Reading the Xing et al. 2026 Network Meta-Analysis: Methods, Dose Definitions, and the Surrounding Literature

---

Question 1 — What the statistics actually do, in plain language

Xing and colleagues took 19 randomized trials enrolling 11,576 children and combined them through a chain of statistical machinery. Here is what each step does, stripped of jargon.

A regular ("pairwise") meta-analysis pools studies that compared the *same* two things — say, vitamin D vs placebo. **A network meta-analysis (NMA)** does more: it links together trials that compared *different* pairs of things (one trial of vitamin D vs placebo, another of zinc vs placebo, another of vitamin A vs zinc, and so on) into one connected "network." Once those direct links exist, the math lets you estimate every pairwise comparison — including comparisons that no single trial ever performed directly (e.g., high-dose vitamin D vs probiotics). The result is a single ranked table where every supplement can be compared to every other.

Bayesian methods are the engine used to fit that network. In the older "frequentist" approach, you compute a fixed point estimate and a confidence interval that reflects long-run sampling behavior. In the Bayesian approach, you start with a "prior" belief (here a vague, neutral prior because the authors had no reason to favor any one supplement) and then let the data update it into a "posterior" distribution that you can sample from. **JAGS** is the software that does the sampling; **MCMC (Markov Chain Monte Carlo)** is the sampling algorithm; the **4 chains × 50,000 iterations with 20,000 burn-in** simply means they ran four independent simulations of 50,000 draws each, discarded the first 20,000 of each (to let the chain "settle"), and used the remaining 120,000 draws to summarize the posterior. The reason this matters: the final "credible interval" is read off this simulated distribution, so it has to be long enough and stable enough to be trustworthy.

Odds ratio (OR) is the effect size. If a placebo group has 300 sick kids per 1000 and the treatment has 246 per 1000, the *odds* of being sick are 300/700 in placebo and 246/754 in treatment, and the OR ≈ 0.76. An OR of 1.0 means no difference; below 1.0 favors the supplement; above 1.0 favors placebo. **Credible interval (CrI)** is the Bayesian cousin of the confidence interval — a 95% CrI of (0.61, 0.94) means that, given the data and the model, there is a 95% probability the true OR lies in that range. Because the CrI for high-dose vitamin D (HVD) does not cross 1.0, the authors call this result statistically significant; every other pairwise comparison in their network crosses 1.0 and is therefore inconclusive.

SUCRA (Surface Under the Cumulative Ranking Curve) is a ranking summary. For each iteration of the simulation, the model produces a complete ranking of all eight interventions (which one performed best, second best, etc.). SUCRA averages those ranks into a 0–100% score: 100% would mean a treatment was always ranked best, 0% always worst, 50% means it landed in

the middle of the pack on average. HVD's 82.44% means it tended to top the league table across the simulations; vitamin A's 23.31% means it tended to land near the bottom. Crucially, SUCRA *ranks even when the underlying differences are not statistically distinguishable.* A treatment can score 65% without ever beating placebo with statistical significance — so SUCRA should always be read together with the CrIs.

DIC (Deviance Information Criterion) is a goodness-of-fit number used to compare alternative models — for example, "should I treat between-study differences as fixed or random?" Lower DIC = better fit penalized for complexity. It's the Bayesian analogue of AIC. The authors used it to justify their choice of a random-effects model.

Random-effects model assumes the true effect varies a bit from one trial to another (because populations, doses, latitudes, and seasons differ) and estimates both an average effect and the spread around it. A *fixed-effect* model would assume every trial is estimating exactly the same underlying effect, which is rarely realistic when you're combining trials done in Mongolia, Canada, and India with different doses and durations.

Node-splitting is the consistency check. The network can give two estimates of, say, "zinc vs probiotics": one straight from any direct head-to-head trials and one indirect, routed through placebo. Node-splitting asks whether those two estimates agree. If they disagree badly, the network's assumption that everything is comparable breaks down. The authors report no significant inconsistencies.

I^2 (I-squared) is the heterogeneity statistic. It is the percentage of variability across trials that is due to real differences in effect rather than random sampling noise. 0% means trials agree perfectly; 25% is low; 50% moderate; 75%+ high. The **I^2 of 76% in the zinc-vs-placebo direct comparison** is a red flag: the zinc trials disagreed with each other a lot, probably because they used different doses, durations, baseline zinc status, and age groups. That instability is part of why zinc's SUCRA (42.5%) sits in the middle even though some individual trials are very positive.

