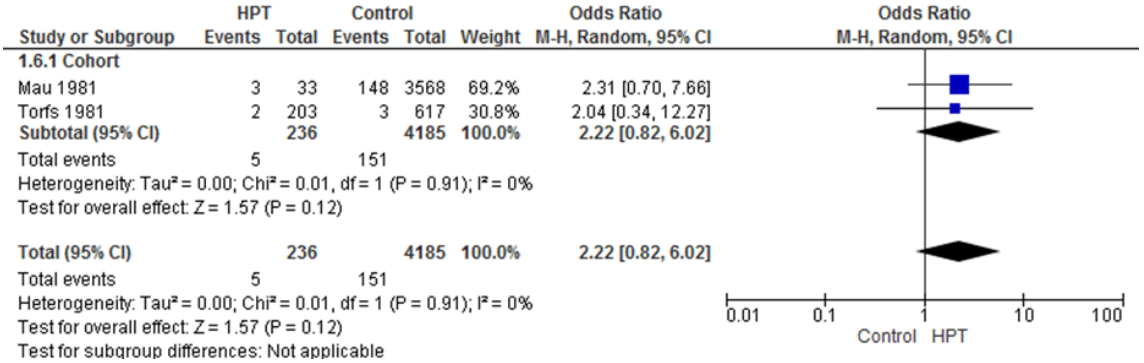
Any Congenital Malformation

The analysis of any congenital malformation included four case-control studies and eight cohort studies. The pooled estimate demonstrated a statistically significant association: OR = 1.34 (95% CI 1.13–1.60; P = 0.0008). No statistical heterogeneity was detected between the included studies (I^2 = 0%).



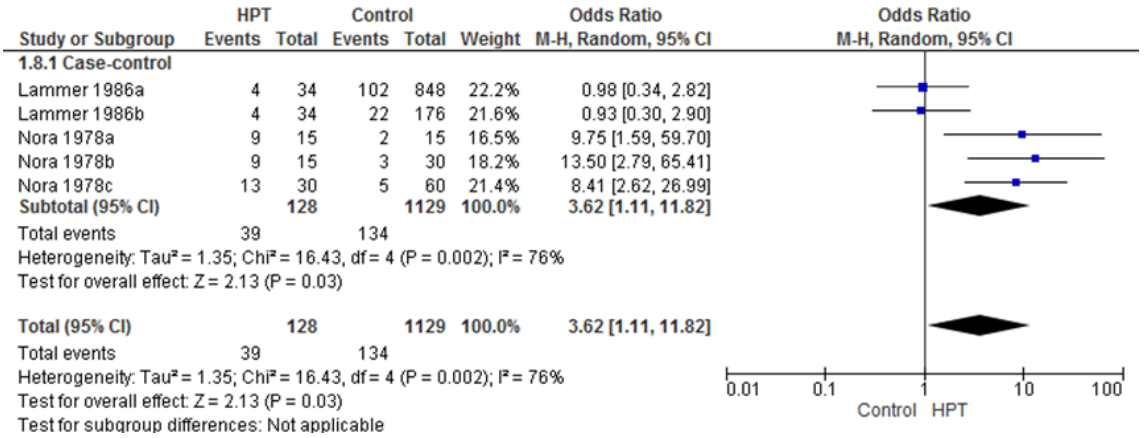
Urogenital Defects

The analysis of urogenital defects included two cohort studies. The pooled estimate was: OR = 2.22 (95% CI 0.82–6.02; P = 0.12). Although the point estimate suggested higher odds of urogenital defects following exposure to oral HPTs, the association was not statistically significant. No statistical heterogeneity was detected (I^2 = 0%).



Other Congenital Defects

Five case-control studies contributed to the analysis of congenital abnormalities classified as "other defects". The pooled estimate was: OR = 3.62 (95% CI 1.11–11.82; P = 0.03). Although the pooled estimate was statistically significant, substantial heterogeneity was observed between studies (I^2 = 76%). This high level of heterogeneity suggests that the studies may have been examining different outcomes or estimating substantially different underlying effects. The broad classification of "other defects" may have combined congenital abnormalities that were not sufficiently comparable to justify a single pooled estimate.



Despite a statistically significant pooled estimate, the result should not be interpreted as reliable evidence of a single overall association given the substantial heterogeneity.

