Does antiretroviral therapy cause congenital malformations?

A systematic review and meta-analysis.

Fekadu Mazengia Alemu, MPH; Alemayehu Worku Yalew, PhD.

* Corresponding author

Fekadu Mazengia Alemu *

School of Public Health, College of Health Sciences, Addis Ababa University
E-mail: mazengiafek23@gmail.com (FM)
Telephone: +211927000315, +251924542858
P.O. Box 100796
Addis Ababa, Ethiopia

ORCID ID: 0000-0002-7332-7309

Alemayhu Worku Yalew

School of Public Health, College of Health Sciences, Addis Ababa University
E-mail: alemayhuwy@yahoo.com
Telephone: +251911405652
P.O. Box 9086
Addis Ababa, Ethiopia

1. School of Public Health, College of Health Sciences, Addis Ababa University.

† Both authors contributed equally.
INTRODUCTION

Human immunodeficiency virus (HIV) is a one of the major indirect causes of maternal and neonatal morbidity and mortality, particularly in Sub-Saharan African countries [1]. Preventing mother-to-child transmission of HIV (PMTCT) has been one of the greatest successes in the prevention efforts to combat the HIV epidemic [2]. Largely, the reduction in transmission of HIV has been sharply reduced due to the use of PMTCT approaches combined with strategic use of highly active antiretroviral therapy (HAART) [3-5]. The benefits of antiretroviral therapy (ART) use during pregnancy considerably outweigh the potential risks. However, immediate and long-term harm in mothers and children, including congenital anomalies, remains a serious issue of concern [3, 6, 7].

A congenital anomaly is defined as the occurrence of a major structural or chromosomal defect or any group of two or more minor defects occurring in a newborn baby after 20 weeks of gestational age [8]. In under-resourced countries, various risk factors such as malnutrition, anemia, multi-parity, teenage pregnancy, limited healthcare access, co-morbidities such as malaria, and irregular use of ART may increase the risk of birth defects. However, the role of such interactions in the development of birth defects has not been well studied [9].

A cohort study revealed that anemia (84%), neutropenia (64%), and thrombocytosis (24%) were the most common disorders among children exposed to ART. Long-term follow-up also revealed various neurological, cardiac, and ophthalmological pathologies [10].

The most commonly reported birth defects were found in the genital and urinary system (30.6%), the cardiovascular system (27.4%), the musculoskeletal system (12.9%), and the digestive system (9.7%) [11]. In zidovudine-exposed but HIV-uninfected infants, transient anemia, long-term hematological anomalies (neutropenia, thrombocytopenia, and lymphopenia), and hyperlactatemia have been reported. Pre-clinical studies showed that zidovudine had a carcinogenic effect [12]. Literature Research in the literature has also documented the timing of initiation and type of ART regimen as some of the factors associated with adverse pregnancy outcomes [13].
For safety reasons, it is essential to monitor pregnancy outcomes in resource-limited settings during the design and implementation of the World Health Organization (WHO) ART guidelines [14]. The time of the highest sensitivity to teratogenic exposures is the stage of organogenesis, 18–60 days after conception or 4–13 weeks after the beginning of the last menstrual period [8]. However, congenital defects due to teratogenic exposure can happen any time through the course of pregnancy [15].

New drugs are coming to the market and effective ART is available, yet toxicity issues and harmful outcomes during pregnancy, including major congenital anomalies, preterm delivery, anemia, and low birth weight, are of increasing concern among clinicians and program managers [7, 13, 16-18]. There is a counter-argument that ART does not have an effect on the formation of congenital anomalies [3-5]. The main aim of this review, therefore, was to investigate the effect of ART on congenital anomalies and to come up with a summary estimate.

METHODS AND MATERIALS

The study designs reviewed include interventional studies (randomized controlled trials and controlled clinical trials). As Chou et al. recommended gathering evidence on harms from a broad range of sources including observational studies, notably when clinical trials are lacking; observational studies (cohort and case-control studies) were also included.

