Annex to:

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), Turck D, Bresson J-L, Burlingame B, Dean T, Fairweather-Tait S, Heinonen M, Hirsch-Ernst KI, Mangelsdorf I, McArdle HJ, Naska A, Nowicka G, Pentieva K, Sanz Y, Siani A, Sjödin A, Stern M, Tomé D, Van Loveren H, Vinceti M, Willatts P, Fewtrell M, Lamberg-Allardt C, Przyrembel H, Arcella D, Dumas C, Fabiani L, Martino L, Tomcikova D, and Neuhäuser-Berthold M, 2018. Scientific opinion on the update of the tolerable upper intake level for vitamin D for infants. EFSA Journal 2018;16(7):5365, 25 pp. doi:10.2903/j.efsa.2018.5365

Annex A – Statistical methods used to estimate the intake-response of serum 25(OH)D concentration on daily supplemental intake of vitamin D and to derive the percentage of infants exceeding a serum 25(OH)D concentration

The objective of the statistical analysis was to characterise the dose-response relationship between the exposure to 'high' intake levels of vitamin D in the healthy population of infants (aged 0 to 12 months) and achieved serum concentrations of 25(OH)D. The Panel considers 200 nmol/L to be a serum concentration of 25(OH)D below which it is unlikely that adverse effects (hypercalciuria, hypercalcaemia, nephrocalcinosis, abnormal growth patterns) would occur in infants (Section 3.3.6.5. of the scientific opinion). The analysis was based on the data collected during a systematic literature review (Section 3.1. of the scientific opinion). This analysis is described in brief in Section 3.5. of the scientific opinion and in more details in the present Annex. The steps which were followed in the statistical analysis are described in detail in the following sections:

- ✓ A **meta-analytical mixed-effect model** was set up to explain the relationship between **supplemented vitamin D intake and study-arm mean serum 25(OH)D concentration**. Background intake from food was not considered since it was seldom measured in the retrieved studies. The Panel considered that this leads to an underestimation of the true intake corresponding to potential adverse effects and concluded that this was acceptable since leading to a conservative UL estimate;
- ✓ The model was adjusted for a set of explanatory factors (fixed effects) and a set of factors explaining the hierarchical structure in the data (random effects);
- ✓ The **distribution of the study-arm mean achieved serum 25(OH)D concentration under realistic combinations of vitamin D intake and other explanatory factors** was simulated based on the predictive meta-analytical mixed-effect model previously set;
- ✓ **Individual responses were simulated** for each mean response value predicted by the model under the assumption that a truncated normal distribution describes the variability of an individual response around the study population mean. Since inter-individual variability was unknown, it was estimated using within study variability extracted from each study-arm-measurement occasion;
- ✓ The simulated individual distribution of serum 25(OH)D was **stratified by classes of vitamin D intake** (between 5 and 50 µg/day with a step size of 5 µg)**, baseline concentration of the biomarker** (serum 25(OH)D below 30 nmol/L; 30–60 nmol/L; 60–90 nmol/L) **and age class** (below and above 6 months of age). For each group defined by age-dose-baseline, the percentages of infants expected to exceed a pre-defined concentration of the biomarker were computed. To address the uncertainty surrounding such concentration (Section 3.3.6.1. of the scientific opinion), two concentrations (150 and 200 nmol/L) were considered in order to investigate how this would change the results of the meta-regression analysis.

1.1. Mixed effect meta-regression model: dose-response relationship of study-arm mean serum 25(OH)D on vitamin D intake

Among the possible meta-analytical approaches, the meta-regression has the advantage of permitting the assessment of the **influence of a set of explanatory variables** when exploring the relationship between the exposure to a potential hazard and an effect (van Houwelingen et al., 2002). This allows explaining at least part of the total heterogeneity among studies.

1.1.1. Dose-response approach

First, this analysis refers to a nutrient, considerations about balancing risk of inadequacy and risk of adverse effects are needed. Secondly, the usual toxicological approach of setting the effect of concern (the so-called critical effect or Benchmark Response (EFSA Scientific Committee et al., 2017)) based on the definition of a threshold for the Relative Risk (e.g. risk ratio, odds ratio) is not necessarily applicable for a nutrient and related biomarker(s), for which e.g. absolute 'thresholds' might also be biologically relevant.

