

Bayesian Meta-Analysis with Application in Dental Studies

Sara Antonijevic, Danielle Sitalo, and Brani Vidakovic

Department of Statistics, Texas A&M University

June 2025

Abstract

Dental caries remain a persistent global health challenge, and fluoride varnish is widely used as a preventive intervention. This study synthesizes evidence from multiple clinical trials to evaluate the effectiveness of fluoride varnish in reducing *Decayed–Missing–Filled (DMF)* surfaces. The principal measure of efficacy is the *Prevented Fraction (PF)*, representing the proportional reduction in caries relative to untreated controls.

A comprehensive meta-analysis was conducted using fixed-effect and random-effects models, complemented by hierarchical Bayesian inference. The Bayesian framework incorporated multiple prior distributions on between-study variance—including Pareto, half-normal, uniform, beta, and scaled chi-square forms—to assess robustness under alternative heterogeneity assumptions. Across all specifications, the pooled estimate indicated an approximate 43% reduction in caries incidence, with credible intervals consistently excluding the null.

Compared to classical methods, the Bayesian approach provided richer uncertainty quantification through full posterior distributions, allowed principled incorporation of prior evidence, and offered improved inference under heterogeneity and small-sample conditions. The stability of posterior estimates across diverse priors reinforces the robustness and reliability of the conclusions.

Overall, findings confirm fluoride varnish as an effective and consistent preventive measure, and demonstrate the value of Bayesian hierarchical modeling as a powerful complement to traditional meta-analytic techniques in dental public health research.

1 Introduction

Dental caries remain a major global public health concern, representing one of the most prevalent chronic conditions across all age groups (World Health Organization, 2022). The disease results from bacterial demineralization of tooth enamel and dentin and can lead to pain, infection, and tooth loss when untreated. The burden of disease is commonly assessed using the **Decayed–Missing–Filled (DMF)** surfaces index, which quantifies the number of tooth surfaces affected by decay, extracted due to caries, or restored as a consequence of prior lesions (Klein & Palmer, 1938).

Fluoride-based interventions, including fluoride varnish, have long been recognized as effective preventive measures against dental caries (Marinho et al., 2013). However, reported

magnitudes of effect vary across populations, age groups, baseline risk levels, and study designs. To obtain an overall estimate of effectiveness and reconcile heterogeneous findings, a quantitative evidence synthesis was conducted using *meta-analysis*.

In this study, effectiveness is summarized using the **Prevented Fraction (PF)**, a measure representing the proportion of dental caries prevented in the treatment group relative to the control group. The PF quantifies the fraction of disease risk that was averted due to the intervention, with values near zero indicating little or no preventive benefit, and larger negative values denoting stronger caries reduction. A formal definition and estimation procedure are presented in Section 2.1.

Two complementary classical approaches were employed to combine study-level estimates: the **fixed-effect** and **random-effects** models (DerSimonian & Laird, 1986a). The fixed-effect model assumes a common true effect across studies and pools estimates as inverse-variance weighted averages. In contrast, the random-effects model incorporates between-study heterogeneity through an additional variance component τ^2 , accommodating differences in populations, protocols, and settings. The degree of heterogeneity was evaluated using Cochran’s Q statistic and the I^2 measure, which quantify excess dispersion beyond chance (Higgins & Thompson, 2002).

To complement classical inference, a **Bayesian hierarchical meta-analysis** was performed to obtain posterior distributions for PF and to assess sensitivity to prior assumptions. Bayesian approaches offer coherent uncertainty quantification and allow incorporation of external information (Gelman, 2006). Three alternative priors for the between-study variance were considered:

- **Model A (Pareto prior):** permits substantial heterogeneity and yields heavy-tailed support for τ^2 ;
- **Model B (Uniform prior on variance ratio):** models the proportion of total variance attributable to between-study variation;
- **Model C (Scaled Chi-Squared prior):** imposes moderate shrinkage, providing more stable estimates when sample sizes are limited.

By comparing results across fixed-effect, random-effects, and Bayesian frameworks, this study offers a comprehensive assessment of the preventive effectiveness of fluoride varnish. The findings contribute to evidence-based dentistry and inform public health strategies for caries prevention.

2 Meta-Analysis

Meta-analysis is a quantitative framework for synthesizing results from multiple independent studies to obtain more precise and generalizable estimates of effect (Normand, 1999). It is especially valuable in medical and epidemiological research, where individual studies may be underpowered or context-specific. As emphasized by Normand (1999), meta-analysis is not a mere aggregation of outcomes but a structured inferential process involving systematic study identification, standardized effect-size estimation, and statistically rigorous pooling.

A central methodological choice concerns whether to employ a *fixed-effect* or *random-effects* model. The fixed-effect model assumes a single true underlying effect, attributing observed variability to sampling error. The random-effects model allows the true effect to vary across studies

and introduces a between-study variance component, denoted τ^2 , to account for heterogeneity (DerSimonian & Laird, 1986a). Inference may be based on maximum likelihood estimation (MLE), restricted maximum likelihood (REML), or Bayesian hierarchical formulations, each providing distinct advantages depending on sample size and heterogeneity structure.

Assessment of heterogeneity is a crucial diagnostic step. Standard measures include Cochran’s Q statistic, which tests the null hypothesis of homogeneity, and the I^2 statistic, which expresses the proportion of total variation attributable to genuine heterogeneity rather than chance (Higgins et al., 2003a). Moreover, potential *publication bias*—the preferential publication of statistically significant or positive findings—can distort pooled estimates. Visual inspection via funnel plots and formal tests such as Egger’s regression are commonly employed to evaluate such bias.