Search Strategy, and Keywords

First, the DARE database (http://www.library.ucsf.edu) was explored in an attempt to confirm whether systematic reviews, meta-analyses, or ongoing projects related to this topic existed. Studies were selected and compared following the guidelines outlined in the Michael Hewitt Trent and Cochrane Reviewers’ Handbooks [19, 20]. Following the implementation of the search strategy, the titles of all appropriate abstracts and titles collected from electronic and hand-searches were entered into EndNote version 6 reference software.

The main sources for the review were electronic bibliographic databases. The MEDLINE/PubMed and EMBASE databases, covering most areas of health care and containing index journals published from around the world, were searched. Further, Web of Science, Scopus, and CINAHL were also searched. AIDSLINE, which focuses on specific
areas of health, was also examined. The Cochrane Collaboration, an electronic database for reports of controlled trials ("CENTRAL"), and search engines such as Google Scholar were searched specifically for gray literature.

The following search terms were used as keywords and/or Mesh MeSH terms:

(((Antiretroviral[All Fields] AND ("therapy"[Subheading] OR "therapy"[All Fields] OR "therapeutics"[MeSH Terms] OR "therapeutics"[All Fields])) AND ("congenital abnormalities"[MeSH Terms] OR ("congenital"[All Fields AND "abnormalities"[All Fields]) OR "congenital abnormalities"[All Fields] OR ("congenital"[All Fields AND "anomalies"[All Fields]) OR "congenital anomalies"[All Fields]) OR "congenital abnormalities"[All Fields])) AND ("congenital abnormalities"[MeSH Terms] OR ("congenital"[All Fields] AND "abnormalities"[All Fields]) OR "congenital abnormalities"[All Fields]) AND ("congenital abnormalities"[MeSH Terms] OR ("congenital"[All Fields AND "anomalies"[All Fields]) OR "congenital abnormalities"[All Fields] OR ("congenital"[All Fields AND "malformations"[All Fields]) OR "congenital malformations"[All Fields])) AND ("congenital abnormalities"[MeSH Terms] OR ("congenital"[All Fields] AND "abnormalities"[All Fields]) OR "congenital abnormalities"[All Fields] OR ("birth"[All Fields] AND "defect"[All Fields]) OR "birth defect"[All Fields])

Eligibility Criteria

Only studies fulfilling our eligibility criteria, defined using the population, intervention, comparator, outcome (PICO) framework, were included [21]:

1. **Patients/Population:** Pregnant women living with HIV and/or the children who were exposed to HIV exposed or infected during pregnancy. Women at any gestational age were included, as the recent literature has suggested that focusing on CA—congenital abnormalities only on the first trimester is outdated [15]. All countries and settings were eligible for inclusion.

2. **Interventions:** All combinations and doses of Antiretroviral Therapies (ARTs). Standard combination ART (cART) consists of the combination of antiretroviral (cART) drugs to maximally suppress the HIV virus and stop the progression of HIV disease. To be included,
The pregnant woman should have taken ARTs for at least one month before delivery. Studies were not included if they examined ARTs that were directly administered to neonates, infants, or children, were pre-clinical animal studies, analyzed prophylaxis-only doses, or did not specify if women were on ART.

3. Comparators: Studies comparing antiretroviral medications administered to HIV-positive mothers to i) HIV-negative women, ii) HIV-positive women not receiving treatment, and iii) HIV-positive women taking another type of ART medication (medication classes as per the WHO 1a-1f category) were included. The last group was considered for a sub-group analysis to determine the effects of individual drugs on the occurrence of congenital anomalies.

4. Outcomes: The presence or absence of congenital anomalies must have been stated and unambiguous. The primary safety outcome is major congenital malformations (overall and by specific type), which is defined by the International Classification of Diseases, 10th revision. ICD-10 includes congenital anomalies/birth defects in Chapter XVII. Congenital anomalies were operationalized for this study as anatomical or functional defects, including metabolic ailments, which were present at birth.

5. Study designs: Experimental (randomized clinical trials and non-RCTs), quasi-experimental (controlled before and after and time series), and observational (cohort, case-control, and drug registry) studies were included for the analysis. However, cross-sectional studies and case reports were excluded.

6. Other limitations: No limitations were imposed on publication status, study site, and the duration of the study. However, non-English literature studies and studies with a very small sample size (<50) used to detect differences were excluded.