1.1.2. Model assumptions: Normality, homoscedasticity and linearity

Normality, uniformity of the residual variance across doses (i.e. homoscedasticity) and linearity are standard assumptions in regression and meta-regression analysis (Viechtbauer, 2010a). Visual inspection of the distribution of the response and the residuals can help identifying important deviations from these assumptions.

The response variable – the serum 25(OH)D concentration - is assumed to come from a population that is **normally distributed**. This assumption was tested both graphically and using formal testing (Shapiro and Wilk, 1965). The normality of the response was assessed both on the original scale and on the natural log-transformed scale (ln-scale in the following).

QQplot on the original scale (Figure 1.a) shows **deviations from normality in the right tail** of the distribution providing indication of some right skewness. Deviations from normality are mainly **resolved** (i.e. dots are better contained into the dotted band after ln-transformation of the response) **when moving to the ln-scale** as it is reasonable to expect in these cases (Figure 1.b).

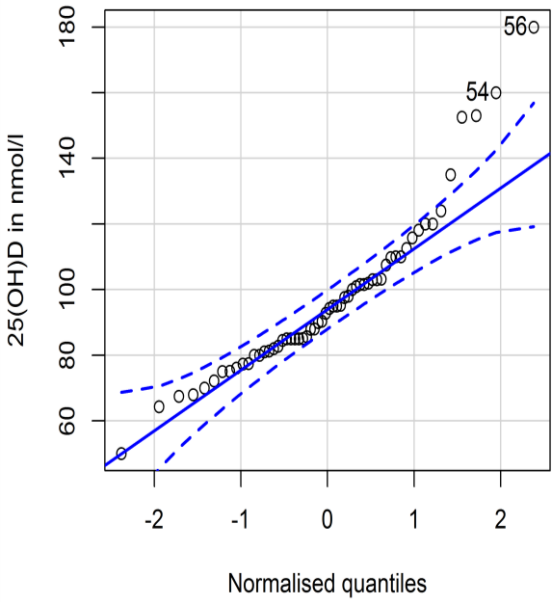

Figure 1: a. Qqplot of data in the original scale

Shapiro Wilk normality test
W = 0.91099, p-value = 0.0004293*
*hypothesis that sample was drawn from a normally distributed population can be rejected with prob<0.05

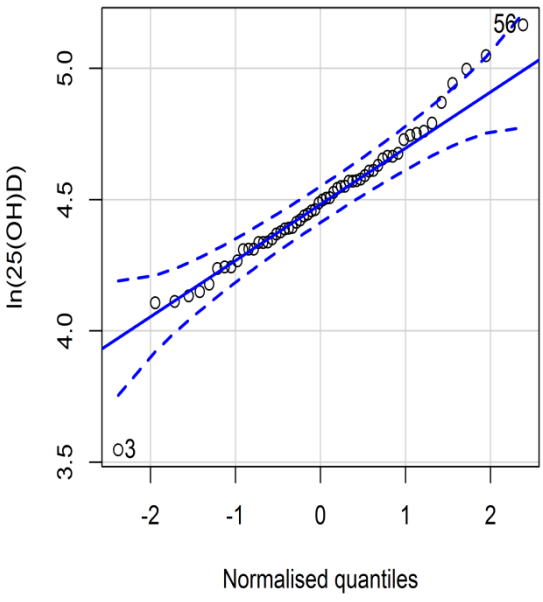

Figure 1: b. Qqplot of data in ln-transformed scale

Shapiro Wilk normality test
W = 0.96274, p-value = 0.0722**
**hypothesis that the sample was drawn from a normally distributed population cannot be rejected with prob<0.05

A visual inspection of the unadjusted relationship of serum 25(OH)D concentration on vitamin D intake showed that **linearity** might fit relatively well the data **except** at high vitamin D intake (e.g. 40 µg/day), where most of the points systematically lay above or below the regression line (Figure 2).