Absolute effect estimate. Odds ratios can mislead when the underlying disease is common. To translate the OR back to something tangible, the authors picked an assumed baseline risk of 300 cases of respiratory infection per 1000 children in the placebo group (a number borrowed from the pooled placebo arms of their included trials). Applying OR = 0.76 to those odds and converting back to a risk gives roughly 246 cases per 1000 — a reduction of **54 cases per 1000**, or equivalently a number-needed-to-treat of about 19. Change the assumed baseline (say, in a low-incidence setting at 100/1000) and the absolute benefit shrinks proportionally.

What the dose categorization actually accomplished. The authors split vitamin D into two "arms" of the network: HVD (>5600 IU/week, i.e. >800 IU/day) and LVD (⩽5600 IU/week, i.e. ⩽800 IU/day). Statistically, this turns a single intervention into two distinct nodes that the network treats as separate entities. The mathematical *consequence* is that any signal that benefits scale with dose ends up split between the two nodes; if the cut-point is too low (more on

this below), then the HVD bin contains a wide span of "anything above 800 IU/day," which may dilute or distort what an actual dose-response curve would look like. The authors explicitly concede that their threshold is a **"pragmatic analytical approach rather than a strict biological threshold"** — i.e., it was chosen because it produced two reasonably populated bins in their dataset, not because there is any biological reason 800 IU/day is a transition point.

Question 2 — What "high-dose vitamin D" actually means in this paper

In Xing et al., **HVD = anything above 5,600 IU per week**, which works out to **more than ~800 IU per day**. That is the entire definitional content. The authors say it explicitly: this is a statistical convenience, not a pharmacologic or biological cut-point.

The trials that populate the HVD node in this network are the following:

Trial (year)	Country	Dose used	Equivalent daily dose
Aglipay (2017)	Canada	14,000 IU/week	2,000 IU/day
Loeb (2019)	Vietnam	14,000 IU/week	2,000 IU/day
Morris (2021)	Canada	28,000 IU/week	4,000 IU/day
Reyes (2024) HVD arm	Mexico	11,200 IU/week	1,600 IU/day
Somnath (2017)	India	10,000 IU/week	~1,430 IU/day

The Reyes 2024 trial is particularly telling because it had an internal contrast: an "LVD" arm at 5,600 IU/week (800 IU/day) and an "HVD" arm at 11,200 IU/week (1,600 IU/day). Both fall squarely inside what most vitamin D investigators would call physiological replacement dosing — neither is what the clinical literature means by "high dose."

By the standards of the vitamin D research community, **"high-dose vitamin D" usually refers to one of:**

- **≥4,000 IU/day** sustained daily dosing (the U.S. Institute of Medicine's Tolerable Upper Intake Level for children >9 and adults; routinely exceeded in trials like VITAL, D-Health, and CORONAVIT);
- **Bolus dosing of 50,000 IU or more** (single or weekly — the dose used in classic Heaney pharmacokinetic studies, in many of Hollis & Wagner's pregnancy work, and in numerous COVID-era trials);
- **Pharmacologic replacement** for deficiency, typically 50,000 IU/week for 6–8 weeks, sometimes 100,000 IU bolus monthly.

Against that yardstick, the so-called "high-dose" arm of the Xing meta-analysis is, with one exception, **modest physiological supplementation**. Four of the five HVD trials sit at 1,400–2,000 IU/day; only Morris 2021 (4,000 IU/day) reaches the lower bound of what most clinicians would call "high." The Aglipay 2017 trial, often cited as a "high-dose pediatric trial," literally compared 2,000 IU/day against a 400 IU/day control — both intakes that the Endocrine Society's guidelines would treat as preventive, not pharmacologic.

Why this matters for interpreting the OR = 0.76 result. Lumping 800 IU/day with 4,000 IU/day into a single category obscures any genuine dose-response signal. If the benefit is truly concentrated at 4,000 IU/day, the HVD pool is being diluted by the lower-dose trials and the true high-dose effect is being understated. If the benefit instead saturates near 1,000–2,000 IU/day, the HVD pool is correctly capturing a plateau but mislabels the relationship. Either way, the binary HVD/LVD split cannot distinguish these scenarios. This is precisely the kind of question that demands either individual-participant-data dose-response modeling (as Jolliffe and Martineau have pursued) or, at minimum, a finer-grained dose categorization (e.g., 800–2,000, 2,000–4,000, >4,000 IU/day equivalent).