Studies Excluded Because of Unavailable Event Data

The EWG report presented odds ratios and confidence intervals for outcomes reported in three additional studies:

- Lammer (1986)
- Sainz (1987)
- Tümmler (2014)

However, the underlying event-rate data required for their inclusion in the subsequent meta-analysis were not reported. These studies were therefore excluded from the pooled calculations. Excluded outcomes included nervous system abnormalities, orofacial clefts, digestive and abdominal wall defects, urinary system abnormalities, and limb defects. This limitation is relevant because the pooled estimates could not incorporate all the outcomes for which the EWG had reported individual effect estimates.

Summary of the Results

The random-effects meta-analysis of data extracted by the EWG identified statistically significant associations between oral HPT exposure and both congenital heart defects and congenital malformations overall. The association with congenital heart defects was particularly notable, with a pooled odds ratio of 1.91 and relatively low statistical heterogeneity. For any congenital malformation, the pooled odds ratio was 1.34, with no statistical heterogeneity detected. The analysis of urogenital defects did not demonstrate a statistically significant association, whereas the pooled estimate for other congenital defects was substantially heterogeneous and considered unsuitable for combination.

Discussion

Overview of the Principal Findings

The principal finding of the subsequent reanalysis of the Expert Working Group's (EWG) epidemiological data is that statistically significant associations were identified between exposure to oral hormone pregnancy tests (HPTs), including Primodos, and both congenital heart defects and congenital malformations overall.

The random-effects meta-analysis undertaken by Heneghan and Aronson reported an odds ratio (OR) of 1.91 (95% confidence interval [CI] 1.36–2.68) for congenital heart defects and an

OR of 1.34 (95% CI 1.13–1.60) for any congenital malformation. Statistical heterogeneity was relatively low for congenital heart defects (I^2 = 22%), while no statistical heterogeneity was detected for congenital malformations overall (I^2 = 0%).

These findings are important because the EWG reviewed the underlying epidemiological evidence in 2017 but concluded that a formal meta-analysis was inappropriate. The EWG's overall conclusion was that the available evidence did not, on balance, support an association between exposure to HPTs during early pregnancy and congenital anomalies.

The subsequent quantitative findings therefore raise a fundamental methodological question: was the EWG justified in rejecting meta-analysis, or could the limitations it identified have been investigated using established systematic review methods? This discussion examines the EWG's stated reasons for rejecting quantitative synthesis, evaluates those reasons against established methodological principles, and considers the implications of the subsequent meta-analysis for interpreting the EWG's conclusions.

The EWG's Reasons for Rejecting Meta-analysis

Section 5.3.4.2 of the EWG's final report, *Development of a quality scoring system*, provides the principal explanation for its decision not to undertake formal meta-analysis.
The EWG advanced three main reasons:

1. The historical studies were considered insufficiently robust.
2. The studies were considered too heterogeneous in their design.
3. Conventional statistical weighting was considered inappropriate because larger studies could receive greater influence despite substantial methodological limitations.

Rather than undertaking a pooled quantitative analysis, the EWG developed a bespoke quality assessment system based on seven criteria. These examined the appropriateness of control groups, the reliability and specificity of exposure measurement, the timing of exposure, adjustment for confounding, sample size, and the consideration of multiple statistical comparisons. Individual studies were assessed using a green, amber, or red classification to indicate good, moderate, or poor quality for each criterion.

The EWG did not calculate an overall numerical quality score because it considered that assigning relative weights to the criteria would introduce additional subjectivity. The EWG's concerns reflected recognised limitations of historical observational research. However, the existence of these limitations does not necessarily preclude quantitative synthesis.[10] Established systematic review methodology provides approaches for investigating the influence of study quality, heterogeneity and weighting on pooled results. The methodological issue is whether those limitations justified rejecting meta-analysis altogether rather than investigating their influence through appropriate analytical methods.

Methodological Robustness and Sensitivity Analysis

The EWG's first justification was that many historical studies were insufficiently robust to support reliable quantitative synthesis. The limitations identified in the EWG report and Annex 26 included inappropriate or poorly described control groups, uncertainty regarding exposure to specific hormonal preparations, incomplete information on gestational timing, recall bias, and inadequate adjustment for confounding factors.