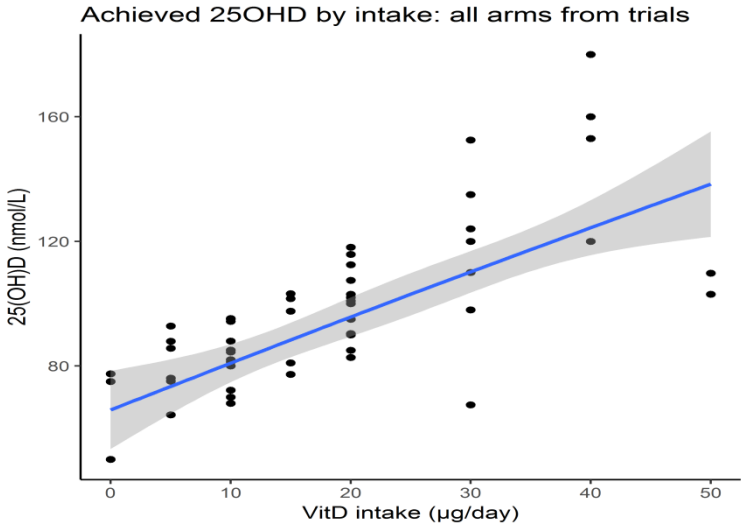

Figure 2: Unadjusted intake-response relationship (no moderator variables) – original scale

Figure 2 indicates mean 25(OH)D response (original scale) in each arm (black dots) of the various studies at different levels of vitamin D intake, the blue line is the fitted line of the mean response, the grey band is the confidence interval around the mean.

Figure 2 also displays some deviation from the assumption of constant variance across doses, higher variability being present for doses between 30 and 40 µg/day.

The ln-transformation of the serum 25(OH)D response is expected to **improve the approximation to a normal distribution and reduce the impact of lack of homoscedasticity. A ln-transformation of the explanatory variables** (specifically of the vitamin D intake and baseline concentration) can improve the **linear** fit considering that a **ln-transformation** of the response could make relationship deviating from linearity. Figure 3 shows the unadjusted intake response relationship of ln-transformed serum 25(OH)D concentration (ln-25(OH)D) on ln-transformed vitamin D intake (ln(VitD intake)). The approach proposed by Higgins et al. (2008) was used to ln-transform study-arm mean and standard deviation values.

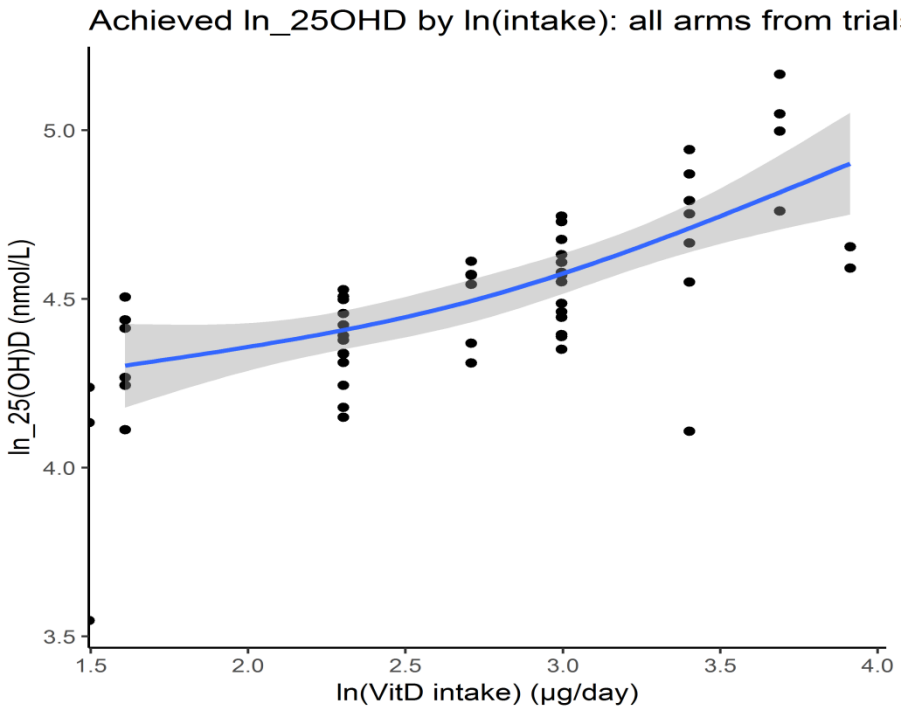

Figure 3: Unadjusted intake-response relationship (no moderator variables) – ln-scale for response and intake

In order to assess the influence of the choice of the scale (original or ln-transformed) and the related uncertainty on the simulated study-arm means of the achieved serum 25(OH)D concentration, **both scales have been considered** for response and intake in the models described below.