Quality Appraisal of Papers and Risk of Bias

A structured template adopted from the Newcastle-Ottawa Scale was used to appraise each paper. To facilitate the improvement in the quality of reporting of observational studies, the STROBE statement was used. The quality criteria included whether sampling procedure and sample size calculations were illustrated, the description of algorithms and results known for both HIV-infected and uninfected women, the process of measurement of
the outcome and the degree of blinding of the investigators about mother’s infection status; the degree of follow-up; and the strategies used to control for confounding.

We used the GRADE approach to assess and grade the confidence of evidence for each outcome in the involved studies. Eight criteria were used to either downgrade or upgrade from each study. As a rule of thumb, GRADE starts with a baseline rating of high for RCTs and low for non-RCTs, including observational studies. The five criteria used to downgrade the research quality are presence of risk of bias, inconsistency, indirectness of evidence, imprecision or lack of reliability, and publication bias.

If a serious concern exists, the evidence was downgraded by one level, e.g., such as from high to moderate (-1). If a very serious concern exists, the evidence was downgraded by two levels, e.g., such as from high to low (-2). The other three criteria (criteria 6 to 8) used to upgrade the grade are a large magnitude of effect, a dose-response effect, and all reasonable confounding factors on reducing the effect (where an effect is observed) or suggesting a false effect (when no effect is observed). We judged whether the evidence should be upgraded once (+1) or twice (+2) for criteria 6-8. We integrated downgrading and upgrading factors to obtain an overall quality of evidence, ranked as high, moderate, low, or very low as specified by the GRADE approach. Finally, information concerning the quality of evidence for all outcomes was concisely combined in the “summary of findings” (Annex I: Supplementary/Table 1) [15]. Subsequently, data from each of the 30 included studies were abstracted and entered into a data abstraction tool that was developed in Microsoft Excel (2013).

**Statistical Analysis**

We saw noted that there was variability between the included studies due to clinical variability (variation in characteristics of participants, such as age and nutritional status, exposure, measurement, and categories of ART), differences in outcomes, and methodological variability (in study design). Using the Der-Simonian-Laird random effects model, an overall estimate of effect was determined [24]. For each study, the relative risk (RR) as the weighted measure of association was computed by comparing exposed to unexposed women or by comparing various ART regimens. This analysis also focused on the evaluation of the levels of discrepancy, the extents of dispersion, and the causes of heterogeneity. The

메모 오전 5: As a minor note, this slight revision for clarity was made based on https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/.
heterogeneity assessment included study design, publication status, study setting, and drug categories. STATA version 14 (StataCorp. 2015, College Station, TX, USA) statistical software was used to analyze the data.

**Assessment of the Extent of Inconsistency**

We used the eyeball technique on the forest plots to examine the dispersion of observed RRs. Along with the forest plot, we computed Cochran’s Q at $p < 0.10$ and the $I^2$ statistics with a 95% CI to formally quantify the proportion of variance in observed RRs that reflects reflected the true heterogeneity between studies rather than mere chance [25]. A heterogeneity value $P$ value with $p < 0.10$ demonstrated the existence of heterogeneity [27]. According to the criteria established by Higgins et al., we considered $I^2 < 25\%$ as indicating low heterogeneity, 50%-75% as indicating substantial heterogeneity, and >75% as indicating significant heterogeneity[26].

**Assessment of the Amount of Dispersion**

In addition to $I^2$, we also reported the $I^2$ statistic to indicate the amount of dispersion of true effects, mainly during in the subgroup analysis. We computed the 95% confidence interval CI to estimate the range within which a hypothetical new true measure of association was expected to be found in 95% of cases [28].