2.1 Prevented Fraction

The **Prevented Fraction (PF)** quantifies the proportion of dental caries prevented in the treatment group relative to the control group (Greenland & Drescher, 1987; Rothman et al., 2008). It is defined in terms of the **Relative Risk (RR)** as

$$PF = RR - 1 = \frac{X_t - X_c}{X_c} = \frac{X_t}{X_c} - 1,$$

where **RR** denotes the relative risk of dental caries in the treatment group compared to the control group, and X_t and X_c denote the mean numbers of DMF surfaces in the treatment and control groups, respectively. Under this definition, $PF = 0$ implies no difference between groups, while negative PF values indicate a beneficial preventive effect (a reduction in disease incidence).

Some authors alternatively define $PF = 1 - RR$, which simply reverses the sign; both forms are mathematically equivalent, but the $RR - 1$ convention used here yields negative values for caries prevention and aligns with standard epidemiological interpretation.

Assuming approximate normality of sample means, we let X_t and X_c represent random variables corresponding to outcomes of interest (e.g., mean DMF surfaces) in treatment and control groups, with means μ_t, μ_c and variances σ_t^2, σ_c^2 . Under large-sample assumptions,

$$X_t \sim N\left(\mu_t, \frac{\sigma_t^2}{n_t}\right), \quad X_c \sim N\left(\mu_c, \frac{\sigma_c^2}{n_c}\right),$$

where n_t and n_c are sample sizes.

Its sampling variance, required for confidence intervals and weighting, is

$$\text{Var}(PF) = \text{Var}\left(\frac{X_t}{X_c}\right).$$

For a fixed-effect synthesis across k studies, the pooled estimate is an inverse-variance weighted mean:

$$\widehat{PF}_{FE} = \sum_{i=1}^k w_i PF_i, \quad w_i = \frac{1/V_i}{\sum_{j=1}^k 1/V_j},$$

where V_i denotes the estimated variance of PF_i . When between-study variability is non-negligible, a random-effects model incorporating τ^2 yields more appropriate inference. Bayesian hierarchical models further extend this framework by introducing priors on τ^2 and enabling full posterior inference, thereby integrating prior information with observed data (Gelman, 2006).

3 Description of the Data

The data analyzed in this study comprise a set of clinical trials evaluating the effectiveness of topical fluoride varnish interventions in preventing dental caries. Across all included studies, the primary outcome measure is the number of **Decayed–Missing–Filled (DMF)** surfaces. Each study reports results for both treatment and control groups, with the treatment arm receiving fluoride varnish applications and the control arm receiving either placebo or no treatment. For each study, mean DMF values, standard deviations, and sample sizes were extracted to compute the *Prevented Fraction (PF)*, defined as the proportional reduction in DMF surfaces in the treatment group compared with controls. The variance of PF was derived from the reported standard errors and used to determine study weights in the meta-analysis, ensuring that larger and more precise studies contribute more to the pooled estimate.

A summary of the individual study results is presented in Figure 1. Each red square represents the point estimate of PF for an individual trial, with the size of the square reflecting the study’s sample size. Horizontal lines denote 95% confidence intervals, and the dashed vertical line at zero indicates the null hypothesis of no treatment effect. The pooled estimate appears as the red diamond at the bottom of the plot, representing the overall magnitude and precision of fluoride varnish effectiveness. Most point estimates lie to the left of zero, indicating a consistent preventive effect across studies, with the exception of the trial by Milsom et al.

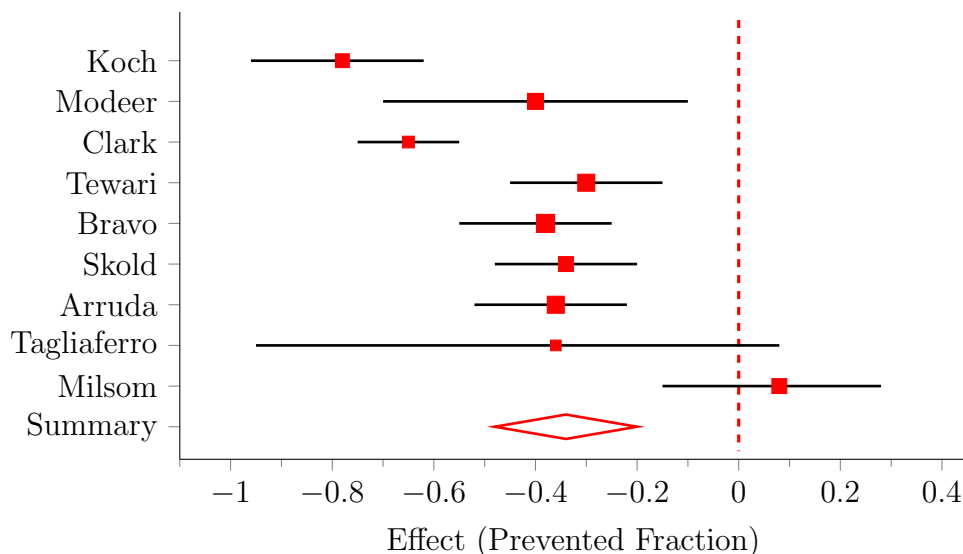


Figure 1: Fixed-effect meta-analysis on prevented fraction

The trial conducted by Milsom et al. (Milsom et al., 2011) did not demonstrate a clinically significant effect of fluoride varnish on first permanent molars among 7- to 9-year-old children. The mean number of **Decayed and Filled Surfaces (DFS)** was nearly identical between the varnish and control groups: 0.65 (SD = 2.15) versus 0.67 (SD = 2.10). The authors concluded that there was “insufficient evidence of an effect of intervention on caries at follow-up for either DFS... or DFT” (Milsom et al., 2011). Several factors likely contributed to this result. Approximately 80% of children had DFS = 0 at baseline, limiting the study’s ability to detect significant reductions. The authors further noted that “it is questionable whether the treatment benefits of topical fluoride evidenced in trials undertaken in high-caries populations with relatively low fluoride exposure will be seen in trials with participants with frequent

exposure to fluoride” (Milsom et al., 2011). Thus, the limited incremental benefit may reflect modern fluoride exposure levels. In addition, only half of eligible parents provided consent, further reducing effective sample size and study power. Collectively, these factors may explain the absence of a statistically or clinically meaningful effect in this trial.