1.1.3. Selection of the explanatory variables

The background intake of vitamin D from diet (i.e. vitamin D from formulae and other foods for infants, fortified and not fortified) was rarely measured/reported in the studies. Therefore, the intake-response relationship was established only on the basis of the additional dose of vitamin D provided (trials), which was always **through a supplement** (and not a fortified food) in the dataset used (Section 3.5.1. of the scientific opinion). This was done considering that difference of bio-availability of vitamin D when supplemented, naturally present or added to food could be considered limited, as only scarce data on this aspect is available (EFSA NDA Panel, 2016) (Section 7. of the scientific opinion).

A series of factors were identified as **potential confounders/moderators** that could be able to modify either the response (serum 25(OH)D concentration) or both the response and the exposure (vitamin D intake) (Sections 3.2.3. and 3.5.2.1. of the scientific opinion). They included:

- Serum 25(OH)D baseline concentration;
- Latitude;
- Feeding type at start;
- Body weight/age
- Categories of duration of gestation;
- Supplementation duration;
- Vitamin D form (D_2 versus D_3);
- Analytical method used to measure serum 25(OH)D concentration.

Transformation and re-categorisation of some of these variables are described in Table 5 of the scientific opinion.

A **visual investigation** (Figure 4) was performed in order to identify factors that, based on data, might have a stronger impact on the intake-response relationship of serum 25(OH)D concentration on vitamin D intake. Some variables showed a potential for interaction with vitamin D intake level (e.g. supplementation duration). Because of the limited size of the sample and the need to balance complexity and interpretability of the results (i.e. parsimony principle), it was decided to include only the main effects in the model and not the interactive ones.

A graphical investigation of the dose-response by concentration of the biomarker serum **25(OH)D at baseline** (Figure 4.c) highlighted that higher values are achieved when the concentration at baseline is higher. The effect of the initial concentration is more evident at more extreme vitamin D intake levels (below 10 and above 30 µg/day). The variable was included by default in the model for biological reasons (Section 1.8.1. of the scientific opinion). It logically replaced the intercept from the model that therefore was eliminated.

Figure 4.a depicts the intake-response relationship stratified by **classes of latitude**: class 3 corresponding to countries located above 50° parallel (North or South), class 1 for countries closer to the equator line (between 40°South and 40° North), class 2 for the countries in between (Table 5 of the scientific opinion). The results indicate higher concentrations of serum 25(OH)D for infants living in northern countries. This would conflict with the expectation of a lower endogenous vitamin D production at higher latitudes. A possible justification for this result was that the latitude was masking other factors favouring higher concentrations of serum 25(OH)D in northern countries (e.g. country specific practices for maternal and infantile vitamin D supplementation) and/or infant sun exposure was too limited to expect an effect on the biomarker (Section 1.7.1.1. of the scientific opinion). **The Panel decided to discard this factor from the model since it was lacking biological relevance.**

The intake-response relationship stratified by **type of feeding at the start of the study** (Figure 4.b) shows that achieved serum 25(OH)D concentration is higher for infants receiving a mixed feed at the start of the study as compared to exclusively breastfed infants. This hierarchy is inverted for low levels of vitamin D intake (up to 10 µg/day). The Panel considered that feeding status can change quickly at this stage of life and observations taken only at baseline are not much indicative of the feeding type in the following weeks. Therefore, the **Panel decided to discard this variable from further analysis.**

The Panel discussed whether mean body weight or mean age was more relevant to explain the achieved mean concentration of serum 25(OH)D. The two variables were highly correlated ($r = 0.99$) in the available body of evidence. Therefore, from a statistical perspective, the inclusion of both of them in the model would have not been advisable. Indeed high correlation among fixed effects in a model (known as multicollinearity issue) can increase the variance of the coefficient estimates and make the estimates very sensitive to minor changes in the model. Eventually, **age** was selected

because always reported for the study participants, whereas body weight was sometimes missing. The possibility to obtain model predictions stratified by age classes was also considered a plus.

The intake-response relationship does not highlight a clear pattern for different **durations of gestation** (i.e. groups of only full-term infants vs groups of mixed/unclear/unspecified duration of gestation, Figure 4.d). At extreme doses full terms infants are over-performing, the opposite is observed at intermediate doses.