These concerns are relevant to the reliability of the epidemiological evidence; however, methodological weaknesses are frequently encountered in systematic reviews of observational studies. Their presence does not automatically mean that meta-analysis is inappropriate.
A recognised approach is to assess the risk of bias in individual studies and investigate its influence through sensitivity analyses. [11] For example, pooled estimates may be recalculated

after excluding studies considered to be at higher risk of bias. The resulting estimates can then be compared with those obtained from the complete dataset. This enables investigators to assess whether an observed association depends primarily on studies with substantial methodological limitations.

Other approaches include restricting analyses to studies with clearly defined HPT exposure, examining cohort and case-control studies separately, and investigating whether differences in exposure timing or outcome definitions influence the findings. These approaches are relevant to the systematic review by Heneghan et al., which used explicit eligibility criteria and methodological assessment to examine the historical evidence.

The EWG's concerns about study quality could therefore have been investigated within a quantitative synthesis framework. Sensitivity analysis cannot eliminate bias from the underlying studies, and excluding studies does not guarantee that the remaining evidence is reliable. Nevertheless, it provides a transparent means of assessing how methodological decisions influence pooled estimates.

The central criticism is that the EWG treated limitations in study quality as grounds for not conducting a meta-analysis, rather than demonstrating whether those limitations materially affected the quantitative synthesis.

Heterogeneity Between the Historical Studies

The EWG's second justification was that the historical studies were too heterogeneous in design to be combined. Heterogeneity is an important consideration in systematic reviews, but it is necessary to distinguish between statistical, clinical and methodological heterogeneity. Statistical heterogeneity refers to variation in the observed effect estimates beyond that expected from sampling error. Clinical heterogeneity concerns differences in populations, exposures and outcomes, while methodological heterogeneity concerns differences in study design, measurement and risk of bias.

Meta-analysis provides methods for estimating statistical heterogeneity and investigating potential sources of variation. The I^2 statistic is commonly used to describe the proportion of observed variability attributable to between-study heterogeneity rather than sampling error. A random-effects model allows for variation in the underlying effects between studies.

The subsequent meta-analysis of EWG-extracted data reported $I^2 = 22\%$ for congenital heart defects and $I^2 = 0\%$ for any congenital malformation. These results indicate that relatively little statistical heterogeneity was detected for the two principal outcomes.

This finding is particularly relevant because it demonstrates that substantial statistical inconsistency was not observed when the historical data were combined. However, low statistical heterogeneity does not establish that the studies were methodologically comparable or free from bias. An I^2 value of 0% may occur even when studies differ in important respects, particularly when the number of studies is limited or estimates are imprecise. Furthermore, combining studies with fundamentally different exposures or outcomes may produce a statistically precise estimate that lacks a clear clinical interpretation.

Consequently, the subsequent findings do not eliminate the EWG's concerns about methodological heterogeneity. They do, however, challenge any suggestion that statistical heterogeneity itself made pooling inappropriate.

The appropriate methodological question was whether the studies were sufficiently comparable to estimate a meaningful pooled association, and whether the resulting estimate was robust to differences in study design and exposure classification. These questions could have been

investigated through subgroup analyses, sensitivity analyses and explicit assessment of clinical and methodological comparability.

The EWG's Concerns About Statistical Weighting

The EWG's third justification concerned the statistical weighting used in conventional meta-analysis. The EWG argued that larger studies could receive greater statistical weight despite substantial methodological weaknesses. This concern reflects an important distinction between statistical precision and methodological validity.

Conventional inverse-variance meta-analysis generally gives greater weight to studies with more precise effect estimates. However, greater precision does not necessarily indicate that a study is less vulnerable to confounding, selection bias or exposure misclassification. Nevertheless, established meta-analytical methods provide approaches for examining this issue.

Random-effects meta-analysis incorporates an estimate of between-study variance into the weighting process. When between-study variance is appreciable, this can reduce the relative dominance of the largest studies compared with a conventional fixed-effect inverse-variance model. However, random-effects modelling does not automatically correct for methodological bias or guarantee that lower-quality studies receive less weight.
Where between-study variance is estimated to be zero, random-effects and fixed-effect inverse-variance weights may be identical. This distinction is particularly relevant to the analysis of congenital malformations overall, for which I^2 was reported as 0%. Consequently, random-effects modelling should not be regarded as a complete solution to the EWG's concerns about weighting. Nevertheless, those concerns could have been investigated through sensitivity analyses, subgroup analyses and examination of the influence of individual studies on the pooled estimates.