**Investigating the Causes of Heterogeneity**

Subgroup analyses were done to explore how drug intervention, clinical variability, and study design variability influenced the pooled estimate. Even though both clinical and study design diversity leads to statistical heterogeneity, previous literature indicates that clinical aspects may vary more across studies to a greater extent than design factors. The subgroup analyses in this study, therefore, focused on the assessment of assessing the effects of key characteristics of participants which that may be associated with the main outcomes of interest. Thus, we conducted subgroup analyses based on, but not limited to, antiretroviral drug class, publication status, fetal outcome, study setting, and methodological quality. We conducted a meta-regression from using the study average or proportion values of the main participant’s characteristics [28].
**Meta-Bias Assessment**

We used a funnel plot to explore the potential of publication bias and small study effect. Additionally, the Egger's test and Harbord test for funnel plot asymmetry were conducted [27]. **Meta-trim** A trim-and-fill analysis was also conducted to see if whether one or more studies influenced the estimate.

**RESULTS**

The database searches produced 765 articles after duplicates were removed. The selection of titles and abstracts resulted in 182 potentially relevant articles, of which 133 references were excluded due to the reasons presented in Supplementary Table 2 ([S2. annexed]). Finally, 49 studies met the predefined inclusion criteria and PICO assessment. Thirty studies were suitable for a quantitative synthesis (meta-analysis). Most of the selected studies were designed as cohort study studies as a study design. **Figure 1** displays the flow of data through the different stages of the systematic review.

The study-specific estimates appeared considerably heterogeneous (e.g., the CIs of the following studies, [7] and [3], did not overlap); hence, the fixed effect assumption might not be plausible for this dataset. This was reinforced by the Q-test, which also showed the presence of heterogeneity ($p=0.014$). The mean of the $I^2$ measure, which measures the amount of heterogeneity across studies, suggested the presence of moderate heterogeneity (39.8%). Therefore, we choose a random effect model for this particular study.

Of the 30 studies that were included in this review, 1,461 congenital anomalies were reported among a total of 53,186 births in women exposed to cART during pregnancy. For cART exposure during pregnancy, the pooled relative risk showed that those receiving cART had about a 10% increase in risk of having a child develop a congenital malformation compared to the non-exposed category (OR=1.09, 95% CI (1.04–1.14)) (Fig. 2).

**Heterogeneity Assessment**

**Meta-Regression**

The variables entered into the model were several variables were entered into the model. Studies were compared according to their setting (developing or developed countries). setting
compares studies done from developing and developed countries. Study design compares (RCTs to versus observational studies—), drug category—cARTs (Zidovudine—based, Nevirapine—based, Efavirenz—based, Protease inhibitor [PI]—based, or Integrase inhibitor—based cARTs—), methodological quality refers to high-grade [3+] and above vs. to low-grade [1-2], and finally and publication status (published versus unpublished) studies.

Some studies were done in developing countries, while others were done in developed counties. There were also differences in the ART regimen used among studies. Further, we suspected that study design, publication status, and methodological quality could moderate the differences observed in findings. As shown in Table 1, $\tau^2$ was 0.90, and it follows that $I^2$ was 96.8%. The joint test for all five covariates gave a $P_{p}$-value of 0.0059, indicating some evidence for an association of at least one of the covariates with the size of the treatment effect (Table 1).

**Galbraith’s Graph**

This graph represents the accuracy of each study versus the standardized effects. It also shows the adjusted regression line and sets two confidence bands. Five studies contributed to heterogeneity (Fig. 3).

**Subgroup Analysis**

The subgroup analysis found no differences in risk of congenital anomalies between these two groups (RR = 1.04; 95% CI = 0.97, 1.11). Heterogeneity between the studies was moderately low ($I^2$=29%). Similarly, efavirenz exposure seems to have an association with congenital anomalies. However, a 10% increase in risk of congenital anomalies was shown for Zidovudine and PI-based antiretroviral therapies (RR = 1.09; 95% CI = 1.00, 1.19 and RR = 1.10; 95% CI = 1.02, 1.18, respectively) (Fig. 4). Three studies have reported congenital anomalies with a new group of drugs (Integrate inhibitors) recently approved by the Food and Drug Administration (FDA) and WHO. The subgroup analysis revealed a 60% increased risk of congenital anomalies among associated with the use of integrate inhibitors (RR = 1.61; 95% CI = 1.60, 2.43). In addition to the sample size, the results of for the integrate inhibitors treatment should be interpreted cautiously because of both the small sample size and the moderately high heterogeneity ($I^2$ =58%).