It is also worth flagging that several of the included HVD trials (Aglipay, Reyes) used a *lower-dose vitamin D control*, not a true placebo. In the present NMA, those low-dose arms are nominally categorized as LVD, and the HVD-vs-placebo result is therefore generated partly through the network rather than from head-to-head HVD-vs-placebo data — a perfectly legitimate NMA construction, but one that should be remembered when interpreting the OR.

Question 3 — How this paper fits with the broader literature

3a. Other reviews of vitamin D and pediatric respiratory infection

Martineau et al., BMJ 2017 (the IPD meta-analysis that defined the field). Twenty-five RCTs, 10,933 participants of all ages. Vitamin D supplementation reduced acute respiratory tract infections overall: OR 0.88 (95% CI 0.81–0.96). The protective effect was *strongest at low baseline 25(OH)D* (<25 nmol/L: OR 0.30, 95% CI 0.17–0.53) and weaker at higher baseline (≥25 nmol/L: OR 0.75, 95% CI 0.60–0.95). Daily or weekly dosing worked; large infrequent boluses did not. This is the result that anchored a decade of clinical interest.

Jolliffe et al., Lancet Diabetes & Endocrinology 2021 (aggregate-data update). Forty-three RCTs, 48,488 participants. Overall OR 0.92 (95% CI 0.86–0.99) — modest, but still statistically significant. Subgroup signals consistent with the 2017 IPD analysis: protective effects concentrated in daily dosing, in doses of 400–1,000 IU/day, and in regimens lasting ⩽12 months.
Endocrinologyadvisor

Jolliffe et al., Lancet Diabetes & Endocrinology 2025;13(4):307–320 (stratified aggregate-data update — published February 21, 2025). This is the major recent update that Xing's paper should

be read against. Six new RCTs were identified (19,337 participants), of which three (16,085 participants) provided aggregate data; these were pooled with the 43 prior trials, for **40 studies and 61,589 participants in the primary analysis**. The headline result reversed the 2021 conclusion: **OR 0.94 (95% CI 0.88–1.00), p = 0.057, I^2 = 26.4%** — i.e., the 95% CI now just crosses 1.00 and the effect is no longer statistically significant. Pre-specified subgroups by **age (including children), baseline 25(OH)D, dosing frequency, and dose size** showed **no effect modification**, which is a striking shift from the 2017 IPD findings. A funnel plot showed left-sided asymmetry (Egger p = 0.0020), pointing to possible small-study/publication bias. Serious adverse events were unchanged (OR 0.96, 95% CI 0.90–1.04). The Lancet authors interpret this as "no statistically significant protection," and the linked editorial framed it as the closest thing yet to a definitive null answer. PubMed + 5

The contrast with Xing et al. is sharp: Xing's NMA, restricted to pediatric trials and using a different statistical apparatus, finds a significant HVD-vs-placebo benefit (OR 0.76, 95% CrI 0.61–0.94) at the very moment the largest available aggregate-data analysis finds no effect overall and no signal in children specifically. The most plausible explanations are (i) different trial inclusion sets, (ii) Xing's restriction to pediatric data, (iii) the dichotomized HVD vs LVD splitting (which Jolliffe 2025 found no support for), and (iv) network-derived borrowing of strength across nodes.

Wang L. et al., BMC Pediatrics 2025;25:985 (cited as ref 45 by Xing). Seventeen RCTs, 18,372 pediatric participants, searched to April 2025. Overall vitamin D vs placebo: **RR 0.82 (95% CI 0.64–1.06) — null**. The interesting subgroup finding ran in the opposite direction from Xing: **daily low-dose regimens of ⩽1,000 IU/day showed benefit (RR 0.42, 95% CI 0.27–0.66), while higher doses did not.** Hospitalization duration, recovery time, and all-cause mortality were unaffected. Serious adverse events were modestly reduced (RR 0.83, 95% CI 0.71–0.96). This finding — that the *lower*-dose daily regimens, not high-dose, drove the protective signal — directly contradicts Xing's HVD-dominance conclusion and is more aligned with the 2017 Martineau IPD subgroup analyses. PubMed Central + 2