The EWG's rejection of the meta-analysis was not the only available methodological response. A more informative approach would have been to investigate how alternative analytical assumptions and study-selection decisions affected the pooled estimates, rather than treating conventional weighting as sufficient grounds for excluding quantitative synthesis.

The Significance of the EWG's Forest Plots

An additional methodological question arises from the forest plots presented in the EWG report. Section 5.3.4.3 of the final report explains that summary statistics were presented graphically and that effect estimates were calculated from proportions where necessary. The forest plots were intended to assist interpretation of the individual studies alongside the EWG's quality assessments.

However, the subsequent analysis identified a footnote in Figure 2 of the EWG report, relating to congenital heart defects, stating: "Weights are from random effects analysis."

Despite this reference, the EWG report did not present corresponding pooled estimates. This raises a question about the relationship between the statistical calculations used to construct the forest plots and the EWG's decision not to undertake formal meta-analysis.
The presence of random-effects weighting does not establish that a complete pooled analysis had been undertaken and withheld. Nevertheless, it indicates that random-effects methods were referenced in the presentation of the evidence, making clarification of the analytical procedures particularly important. If study weights were calculated using a random-effects approach, readers should be able to establish which statistical model was used, how between-study variance was estimated and why pooled estimates were not presented. The

absence of this information limited the transparency and reproducibility of the EWG's published analysis.

Evidence from Annex 26

Annex 26 provides detailed assessments of the epidemiological studies considered by the EWG. [12] These assessments demonstrate that the EWG's concerns about study quality were based on identifiable methodological problems rather than unsupported assumptions.
For example, the Annex describes differences between studies examining congenital heart defects, including variations in exposure classification, control selection, adjustment for confounding and timing of exposure. Some studies reported statistically significant associations, while others did not. They also demonstrate why quantitative synthesis should be accompanied by careful assessment of study comparability and risk of bias. Importantly, the Annex documents the individual findings and methodological limitations that could have informed subgroup and sensitivity analyses.
Its detailed assessments therefore provide a potential foundation for investigating whether the associations observed in the subsequent meta-analysis remain robust when studies are classified according to methodological characteristics.

Comparison of the Unpublished Draft and Final EWG Report

A comparison of the unpublished draft and the final published versions of the EWG report provides further context for understanding the decision not to conduct a meta-analysis. Both versions contain substantively the same justification for rejecting formal quantitative synthesis. The EWG's concerns about study robustness, heterogeneous designs and statistical weighting were therefore retained in the final report. However, changes were made to the presentation of the epidemiological findings and the language used to express the overall conclusions. In particular, the draft described limited evidence for no association with overall congenital anomalies, whereas the final report stated that the evidence reviewed did not support an association, while acknowledging limitations in evidence quality.

The final report also made it explicit that the available studies could not establish either a causal association or its absence. Furthermore, the overall conclusion included the qualification "on balance" when stating that the available evidence did not support a causal association.

These changes are relevant because they affect how uncertainty is communicated. They do not, however, alter the underlying methodological decision not to pool the historical evidence. The comparison therefore indicates that the EWG's rejection of meta-analysis remained a consistent feature of the review across the examined versions. It does not establish that pooled results were deliberately suppressed or that the EWG had completed and withheld a formal meta-analysis.

Implications of the Subsequent Meta-analysis and limitations

The subsequent random-effects meta-analysis demonstrated that quantitative synthesis of the EWG's extracted data was feasible. It identified statistically significant pooled associations for congenital heart defects and congenital malformations overall. The pooled odds ratio for congenital heart defects was 1.91 (95% CI 1.36–2.68), while the corresponding estimate for any congenital malformation was 1.34 (95% CI 1.13–1.60).

These findings are relevant because they provide quantitative estimates absent from the EWG's published synthesis, and the analysis demonstrates that statistically significant associations could be identified in the epidemiological evidence assembled for the EWG's own review. The remaining question is whether these associations are sufficiently robust to methodological limitations to influence the overall interpretation of the evidence.

The reanalysis was dependent on the completeness and accuracy of the numerical data extracted by the EWG. Studies lacking sufficient event data could not be incorporated into the pooled estimates. The underlying observational studies were subject to potential selection bias, exposure misclassification, confounding and differences in study design and outcome definitions. Random-effects modelling accounts for variation in underlying effects between studies but does not eliminate methodological bias. Accordingly, the pooled estimates were interpreted in the context of the quality and comparability of the underlying evidence.