3.1 Study Characteristics and Risk of Bias

The clinical trials included in this meta-analysis vary in design, sample size, and geographic setting, introducing both heterogeneity and potential sources of bias. The study by Milsom et al. (Milsom et al., 2011) was the largest and among the most methodologically rigorous, enrolling over 2,600 schoolchildren in the United Kingdom. It employed computer-generated randomization stratified by school size and location, achieving strong baseline comparability and balanced attrition rates. Due to its large sample, this trial heavily influences the pooled estimate.

In contrast, Tewari et al. (Tewari et al., 1990) conducted a four-arm randomized controlled trial in India with a distilled-water placebo and stratified randomization by demographic and caries-related factors. Both participants and examiners were blinded, enhancing internal validity. These methodological strengths reduce bias and lend credibility to observed treatment effects.

Other studies were smaller in scale or less rigorous. Tagliaferro et al. (Tagliaferro et al., 2011) used quasi-randomization based on systematic alternation rather than true random allocation, introducing potential selection bias. Bravo et al. (Bravo et al., 1997) employed a cluster quasi-randomized design without detailed reporting of allocation concealment, raising concerns about baseline equivalence. Modeer et al. (Modeer et al., 1984) reported unequal dropout rates—26% in the treatment group versus 9% in controls—potentially biasing results if attrition correlated with outcomes. These limitations suggest that while the overall trend supports fluoride varnish efficacy, individual study findings should be interpreted with caution.

Outcome assessment methods also differed. Radiographic techniques with intra-examiner calibration were used in studies such as Modeer (Modeer et al., 1984) and Sköld (Sköld et al., 2005), whereas large-scale trials like Milsom (Milsom et al., 2011) relied on clinical examinations by multiple examiners. Definitions of dental caries were not uniform: some studies measured increments in DMFs, while others, such as Arruda et al. (Arruda et al., 2012), included both cavitated and non-cavitated lesions using ICDAS criteria. These diagnostic discrepancies introduce heterogeneity and complicate direct comparison. Additionally, contamination of control groups was documented in Sköld et al. (Sköld et al., 2005), where 95% of controls received at least one fluoride varnish application, potentially underestimating treatment effects.

Despite such variability, the fixed-effect analysis demonstrates a clear overall benefit of fluoride varnish in reducing dental caries incidence. The pooled estimate, depicted by the diamond in Figure 1, corresponds to a PF of approximately 40–45%, consistent with prior methodological discussions (DuMouchel, 1994; Normand, 1999). The confidence interval around the summary effect excludes the null, confirming statistical significance and clinical relevance. Estimated numbers needed to treat (NNT) range from 1.4 to 3.2, depending on baseline caries levels, underscoring the preventive potential of fluoride varnish.

Overall, this dataset integrates evidence from large, high-quality randomized controlled trials alongside smaller studies, yielding variation in design and outcome measurement. When considered collectively, the evidence strongly supports a protective effect of fluoride varnish. Methodological differences and potential biases may affect precision but are unlikely to overturn

the consistent pattern of caries reduction observed across diverse populations. The fixed-effect results provide a precise pooled estimate, while complementary random-effects and Bayesian analyses further account for heterogeneity, reinforcing confidence in the overall conclusions.

4 Methodology

Given the methodological variability among the clinical studies, both **random-effects** and **Bayesian hierarchical models** were employed to account for between-study heterogeneity in the analysis. This approach ensures that variation in study design, population characteristics, and baseline caries risk is properly incorporated into the estimation of the overall effect.

4.1 Classical Definition

In classical statistical inference, the **Prevented Fraction (PF)** quantifies the effectiveness of an intervention in reducing risk. As defined in Section 3, PF is expressed in terms of the relative risk (RR) as

$$PF = RR - 1.$$

This measure reflects the proportionate reduction in adverse outcomes attributable to the intervention. Estimation of PF relies on standard frequentist methods, with variances obtained from conventional error-propagation formulas to facilitate confidence interval construction and weighting in meta-analysis.

4.1.1 Fixed-Effect Analysis

As discussed in Section 2, the **fixed-effect model** assumes that all studies in the meta-analysis estimate a common true effect size. Under this model, any observed differences among study estimates are attributed solely to sampling variability rather than genuine heterogeneity. The summary prevented fraction (SPF) is computed as an inverse-variance weighted average of study-specific PF estimates:

$$SPF = \sum_{i=1}^k w_i PF_i, \quad w_i = \frac{1/V_i}{\sum_{j=1}^k 1/V_j},$$

where w_i denotes the weight for study i and V_i its variance.

The variance of the pooled estimate is given by

$$\text{Var}(SPF) = \frac{1}{\sum_{i=1}^k 1/V_i},$$

leading to the 95% confidence interval (CI):

$$\left[SPF - 1.96 \sqrt{\text{Var}(SPF)}, SPF + 1.96 \sqrt{\text{Var}(SPF)} \right].$$

Fixed-effect models are appropriate when between-study heterogeneity is negligible, studies share similar designs and populations, or when the objective is to estimate a single common effect size.