**Summary of Treatment Effect**
The L’Abbé plot (Fig. 5) is another way to depict a summary of the effects of ART on congenital malformation. In the L’Abbé plot, studies are represented as follows: the risk of congenital anomalies (events) in the ART group (intervention group) is displayed on the y-axis and the event rate in the ART–naïve group (the control group) on the x-axis. Each circle represents an individual study and the size of the circle is proportional to study size. The 45° line is the line of no effect. The summary effect shows that ART has a small but significant causal effect on congenital anomalies.

Publication Bias

Funnel Plot and Counter-Funnel Plot

Funnel plots are scatter plots that display treatment effects from individual studies on the x-axis against a measure of study precision on the y-axis. Figure 6a plots the effect size against the variance of each study. Looking at the lower corner of the funnel plot, the slight emptiness of negative or null studies indicates a probability of publication bias [29]. The contour-enhanced funnel plot (Fig. 6b) helps to differentiate between publication bias and other causes of observed asymmetry. It displayed that small studies were found not only in the shaded areas (statistical significance) but also in the white areas (areas of non-statistical significance). Thus, several factors, and not solely publication bias, might be responsible for the asymmetry observed in the funnel plot. The areas where missing studies are perceived included regions of both low and high statistical significance (i.e., the area crossed over the contours), suggesting that both studies showing ART and congenital anomalies to be non-significantly and significantly associated were missing. Therefore, publication bias cannot be established as the only cause of funnel asymmetry.

Regression Analysis for Publication Bias

Harbord’s meta-regression model was calculated to measure the scale and statistical significance of the association between observed effect sizes and the size of studies (Fig. 7). The test points out that smaller studies did not tend to give different results when compared with larger trials, as the CI of the intercept contained the zero value. In addition, Harbord’s modified test for small-study effects showed that the estimated bias coefficient is 1.75 with a standard error of 0.905, giving a p-value of 0.065. The test provided weak evidence for the presence of small-study effects.
Trim and Fill Analysis

A sensitivity analysis was performed to identify studies that have had a larger influence on the estimates. Figure 8 visually provides a visual estimate with a 95% confidence interval. The sensitivity analysis revealed that no individual studies affected the estimate excessively. Exclusion—The exclusion of reference [7] was influential, but the influence impact was not statistically significant. The trim-and-fill analysis revealed no significant asymmetry of the funnel plot (estimated effect size = 1.087; [95% CI, 1.14–3.56] versus observed RR = 1.09; 95% CI, 1.04–1.14).

DISCUSSION

This meta-analysis aimed to assess the extent to which exposure to ART during pregnancy is associated with congenital anomalies through a systematic review of published papers. This was assessed through measures of effect on the association and overall pooled estimates of these measures through a meta-analysis.

The pooled relative risk showed that mothers receiving cART were at about a 10% higher risk of having a child with a congenital malformation (RR = 1.09; 95% CI, 1.04–1.14) compared to the non-exposed category.

The connection between exposure to ART and the risk of adverse pregnancy outcomes, including congenital anomalies, has not been clearly explained. Studies are commonly conducted in developed countries and are primarily limited by sample size, observational design, using patient registries or different comparison groups, and conflicting data [30]. In developing countries, for example, a study reported that micronutrient insufficiencies and exposure to environmental pollutants, as well as co-infections, may be more common in HIV-infected pregnant women, and this could result in increased prevalence of certain birth defects [16]. The other most likely hypothesis is that ART and immune reconstitution could modulate the Th1 to Th2 shift required by normal pregnancy [31]. Another proposed hypothesis is an etiological association between in utero NRTI exposure and...
mitochondrial dysfunction [6, 32, 33]. Mitochondrial toxicity is associated with cognitive development and congenital anomalies [34].

The subgroup analysis yielded more specific results. Our findings support previous reports and meta-analyses stating that there is no association between efavirenz exposure and congenital anomalies [35]. However, a 10% increased risk of congenital anomalies was shown for Zidovudine and PI-based antiretroviral therapies (RR = 1.09; [95% CI = 1.00, 1.19]) and (RR = 1.10; [95% CI = 1.02, 1.18]), respectively.