Jolliffe et al., EClinicalMedicine 2025;90:103682 (with corrigendum, January 2026; cited by Xing as ref 46). Aggregate-data meta-analysis of RCTs of maternal vitamin D supplementation in pregnancy with offspring ARI as a prospectively pre-specified outcome, search to August 5, 2025. Sub-group analyses were planned by maternal baseline 25(OH)D (<25, 25–49.9, 50–74.9, ⩾75 nmol/L). This is the first up-to-date synthesis of maternal-pregnancy vitamin D for offspring respiratory protection — a question Xing's pediatric NMA cannot answer because none of its included trials supplemented mothers. nih + 2

Patchen et al., Cochrane Database Syst Rev 2025, Issue 8, CD013396 (cited by Xing as ref 18). The Cochrane review of vitamin D in pregnancy, breastfeeding, or early childhood for *asthma* prevention. Key result: **high-dose vs low/standard-dose vitamin D in pregnant or breastfeeding women likely reduces childhood wheeze (RR 0.79, 95% CI 0.64–0.98; 3 studies, 1,439 participants; moderate-certainty evidence)**. The asthma signal was weaker (RR 0.81, 95% CI 0.63–1.04). Early-childhood vitamin D had little effect. This Cochrane review

reinforces the principle that maternal high-dose timing may matter more than postnatal pediatric supplementation for downstream airway outcomes — a different but adjacent question. Cornell + 4

Fang et al., Frontiers in Nutrition 2023 (pediatric meta-analysis of healthy children). Four trials, 13,367 participants. Vitamin D vs placebo: OR 0.98, 95% CI 0.90–1.08 — null. Notably, this dataset included Ganmaa 2020 (Mongolia, 14,000 IU/week × 3 years, ~8,800 children) and Holland 2012 (Afghanistan, 100,000 IU bolus dosing), neither of which showed protective effects in their primary endpoints.

3b. Network meta-analyses comparing multiple supplements

The Xing 2026 paper is, as far as the published record shows, **the first network meta-analysis explicitly comparing vitamin D, vitamin A, zinc, iron, probiotics, and a VA+Zinc combination for respiratory-infection prevention in children.** The closest analogues in scope are:

- **A 2025 EClinicalMedicine network meta-analysis in adults** (Lancet group) that compared low/moderate/high-dose vitamin C, standard- and high-dose vitamin D, vitamin E, zinc, flavonoids, probiotics, and synbiotics across **107 RCTs, 101,751 participants** — the authors explicitly noted that no prior NMA had compared the full range of nutritional interventions for RTI prevention. The Lancet
- **Amati et al., 2017 (Journal of Pediatric Gastroenterology and Nutrition)** — a network meta-analysis of probiotics for respiratory infection prevention in children and adolescents (21 trials, 6,603 participants) that ranked individual probiotic strains rather than supplements. *Lactobacillus casei rhamnosus* had a 54.7% probability of ranking first. PubMed
- **Milani et al., Allergy 2024 (EAACI taskforce systematic review)** — a non-network systematic review of nutritional and dietary interventions for *treatment* (not prevention) of pediatric ARI, 50 studies. The meta-analysis found a marginal effect of zinc on length of hospitalization for pneumonia but no effect of vitamin D in that context, and concluded that no single supplement could be routinely recommended. Wiley Online Library

So Xing et al. is genuinely the first pediatric-prevention NMA across this supplement set — that is a real contribution, even though the conclusions are constrained by the dose-categorization issue described above.

3c. Cochrane and other single-supplement reviews

Zinc. Lassi et al., Cochrane 2016 (CD005978) — 6 trials, 5,193 children aged 2–59 months. Zinc reduced pneumonia incidence by 13% (RR 0.87, 95% CI 0.81–0.94, fixed-effect, low-quality GRADE) and prevalence by 41% (one trial, low quality). Effects were stronger when pneumonia was radiologically confirmed (RR 0.79, 95% CI 0.71–0.88). A 2023 Cochrane update was more equivocal. The high heterogeneity (I^2 = 76%) that Xing flags in the zinc direct comparison is

consistent with what the zinc literature has always shown: results depend heavily on baseline zinc status, age, and case definition.