4.1.2 Random-Effects Analysis

A **random-effects model** is applied when effect sizes are expected to vary across studies due to differences in sample characteristics, intervention implementation, or methodological factors. As outlined in Section 2, the random-effects formulation introduces a between-study variance component τ^2 , representing heterogeneity beyond sampling error. The total variance for each study is therefore:

$$V_i^* = V_i + \tau^2.$$

Weights under this model incorporate both within- and between-study variability:

$$w_i^* = \frac{1/V_i^*}{\sum_{j=1}^k 1/V_j^*}.$$

Estimation of τ^2 can be performed using the DerSimonian–Laird method (DerSimonian & Laird, 1986b), restricted maximum likelihood (REML), or a Bayesian hierarchical framework. The summary PF under the random-effects model is computed analogously:

$$SPF = \sum_{i=1}^k w_i^* PF_i,$$

with variance

$$\text{Var}(SPF) = \frac{1}{\sum_{i=1}^k 1/V_i^*},$$

and 95% confidence interval

$$\left[SPF - 1.96 \sqrt{\text{Var}(SPF)}, SPF + 1.96 \sqrt{\text{Var}(SPF)} \right].$$

The degree of heterogeneity is assessed using Cochran’s Q statistic and the I^2 index (Higgins et al., 2003b), where I^2 represents the proportion of total variation attributable to between-study differences. Values of I^2 exceeding 50% generally indicate substantial heterogeneity, justifying the use of a random-effects model.

4.2 Bayesian Formulation

In contrast to the frequentist framework, which treats the prevented fraction (PF) as a fixed but unknown quantity, Bayesian statistics expresses uncertainty through probability distributions. Rather than providing a single-point estimate with a sampling-based confidence interval, Bayesian inference yields a full posterior distribution for PF, reflecting updated beliefs after observing data. This probabilistic approach allows for natural incorporation of prior knowledge, coherent uncertainty quantification, and robust inference even with small sample sizes or limited study counts.

4.2.1 Hierarchical Bayesian Model

Hierarchical Bayesian models are widely used in meta-analysis because they explicitly represent both within-study and between-study variation. Under this framework, each study-specific effect θ_i (the PF estimate for study i) is assumed to follow a normal distribution centered on the overall mean effect μ with between-study variance τ^2 :

$$\theta_i \sim N(\mu, \tau^2).$$

The overall mean μ itself is assigned a prior distribution:

$$\mu \sim N(\mu_0, \sigma_0^2),$$

where μ_0 and σ_0^2 are hyperparameters reflecting prior uncertainty. The variance τ^2 captures heterogeneity across studies and is also assigned a prior. This structure allows “borrowing of strength” across studies—information from one study helps refine estimates for others—leading to more stable pooled results.

Bayesian meta-analysis offers several advantages. It unifies fixed- and random-effects ideas under a single hierarchical model, allows incorporation of prior evidence, and yields interpretable posterior distributions. Additionally, Bayesian methods often perform well when the number of studies is small, a setting in which frequentist large-sample approximations may be unreliable.

4.2.2 Prior Variance Models in Bayesian Meta-Analysis

Modeling the between-study variance is central to Bayesian meta-analysis, as prior assumptions about τ^2 influence posterior estimates of both heterogeneity and the pooled PF. To evaluate robustness, three prior specifications (Models A, B, and C) were considered, each reflecting a distinct perspective on heterogeneity.

Model A: DuMouchel’s Pareto Prior Model A adopts the Pareto prior introduced by DuMouchel for the standard deviation of true effects, τ :

$$\pi(\tau) = \frac{s_0}{(s_0 + \tau)^2},$$

where s_0^2 is the harmonic mean of within-study variances. The harmonic mean scale parameter ensures that studies with higher precision exert greater influence on the prior, balancing contributions across the dataset (DuMouchel, 1994). This prior allows substantial heterogeneity while down-weighting implausibly small values, making it well-suited for datasets expected to exhibit moderate to large between-study variation.

Model A Variant: Half-Normal Prior on τ To examine the sensitivity of results to tail behavior, DuMouchel’s Pareto prior was replaced by a half-normal prior $\tau \sim \mathcal{N}(0, 1)$. The half-normal imposes stronger regularization on small τ values and avoids the heavy tails of the Pareto distribution, producing a more concentrated posterior. All other components of the model (likelihood, PF specification, and priors on μ) were held constant for comparability.

Model B: Uniform Prior on Variance Ratio Model B places a prior directly on the proportion of total variance attributable to between-study heterogeneity:

$$B = \frac{\tau^2}{\tau^2 + s_0^2}, \quad B \sim \text{Uniform}(0, 1).$$

Here, B represents the fraction of total variance due to heterogeneity. Values near 1 indicate substantial between-study variation, while values near 0 suggest homogeneity. This formulation provides an intuitive interpretation of heterogeneity and enables adaptive shrinkage.

In this model, τ^2 is derived from B and within-study precision:

$$\tau^2 = \frac{B \cdot s_0^2}{1 - B},$$

where $s_0^2 = \frac{N}{\sum_i P_i}$ and P_i denotes the inverse-variance weight for study i . As $B \rightarrow 1$, heterogeneity increases; as $B \rightarrow 0$, the model reflects greater confidence in study precision. The scale term s_0^2 links observed variability to the heterogeneity component.

To test prior robustness, the noninformative prior `dflat()` on the pooled effect (SPF) was replaced with a weakly informative Normal prior:

$$SPF \sim \mathcal{N}(0, 0.001^{-1}).$$

If estimates of SPF and τ^2 remain stable under this change, it indicates model robustness to mild prior regularization.