A large study in the U.S. revealed a similar finding. Although the pooled estimate shows that the absolute risk of congenital anomalies was relatively low, some individual drugs, such as Atazanavir, showed relative increases in the risk of overall congenital anomalies and specific anomalies, which warrant further study [5, 18]. Our cumulative meta-analysis (Meta-cum) also revealed there was no significant change over time, meaning the newest drugs may also cause congenital anomalies.

For instance, the proportion of neural tube defects (NTDs) among infants of women exposed to dolutegravir at conception is 3.16%. This percentage is higher than the proportion of 0–10% expected in the general population in Sub-Saharan Africa, signifying that the result might not be explained by chance alone. The subgroup analysis revealed a 60% increased risk of congenital anomalies among integrase inhibitor users (RR = 1.61; [95% CI = 1.60, 2.43]). Therefore, integrase inhibitors should be wary studied.

CONCLUSION

One of the limitations of this review is the inclusion of literature written only in English. Furthermore, only a few studies conducted in developing countries were available, and some available ones were excluded at the methodological assessment stage due to the small number of enrolled participants.

The findings of this review show that combined antiretroviral therapy cART may not be a significant risk factor for congenital anomalies. It affirms the findings of previous systematic
reviews that claim efavirenz does not have an association with congenital anomalies. In contrast, there was about a 10% increased risk of congenital anomaly deliveries among women who were exposed to PI-based regimens, ddC, and integrase inhibitor treatments as compared to non-exposed individuals.

RECOMMENDATIONS

- The number of HIV-exposed but uninfected infants who have been exposed to combined ART is rapidly increasing, especially in resource-limited settings and as new drugs are introduced. There is an urgent need to carefully monitor the short- and long-term consequences of antiretroviral drug exposure among pregnant women through surveillance systems.

- During the development of WHO or National ARV ART guidelines, careful consideration should be taken in the inclusion of PI-based regimens, integrase inhibitors, ddC, and newer drugs.

- Further studies are needed in developing countries where environmental, nutrition, adherence, and other factors could influence the occurrence of congenital anomalies.

Acknowledgements.
We acknowledge all authors who did the primary studies and allowed us to use their work.

REFERENCE


Table 1—Meta-regression of the effects of covariates on the association of antiretroviral therapy and congenital anomalies

| Variable          | Coefficient | Std. Err. | t     | P>|t| | 95% Conf. Interval CI |
|-------------------|-------------|-----------|-------|-----|----------------------------|
|                   |             |           |       |     | Lower CI                   |
|                   |             |           |       |     | Upper CI                   |
| Setting           | 1.481641    | .6585354  | 2.25  | 0.036 | .1079601                   |
|                   |             |           |       |     | 2.855322                   |
| Study design      | -.0373077   | .3059196  | -0.12 | 0.904 | -.6754447                  |
|                   |             |           |       |     | .6008293                   |
| Drug category     | .536819     | .1480523  | 2.39  | 0.027 | .0448502                   |
|                   |             |           |       |     | .6625135                   |
| Methodological quality | -.0118461   | .3063184  | -0.04 | 0.970 | -.6508151             |
|                   |             |           |       |     | .627123                    |
| Publication status| .7605734    | .4961117  | 1.53  | 0.141 | -.2742975                 |
|                   |             |           |       |     | 1.795444                   |
| _cons             | -1.904974   | 1.138541  | -1.67 | 0.110 | -4.279928                 |
|                   |             |           |       |     | .4699811                   |

Between studies variation ($\tau^2=0.0049$). Prob > F = 0.0059 (Adj. adjusted $R^2$-squared = 43.40%). Only drug category showed a significant effect.