Overall Interpretation

The EWG's three reasons for rejecting meta-analysis identify recognised challenges in evidence synthesis. However, established systematic review methodology provides approaches for investigating each of these challenges.

Study quality can be examined through risk-of-bias assessment and sensitivity analyses. Statistical heterogeneity can be quantified, while clinical and methodological heterogeneity can be investigated through explicit comparability assessments and subgroup analyses. The influence of statistical weighting can be examined through alternative models, influence analyses and sensitivity analyses.

These methods do not guarantee that pooling will be appropriate, nor do they eliminate the limitations of historical observational evidence. They do, however, provide a structured means of determining whether quantitative synthesis is informative and whether its findings are robust. The subsequent meta-analysis demonstrates that pooled estimates can be calculated from EWG-extracted data and that statistically significant associations were identified for two principal congenital anomaly outcomes. The central methodological criticism is that the EWG rejected quantitative synthesis without demonstrating, through appropriate analytical investigation, that those limitations made pooling uninformative or misleading.

The EWG's qualitative approach provided a detailed assessment of individual study limitations but did not quantify the overall associations identified by the subsequent reanalysis. A more comprehensive approach would have presented the pooled estimates alongside explicit assessments of risk of bias, clinical comparability, statistical heterogeneity and sensitivity to analytical assumptions. Such an approach would have enabled readers to evaluate both the observed associations and the uncertainty surrounding their interpretation.

The findings of the subsequent meta-analysis therefore raise substantive questions about whether the EWG's decision not to pool the historical evidence was adequately justified and whether its published synthesis provided the most complete and transparent account of the available epidemiological evidence.

A comparison of the re-analysis of the EWG data [1] and Heneghan et al. [2] meta-analyses demonstrates a striking consistency in the magnitude and direction of the reported associations. The close agreement between the effect estimates, confidence intervals and measures of heterogeneity provides additional reassurance regarding the robustness and reproducibility of the findings.

Table 2. Comparison of the 2019 re-analysis of EWG-extracted data and Heneghan et al.

Outcome	EWG re-analysis [1]	Heneghan et al. [2]
Congenital heart defects	OR 1.91 (95% CI 1.36–2.68); I^2 = 22%; P = 0.0002	OR 1.89 (95% CI 1.32–2.72); I^2 = 0%; P = 0.0006
Any congenital malformation	OR 1.34 (95% CI 1.13–1.60); I^2 = 0%; P = 0.0008	OR 1.40 (95% CI 1.18–1.66); I^2 = 0%; P < 0.0001
Genital/urogenital malformations	OR 2.22 (95% CI 0.82–6.02); I^2 = 0%; P = 0.12	OR 2.63 (95% CI 0.84–8.28); I^2 = 0%; P = 0.10

The findings are particularly compelling for congenital heart defects, where the pooled odds ratios are almost identical (1.91 versus 1.89), and for overall congenital malformations (1.34 versus 1.40). Both analyses demonstrate statistically significant associations for these outcomes, with low or no observed heterogeneity. For genital/urogenital malformations, both analyses similarly suggest an increased risk, although the confidence intervals are wide and include the null value. These findings therefore remain statistically inconclusive.

Taken together, the close agreement between the EWG and Heneghan et al. estimates strengthens confidence in the consistency of the observed associations and in the reproducibility of the findings across the two analytical approaches. Because the analyses draw substantially on the same evidence base, their agreement is best interpreted as demonstrating analytical reproducibility rather than independent confirmation.

Conclusion

The reanalysis of data extracted by the 2017 Expert Working Group demonstrates that quantitative synthesis of the historical epidemiological evidence was feasible. Statistically significant associations were identified between exposure to oral hormone pregnancy tests and congenital heart defects and congenital malformations overall, with low observed statistical heterogeneity.

The close agreement with the findings of Heneghan et al. strengthens confidence in the analytical reproducibility of these associations.

The EWG's concerns regarding study quality, heterogeneity and statistical weighting were legitimate methodological considerations that could have been investigated through established systematic review methods, including risk-of-bias assessment and sensitivity analyses.

These findings challenge the EWG's decision to reject quantitative synthesis and demonstrate the importance of transparent, reproducible analysis in evaluating evidence and uncertainty.