Impact of Alternative Priors on B The sensitivity of Model B was further examined using $\text{Beta}(0.9, 1)$ and $\text{Beta}(1, 0.9)$ priors as alternatives to $\text{Uniform}(0, 1)$. The $\text{Beta}(0.9, 1)$ prior slightly favors larger B values, permitting greater heterogeneity, while $\text{Beta}(1, 0.9)$ emphasizes shrinkage toward higher B , supporting modestly stronger heterogeneity assumptions when justified by data.

Model C: Scaled Chi-Squared Prior on Precision Model C employs a scaled Chi-Squared prior on the precision of τ^2 :

$$\tau^{-2} \sim \frac{\chi_d^2}{d},$$

where d denotes degrees of freedom, controlling prior flexibility. This prior constrains extreme heterogeneity and prevents any single study from dominating the posterior.

Comparison with AG and ag Priors To explore further sensitivity, the scaled Chi-Squared prior was replaced by two Gamma-based alternatives. The first, the AG prior,

$$\frac{1}{\tau^2} \sim \text{Gamma}(0.001, 0.001),$$

is weakly informative, allowing wide dispersion and exerting minimal influence (Pateras et al., 2021). It serves as a neutral benchmark when prior information is limited, though its flatness can produce high posterior uncertainty in meta-analyses with few studies.

The second, a more informative ag prior,

$$\frac{1}{\tau^2} \sim \text{Gamma}(0.1, 0.1),$$

places greater mass near moderate heterogeneity, stabilizing estimates while maintaining flexibility. The ag prior balances realism and regularization, providing more stable posteriors when the AG prior yields excessive uncertainty, especially in small-sample meta-analyses.

Comparison of Scaled Chi-Squared vs AG Prior on Heterogeneity To assess sensitivity to prior assumptions on between-study variability, the scaled Chi-Squared prior on τ^2 was replaced with the AG prior on precision,

$$\frac{1}{\tau^2} \sim \text{Gamma}(0.001, 0.001).$$

This prior is weakly informative and induces a highly diffuse distribution on τ^2 , permitting substantial mass over both very small and very large heterogeneity. In practice, such flatness can be advantageous when prior information is scarce, providing a neutral reference for posterior learning. However, in meta-analyses with few studies or limited information, the AG prior can yield wider credible intervals and greater posterior uncertainty than more structured alternatives, reflecting its de-emphasis on regularization (Pateras et al., 2021). As a result, it is a useful benchmark for sensitivity analysis but may be less stabilizing when heterogeneity is weakly identified.

Comparison of Scaled Chi-Squared vs ag Prior on Heterogeneity Further sensitivity analysis replaced the scaled Chi-Squared prior with a more informative ag prior on precision,

$$\frac{1}{\tau^2} \sim \text{Gamma}(0.1, 0.1).$$

Relative to the AG specification, this prior places greater mass on moderate precisions and correspondingly down-weights extremely large heterogeneity. The resulting posterior typically exhibits reduced tail sensitivity for τ^2 and more stable interval estimates, particularly in sparse settings. The ag prior thus offers a pragmatic balance: it allows heterogeneity when supported by the data, yet tempers undue dispersion that can arise under highly diffuse priors. In applications with few trials, it often provides improved numerical stability while retaining interpretability.

5 Models and Results

5.1 Model A: Pareto Prior on Variance

Table 1: Bayesian Meta-Analysis Results (Model A: Pareto Prior on Between-Study Variance)

Parameter	Mean	SD	2.5% CI	97.5% CI
Summary PF (SPF)	−0.4341	0.0839	−0.6007	−0.2644
Between-study var (τ^2)	0.0496	0.0443	0.0065	0.1620
Harmonic Mean Var (s_0^2)	0.0095	≈ 0	0.0095	0.0095
Koch	−0.6665	0.1191	−0.9073	−0.4432
Modeer	−0.3557	0.0999	−0.5485	−0.1564
Clark	−0.2609	0.0834	−0.4227	−0.0957
Tewari	−0.6505	0.1138	−0.8805	−0.4368
Bravo	−0.4277	0.0671	−0.5598	−0.2955
Skold	−0.5481	0.0773	−0.7024	−0.3988
Arruda	−0.4076	0.0662	−0.5371	−0.2768
Tagliaferro	−0.4292	0.1606	−0.7494	−0.1069
Milsom	−0.1599	0.1525	−0.4379	0.1535

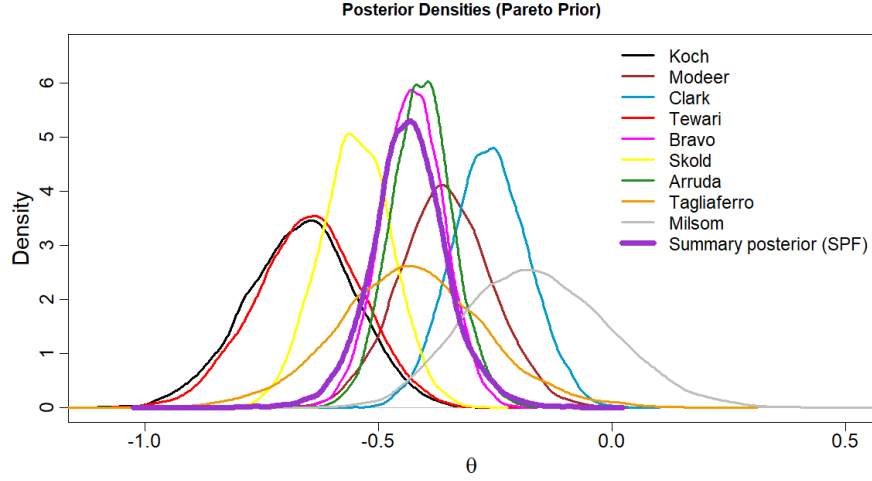


Figure 2: Posterior densities of study-specific effects under the DuMouchel (Pareto) prior. Colored curves correspond to individual studies; the bold purple line represents the posterior distribution of the overall summary effect (SPF).