Vitamin A. Imdad et al., Cochrane 2022 update (CD008524) — 47 trials, 1.2+ million children. Overall mortality benefit (RR 0.88) and diarrhea-mortality benefit (RR 0.88) confirmed, but **no significant effect on respiratory-disease mortality (RR 0.98, 95% CI 0.86–1.12; low-certainty evidence).** A discontinued protocol (Chen 2021, CD014847) for vitamin A in pediatric lower-RTI specifically noted that prior reviews showed no positive effect on existing respiratory disease, and Grotto's classic 2003 review actually flagged that vitamin A supplementation modestly *increased* respiratory-infection incidence in non-deficient children. Yuan et al. (Nutrients 2021) reached the same conclusion: excessive vitamin A in vitamin-A-replete populations increases ARI risk. The Cochrane 2023 review of oral vitamin A specifically for acute upper RTI prevention (CD015306) found "uncertain" effects across three studies and 22,668 participants (RR 1.00, 95% CI 0.92–1.09). This makes the Xing SUCRA finding (vitamin A near the bottom at 23.31%) biologically plausible and consistent with the broader literature. Cochrane + 2

Probiotics. Zhao et al., Cochrane 2022 update (CD006895) — 23 trials, ~6,000+ participants. Probiotics reduced the proportion of participants experiencing at least one acute URTI vs placebo (the prior Hao 2015 review reported OR 0.43, 95% CI 0.29–0.63 for the pediatric subgroup specifically). The 2022 update softened the certainty but maintained the direction. Xing's middling probiotic SUCRA (52.49%) and non-significant OR are weaker than what the Cochrane review reports — likely because the children-only subset is smaller and because probiotic strain effects (which Amati 2017 demonstrated vary substantially across strains) are collapsed in Xing's lumped "probiotics" node. Cochrane Library

3d. Key contrasts and the dose-response question

The most important contrast with Xing's headline finding is the **Jolliffe 2025 Lancet Diabetes & Endocrinology update**, which now finds **no statistically significant overall protection** (OR 0.94, 95% CI 0.88–1.00) and — crucially for the dose-categorization question — **no evidence of effect modification by dose size** when assessed pre-specified across the full pooled dataset. Wang 2025 in BMC Pediatrics goes further and finds that the pediatric protective effect is concentrated at ⩽1,000 IU/day daily dosing, not at higher doses. Martineau's earlier 2017 IPD analysis similarly identified 400–1,000 IU/day as the favorable subgroup. The clinical-trial literature has thus repeatedly suggested that a daily, modest dose in a deficient population is more effective than larger or bolus dosing — the opposite hierarchy from what Xing's HVD-vs-LVD split implies. ResearchGate + 2

To the user's likely question — **has anyone done a proper dose-response analysis?** — the answer is partly yes and partly no. The Jolliffe/Martineau IPD analyses have explored dose categorically (400–1,000, 1,001–2,000, >2,000 IU/day equivalents) but have not modeled a continuous dose-response curve with random splines or non-linear modeling at the participant level. No published pediatric-only IPD dose-response analysis exists, and the closest dose-

response statement in the 2025 Lancet aggregate update is the explicit subgroup null. A genuine dose-response IPD analysis — restricted to children, with continuous daily-equivalent dose, modeled against 25(OH)D achieved, and stratified by baseline status — remains the outstanding methodological need in this field, and it is precisely the kind of analysis that Xing's network meta-analysis cannot substitute for.

Bottom line for VitaminDWiki context. Xing et al. 2026 is a useful first attempt to rank pediatric supplements against each other, and its HVD = top-ranked finding (SUCRA 82.44%, OR 0.76 vs placebo) makes a clean headline. But that headline rests on (a) a dose threshold the authors themselves call pragmatic; (b) an HVD pool that, by the standards of the vitamin D field, mostly contains *modest* daily-equivalent doses (1,400–2,000 IU/day with one 4,000 IU/day outlier); and (c) a divergence from the much larger and methodologically stronger 2025 Jolliffe aggregate-data update that found no overall effect and no dose-size signal. The Wang 2025 BMC Pediatrics analysis — the closest direct comparator — actually finds the protective effect at *low* daily doses rather than at the higher end. Readers should treat Xing's HVD ranking as hypothesis-generating against a literature that, taken as a whole, points either toward a modest daily-low-dose effect concentrated in deficient children (the Martineau-era view) or toward no overall effect (the current Jolliffe-2025 view) — but not toward a "higher is better" gradient that Xing's binary cut-point implies.