References

1. Report of the Commission on Human Medicines' Expert Working Group on Hormone Pregnancy Tests. GOV.UK. (accessed 24 May 2018). https://www.gov.uk/government/publications/report-of-the-commission-on-human-medicines-expert-working-group-on-hormone-pregnancy-tests
2. Report of the Commission on Human Medicines' Expert Working Group on Hormone Pregnancy Tests. Page 63, Section 5.3.4.2. Development of a quality scoring system. https://www.gov.uk/government/publications/report-of-the-commission-on-human-medicines-expert-working-group-on-hormone-pregnancy-tests
3. Heneghan C, et al. Oral hormone pregnancy tests and the risks of congenital malformations: a systematic review and meta-analysis. F1000Res. 2018 Oct 31;7:1725. doi: 10.12688/f1000research.16758.2.
4. Evaluation of systematic review and meta-analysis of studies on oral hormone pregnancy tests, including Primodos. March 2019 Meeting. https://assets.publishing.service.gov.uk/media/5ccc4d2340f0b63325dd0f01/CHM-Ad-Hoc-Expert-Group-paper_Heneghan-meta-analysis_March-2019_redacted.pdf
5. Update to the association between Oral Hormone Pregnancy Tests, including Primodos, and congenital anomalies. CEBM, University of Oxford Reports 2019. 8 March 2019. https://www.cebm.ox.ac.uk/resources/reports/update-to-the-association-between-oral-hormone-pregnancy-tests-including-primodos-and-congenital-anomalies
6. Safety of Medicines and Medical Devices - Motion to Take Note https://www.theyworkforyou.com/lords/?id=2019-02-28a.345.0
7. Report of the Commission on Human Medicines' Expert Working Group on Hormone Pregnancy Tests. Page 71, Section 5.4 Overall conclusion on the evidence for congenital anomalies being associated with exposure to HPTs in pregnancy https://www.gov.uk/government/publications/report-of-the-commission-on-human-medicines-expert-working-group-on-hormone-pregnancy-tests
8. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7(3):177-88. doi: 10.1016/0197-2456(86)90046-2.
9. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003 Sep 6;327(7414):557-60. doi: 10.1136/bmj.327.7414.557.
10. Sterne JA, et al.ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919. doi: 10.1136/bmj.i4919.
11. Cochrane Handbook for Systematic Reviews of Interventions. Higgins JPT, Thomas J, Chandler J, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.5, 2024. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current
12. Expert Working Group on Hormone Pregnancy Test - Annexes to the report. https://www.gov.uk/government/publications/report-of-the-commission-on-human-medicines-expert-working-group-on-hormone-pregnancy-tests

Competing Interests

CH has held grant funding from the NIHR, the NIHR School of Primary Care Research, the NIHR BRC Oxford and previously the World Health Organization for a series of Living rapid reviews on the modes of transmission of SARS-CoV-2 (reference WHO registration No2020/1077093. He has received expenses and fees for his media work. He receives expenses for teaching EBM and is also paid for his GP work in NHS out-of-hours (contract with Oxford Health NHS Foundation Trust). He has also received income from publishing toolkit books and appraising treatment recommendations in non-NHS settings.

TJ was in receipt of a Cochrane Methods Innovations Fund grant to develop guidance on the use of regulatory data in Cochrane reviews (2015–018). In 2014–2016, he was a member of three advisory boards for Boehringer Ingelheim, and a member of an independent data monitoring committee for a Sanofi Pasteur clinical trial on an influenza vaccine. TJ has been a relator in a False Claims Act lawsuit on behalf of the United States that involves sales of Tamiflu for pandemic stockpiling. If resolved in the United States' favour, he would be entitled to a percentage of the recovery. TJ was a co-holder of a Laura and John Arnold Foundation grant for the development of a RIAT support centre (2017–2020) and a Jean Monnet Network Grant, 2017–2020, for The Jean Monnet Health Law and Policy Network. TJ was the consultant scientific coordinator for the HTA Medical Technology programme of the Agenzia per i Servizi Sanitari Nazionali (AGENAS) of the Italian MoH (2007–2019). TJ was funded by NIHR UK and the World Health Organization (WHO) to update Cochrane Review A122, Physical Interventions to interrupt the spread of respiratory viruses. TJ has received fees for articles published by The Spectator and other media outlets. TJ is part of a review group carrying out a living rapid literature review on the modes of transmission of SARS-CoV-2 (WHO Registration 2020/1077093–0). He is a member of the WHO COVID-19 Infection Prevention and Control Research Working Group, for which he receives no funds.