Interpretation of Results: The pooled SPF of -0.4341 indicates an approximate 43% reduction in DMF relative to control, with a 95% credible interval $[-0.6007, -0.2644]$ entirely below zero. The posterior for τ^2 centers at 0.0496 (95% CrI $[0.0065, 0.1620]$), consistent with moderate heterogeneity. Study-level posteriors are uniformly negative except for *Milsom*, whose interval includes zero.

5.1.1 Model A: Half-Normal Prior

Table 2: Bayesian Meta-Analysis Results (Half-Normal Prior on Between-Study Standard Deviation)

Parameter	Mean	SD	2.5% CI	97.5% CI
Summary PF (SPF)	−0.4328	0.0991	−0.6298	−0.2322
Between-study SD (τ)	0.2519	0.1016	0.1066	0.4995
Between-study var (τ^2)	0.0738	0.0695	0.0114	0.2495
Precision of PFs (T)	27.94	188.4	4.008	88.01
Koch	−0.6903	0.1198	−0.9302	−0.4610
Modeer	−0.3480	0.1028	−0.5468	−0.1432
Clark	−0.2493	0.0833	−0.4111	−0.0845
Tewari	−0.6719	0.1148	−0.9011	−0.4521
Bravo	−0.4271	0.0681	−0.5605	−0.2935
Skold	−0.5546	0.0783	−0.7094	−0.4027
Arruda	−0.4061	0.0674	−0.5381	−0.2737
Tagliaferro	−0.4273	0.1749	−0.7729	−0.0778
Milsom	−0.1199	0.1560	−0.4114	0.1968

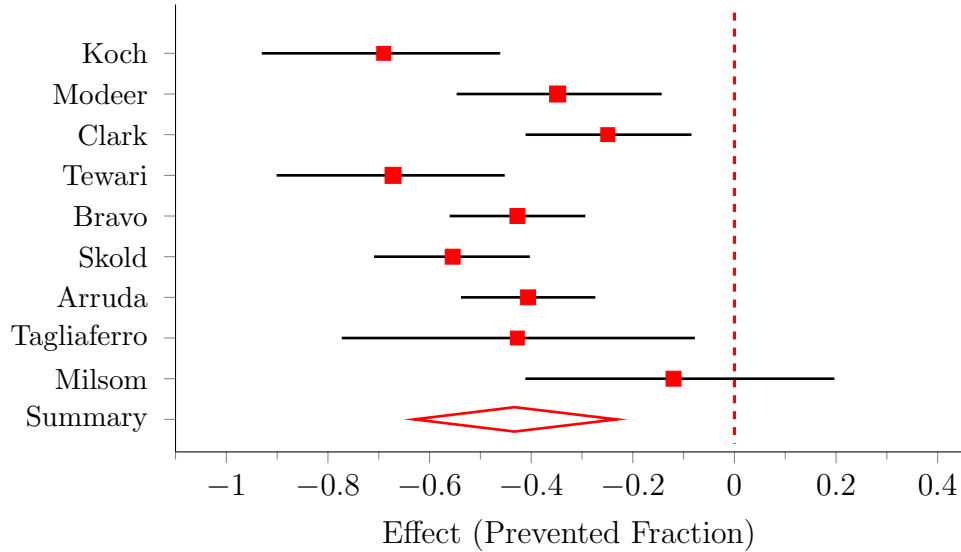


Figure 3: Meta-analysis on prevented fraction (Half-Normal prior): red squares are means; black lines are 95% CIs; red dashed line marks 0.

Interpretation of Results: The SPF is nearly unchanged relative to the Pareto prior, while the half-normal prior yields slightly larger posterior mass for moderate τ and somewhat wider credible intervals at the study level.

Model A Results: Figure 2 depicts posterior densities of the study-specific effects under the DuMouchel (Pareto) prior. The colored curves correspond to the individual study posteriors, while the bold line **purple** represents the posterior distribution of the overall summary effect (SPF). The Pareto and the half-normal prior obtained approximately similar SPF values, indicating consistent central effects across priors. However, as mentioned before, the Pareto prior yields slightly narrower heterogeneity, due to its heavier-tailed structure that allows modest shrinkage toward the common mean.

5.2 Model B

Across Uniform(0, 1), Beta(0.9, 1), and Beta(1, 0.9) priors on the variance ratio B , the pooled SPF remains close to -0.43 . Posterior mass for B concentrates at relatively high values, indicating substantial but not extreme heterogeneity. The parameterization through B is advantageous because it makes the contribution of between-study variability directly interpretable, which in turn facilitates transparent sensitivity analyses. Although the prior shapes differ, their influence is modest: posterior means and intervals for the pooled effect and for τ^2 change little across these specifications, and the substantive conclusions are unaffected.

5.3 Model C: Scaled Chi-Squared Prior on Variance of True Effect

Under the scaled Chi-Squared prior, the pooled effect remains stable, negative, and of similar magnitude, with evidence of moderate heterogeneity. Replacing the scaled Chi-Squared prior by the weakly informative AG prior increases posterior dispersion for τ^2 and slightly widens the intervals for study-level and pooled effects, reflecting the weaker regularization. Using the ag prior reduces tail sensitivity and yields greater numerical stability when information is sparse.