Note: “Setting” compares study studies from developed versus developing countries, “Study design” compares randomized controlled trials with observational studies, “Drug category” compares the 6 WHO World Health Organization drug categories (1a-1f), “Methodological quality” assesses the overall grading of strong versus weak, and “Publication status” compares published versus un-published studies.
Fig 1. RISMA flow diagram for selection of studies identified for systematic review and meta-analysis of association of antiretroviral therapy and congenital anomalies
Fig 2: Meta-analysis of effect antiretroviral therapy on risk of congenital anomalies for those on ART compared to ART naïve

Legend: I²- heterogeneity statistic; ES- effect size; %—percent. For each independent sample included in the meta-analysis, a corrected ES (square) and the associated 95% confidence interval (CI; line) is shown. On the bottom, the diamond shows the meta-analytically inverse variance weighted mean ES. The values associated with the ESs and CIs are located in the right column. Positive values for ES indicate greater positive affect on risk of congenital anomalies. Blue box = effect estimates from single studies. Diamond = pooled result with confidence interval. Vertical line at ‘1’ on the x-axis is the line of no effect. Weight (in %) = influence an individual study had on the pooled result.
Figure 3. Galbraith plot for log odds ratio of congenital anomalies in patients with antiretroviral therapy and ART naïve.

Legend: The slope of the line is equal to the pooled effect. The 95% limits will be 2 units above and below this line. We expect 95% of points to be between these limits if there is no heterogeneity. Studies outside the Galbraith limits are studies where the 95% confidence interval does not contain the pooled estimate and contribute to the heterogeneity.
**Fig 4. Subgroup analysis effect of different categories of ART on risk of congenital anomalies**

D2 - heterogeneity statistic; ES - effect size; %—percent, sub-groups with (D2 = .%, p = .). For each independent sample included in the meta-analysis, a corrected ES (square) and the associated 95% confidence interval (CI, line) is shown. On the bottom, the diamond shows the meta-analytically inverse variance weighted mean ES. The values associated with the ESs and CIs are located in the right column. Positive values for ES indicate greater positive effect on risk of CAs. Black box = effect estimates from single studies. Diamond = pooled result with confidence interval. Vertical line at ‘1’ on the x-axis is the line of no effect. Weight (in %) = influence an individual study had on the pooled result.
Fig 5: L' Abbé plot depict a summary of ART effect on congenital malformation

Legend: Plot of ART naive group risk (x-axis) against treatment group risk (y-axis). The dotted green line represents the line of no effect. The green line represents the combined effect of all studies as RR. The blue circles represent the results of individual studies, with size representing study weight.
**Fig 6a and 6b: Meta funnel and Counter Enhanced Funnel**

**Fig 6a:** Meta funnel with pseudo 95% confidence interval to detect publication bias on the effect of ART on congenital anomalies.

**Figure 6b:** Contour-enhanced funnel plot of the 30 trials comparing ART and ART naive for risk of congenital anomalies. The shaded areas correspond to levels of statistical significance defined by the p value of a z-test for the log (RR, risk ratio).
Figure 7: Harbord's modified test for small-study effects:

Regress $Z / \sqrt{V}$ on $\sqrt{V}$ where $Z$ is an efficient score and $V$ is scored variance.

<table>
<thead>
<tr>
<th>$Z / \sqrt{V}$</th>
<th>Coefficient</th>
<th>Std. Err.</th>
<th>$t$</th>
<th>$P &gt; t$</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{V}$</td>
<td>-1.226333</td>
<td>.2028863</td>
<td>-0.60</td>
<td>0.551</td>
<td>[-.5404854 , .2952188]</td>
</tr>
<tr>
<td>bias</td>
<td>1.750231</td>
<td>.905112</td>
<td>1.93</td>
<td>0.065</td>
<td>[-1.138821 , 3.614344]</td>
</tr>
</tbody>
</table>

Test of $H_0$: no small-study effects $P = 0.065$

**Figure 7: Harbord's modified test Modified Gilberts plot for small study effect**

**Legend:** The estimated intercept is 1.75 with a standard error of 0.905, giving a $p$-value of 0.065. The modified test thus suggests little evidence for small-study effects.
Figure 8: Trim and fill analysis for small study effect (Note 5 studies are trimmed and filled)

Legend: Fig 8 panel is a funnel plot of log RR of Effect of ART on Congenital anomaly. Open circles are original data, Circles inside square are imputed filled values. The trim-and-fill analysis revealed no significant asymmetry of the funnel plot (estimated ES = 1.087 [1.14-3.56] versus observed RR=1.09, 95% CI (1.04–1.14). Indicating publication bias has no significant effect on ART and CA outcome.