In all these cases, the pooled conclusions are robust, with posterior means for SPF consistently close to -0.43 and uncertainty bands excluding zero.

5.4 Comparative Discussion Across Models

Viewed jointly, Models A, B, and C lead to a coherent narrative. The estimated SPF concentrates near -0.43 in every specification, which means that the evidence for a clinically relevant reduction in caries incidence is not driven by prior choice. Differences arise mainly in how heterogeneity is encoded and how uncertainty is expressed. Model A yields slightly more concentrated posteriors under a half-normal prior on τ or τ^2 (SPF SD ≈ 0.099 ; $\mathbb{E}[\tau^2] \approx 0.074$) and a Pareto prior on τ or τ^2 (SPF SD ≈ 0.084 ; $\mathbb{E}[\tau^2] \approx 0.050$), which can be helpful when seeking stable interval estimates without sacrificing fit. Model B, built around the variance ratio $B = \tau^2/(\tau^2 + s_0^2)$, offers an immediately interpretable summary of between-study variation; the posteriors for B cluster at high values, typically around three quarters of total variability, and this feature is remarkably stable across uniform and skewed Beta priors. Uniform B gives SPF SD ≈ 0.087 with $\mathbb{E}[\tau^2] \approx 0.039$, while skewed Beta(0.9, 1)/Beta(1, 0.9) tighten further (SPF SD ≈ 0.077 – 0.078 ; $\mathbb{E}[\tau^2] \approx 0.039$ – 0.041). Model C, through scaled Chi-Squared, AG, or ag priors on the precision, accommodates broader tail behavior for example, scaled- χ^2 gives SPF SD ≈ 0.094 with $\mathbb{E}[\tau^2] \approx 0.065$, and makes the heterogeneity posterior more diffuse when the data are limited (such as Koch, Tewari, and Sköld studies which are strongly preventive, whereas the Milsom occasionally crosses 0, and the Tagliaferro shows the widest confidence intervals), indicating that prior choice modulates heterogeneity and precision.

At the study level, the patterns are similarly consistent. Most trials show credible intervals entirely below zero regardless of model, whereas the Milsom study yields an interval that overlaps zero in all specifications. This persistence across prior families suggests that local design features and background exposure patterns, rather than modeling artifacts, explain the neutral finding in that single trial. Partial pooling stabilizes smaller studies and reduces undue influence from extreme single-study estimates, yet the pooled effect remains essentially unchanged across all frameworks.

The informativeness of the priors affects interval width more than it affects point estimates. Diffuse choices, such as the AG prior, admit heavier tails for τ^2 and consequently widen credible intervals, which is appropriate when prior knowledge of heterogeneity is minimal. Moderately informative options, such as a half-normal prior on τ or the ag prior on the precision, achieve a practical balance, improving stability in sparse settings while leaving the central tendency of the pooled effect largely unaltered. In every case examined here, posterior means of SPF and study effects are driven by the data, and prior changes mainly reshape the uncertainty rather than the conclusion.

Relative to classical fixed-effect and random-effects analyses, the Bayesian hierarchy replicates the central finding while enriching its interpretation. The pooled estimate is closely aligned with the classical random-effects mean, yet the Bayesian approach provides full posterior distributions for pooled and study-specific effects, supports explicit modeling of uncertainty in τ^2 , and enables structured sensitivity analysis to prior assumptions. These features are especially valuable when the number of studies is modest, when designs are heterogeneous, or when decision makers require probability statements about clinically meaningful thresholds.

In practical terms, Model B is attractive for reporting because B transparently communicates the fraction of total variability attributable to between-study differences. When stability is a priority and prior information about heterogeneity is limited but not absent, the half-normal

prior on τ or the ag prior on precision offers a reasonable compromise. If a conservative stance is preferred in the face of sparse or noisy data, the AG prior provides wider intervals without shifting the pooled effect. Overall, the Bayesian results confirm the classical conclusions while adding a principled account of heterogeneity and uncertainty, and the main message remains unchanged: fluoride varnish confers a substantial and robust preventive benefit across diverse study settings.

6 Conclusions

This meta-analysis provides a coherent assessment of the effectiveness of fluoride varnish for preventing dental caries using complementary frequentist and Bayesian frameworks. Across fixed-effect, random-effects, and hierarchical Bayesian models, the estimated prevented fraction (PF) consistently indicates a clinically meaningful reduction in **DMF** surfaces, with a pooled effect near 43% and credible or confidence intervals excluding zero in aggregate. The Bayesian hierarchy yields full posterior distributions for study effects and the pooled PF, offering transparent uncertainty quantification while naturally accommodating between-study heterogeneity.

Sensitivity analyses demonstrate that the principal conclusions are robust to alternative and increasingly diffuse priors on the heterogeneity parameter τ^2 . Substituting DuMouchel’s Pareto prior with a half-normal prior on τ , or reparameterizing heterogeneity through the variance ratio B , yields similar pooled effects. Likewise, replacing the scaled Chi-Squared prior with AG or ag gamma priors alters the tail behavior of the heterogeneity posterior but leaves the central tendency of the pooled PF largely unchanged. Taken together, these results indicate that inference about the overall preventive benefit is data-driven rather than prior-driven within reasonable prior families.

Heterogeneity is nontrivial but moderate. Variation across trials reflects differences in design rigor, diagnostic thresholds, baseline risk profiles, and potential control contamination. Notably, one large, well-conducted study exhibits a posterior interval overlapping zero, consistent with contextual factors such as high background fluoride exposure and outcome sparsity. Even so, the pooled evidence remains stable under multiple modeling choices, emphasizing the external validity of a protective effect for fluoride varnish.

This work has practical implications. First, the magnitude of the pooled PF suggests clinically relevant reductions in caries burden in populations comparable to those studied. Second, the Bayesian framework provides decision-relevant posterior summaries that can support policy and guideline development, including probabilistic statements about achieving predefined prevention targets. Third, the analytic pipeline underscores the importance of reporting heterogeneity explicitly and examining prior sensitivity when synthesizing heterogeneous clinical evidence.

Limitations include variation in outcome definitions and measurement, potential publication bias, and incomplete harmonization of baseline risk and co-interventions across studies. These constraints motivate future research directions. Priorities include: individual-participant data meta-analyses to sharpen adjustment for covariates and explore effect modification; harmonized outcome definitions with calibrated diagnostic thresholds; routine posterior predictive checks and model comparison to assess fit; and prospective data collection that records fluoride exposure from all sources to contextualize effect sizes. Further, hierarchical modeling that jointly addresses cluster designs, contamination, and measurement error would refine inference on both pooled and setting-specific effects.

In summary, fluoride varnish is supported by convergent evidence as an effective caries-preventive intervention. The results are robust across modeling approaches and prior assumptions, and they provide a principled, probabilistic basis for clinical and public health decision making while highlighting avenues for methodological and substantive refinement in future studies.

Reproducibility: All plots, posterior analyses, and numerical results in this paper can be fully reproduced using the R and WinBUGS scripts available at: <https://github.com/saraantonijevic/Fluoride-Varnish-Meta-Analysis-Codes>.

References

- Arruda, A. O., Senthamarai Kannan, R., Inglehart, M. R., Rezende, C. T., & Sohn, W. (2012). Effect of 5% fluoride varnish application on caries among school children in rural brazil: A randomized controlled trial. *Community Dentistry and Oral Epidemiology*, 40(3), 267–276. <https://doi.org/10.1111/j.1600-0528.2011.00657.x>
- Bravo, M., Baca, P., Llodra, J. C., & Osorio, E. (1997). A 24-month study comparing sealant and fluoride varnish in caries reduction on different permanent first molar surfaces. *Journal of Public Health Dentistry*, 57(3), 184–186. <https://doi.org/10.1111/j.1752-7325.1997.tb02970.x>
- DerSimonian, R., & Laird, N. (1986a). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.
- DerSimonian, R., & Laird, N. (1986b). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
- DuMouchel, W. H. (1994). *Hierarchical bayes linear models for meta-analysis* (tech. rep. No. Technical Report No. 27). National Institute of Statistical Sciences. <https://www.niss.org/sites/default/files/technicalreports/tr27.pdf>
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515–533.
- Greenland, S., & Drescher, K. (1987). Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics*, 43, 865–872.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003a). Measuring inconsistency in meta-analyses. *BMJ*, 327, 557–560.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003b). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Klein, H., & Palmer, C. E. (1938). Sex differences in dental caries experience of elementary school children [From Child Hygiene Investigations, Division of Public Health Methods, National Institute of Health, U.S. Public Health Service]. *Public Health Reports*, 53(38), 1685–1690. <https://stacks.cdc.gov/view/cdc/72627>
- Marinho, V. C. C., Worthington, H. V., Walsh, T., & Clarkson, J. E. (2013). Fluoride varnishes for preventing dental caries in children and adolescents [Cochrane Oral Health Group]. *Cochrane Database of Systematic Reviews*, (7), CD002279. <https://doi.org/10.1002/14651858.CD002279.pub2>
- Milsom, K. M., Blinkhorn, A. S., Walsh, T., Worthington, H. V., Kearney-Mitchell, P., Whitehead, H., & Tickle, M. (2011). A cluster-randomized controlled trial: Fluoride varnish in school children. *Journal of Dental Research*, 90, 1306–1311. <https://doi.org/10.1177/0022034511422063>
- Modeer, T., Twetman, S., & Bergstrand, F. (1984). Three-year study of the effect of fluoride varnish (duraphat) on proximal caries progression in teenagers. *Scandinavian Journal of Dental Research*, 92, 400–407. <https://doi.org/10.1111/j.1600-0722.1984.tb00879.x>
- Normand, S.-L. T. (1999). Meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18, 321–359.

- Pateras, K., Nikolakopoulos, S., & Roes, K. C. B. (2021). Prior distributions for variance parameters in a sparse-event meta-analysis of a few small trials. *Pharmaceutical Statistics*, 20(1), 39–54. <https://doi.org/10.1002/pst.2053>
- Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd). Lippincott Williams & Wilkins.
- Sköld, U. M., Petersson, L. G., Lith, A., & Birkhed, D. (2005). Effect of school-based fluoride varnish programmes on approximal caries in adolescents from different caries risk areas. *Caries Research*, 39(4), 273–279. <https://doi.org/10.1159/000086182>
- Tagliaferro, E. P. S., Pardi, V., Ambrosano, G. M., Meneghim, M. C., Silva, S. R., & Pereira, A. C. (2011). Occlusal caries prevention in high and low risk schoolchildren: A clinical trial. *American Journal of Dentistry*, 24, 109–114.
- Tewari, A., Chawla, H. S., & Utreja, A. (1990). Comparative evaluation of the role of naf, apf and duraphat topical fluoride applications in the prevention of dental caries — a 2.5 year study. *Journal of the Indian Society of Pedodontics and Preventive Dentistry*, 8, 28–35.
- World Health Organization. (2022, November). Who highlights oral health neglect affecting nearly half of the world’s population. <https://www.who.int/news/item/18-11-2022-who-highlights-oral-health-neglect-affecting-nearly-half-of-the-world-s-population>