Some differences in the shape of the dose-response relationship and in the achieved concentration of serum 25(OH)D are identified at different **supplementation durations** (Figure 4.e), '6-months' being the duration with the highest study-arm mean concentrations at the highest doses (above 40 µg/day).

Comparison of the intake-response relationship for **vitamin form** D_2 and D_3 (Figure 4.f) could not be performed since study-arms eventually included in the body of evidence after exclusion of infants with rickets and high risk of bias studies enclosed only one study administering vitamin D_2.

Only two **analytical methods** (Figure 4.g) were used in the studies finally included in the body of evidence (LC-MS/MS and RIA). The measurements provided by the RIA method were higher for doses up to around 35 µg/day. The reverse occured for higher doses.

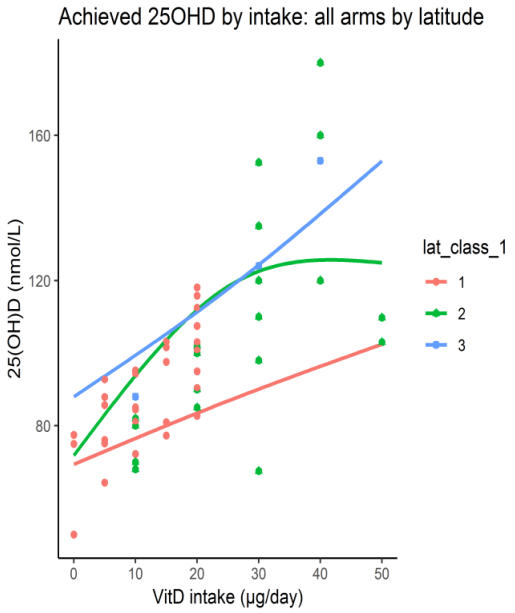

Figure 4.a: intake-response relationship by latitude

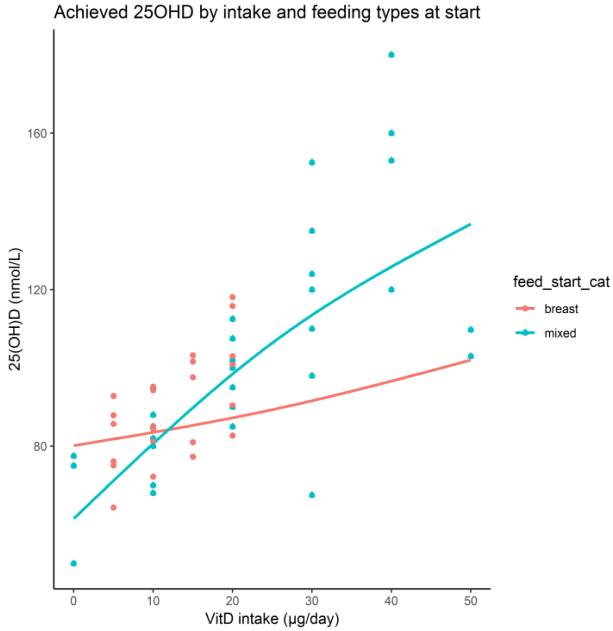

Figure 4.b: intake-response relationship by feeding type at start

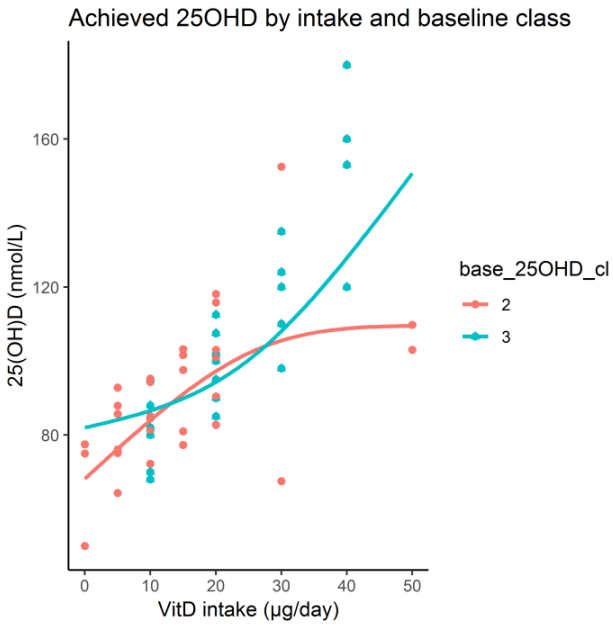

Figure 4.c: intake-response relationship by baseline serum 25(OH)D concentration

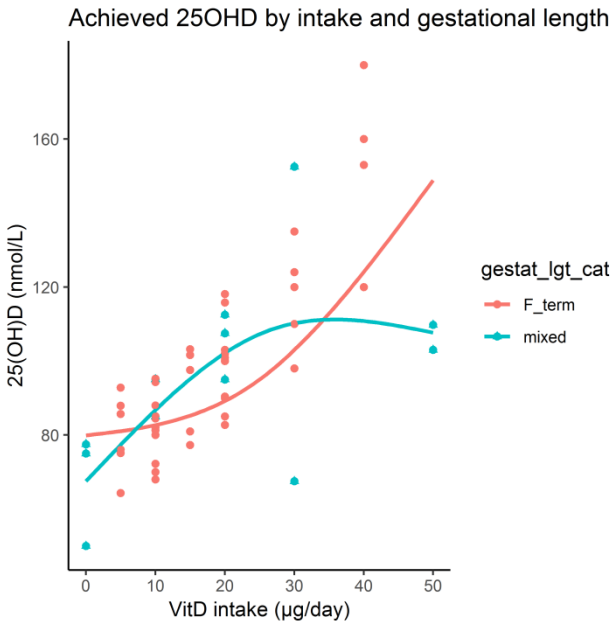

Figure 4.d: intake-response relationship by category defined by duration of gestation

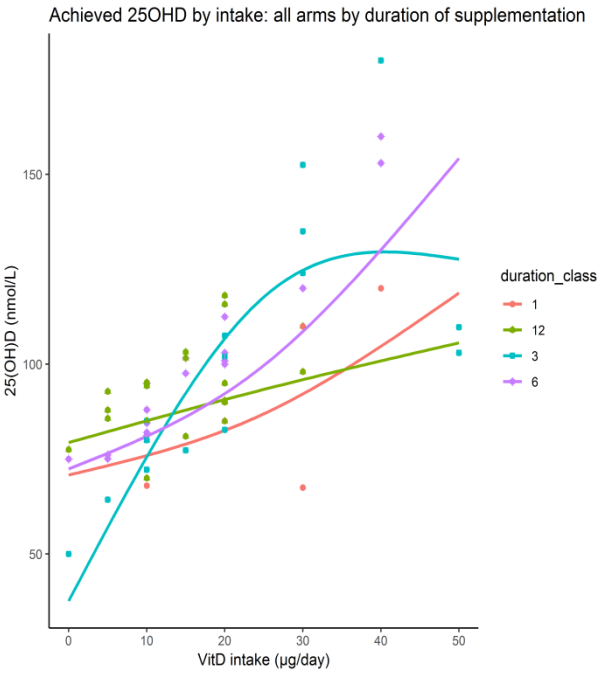

Figure 4.e: intake-response relationship by supplementation duration

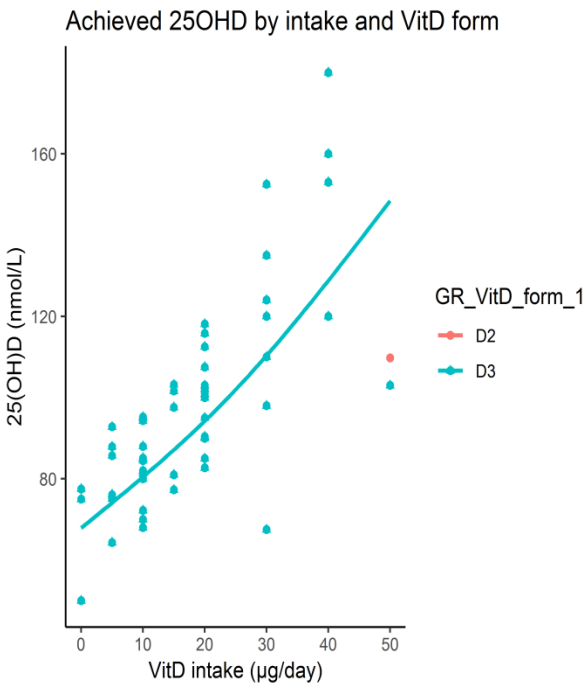

Figure 4.f: intake-response relationship by vitamin D form

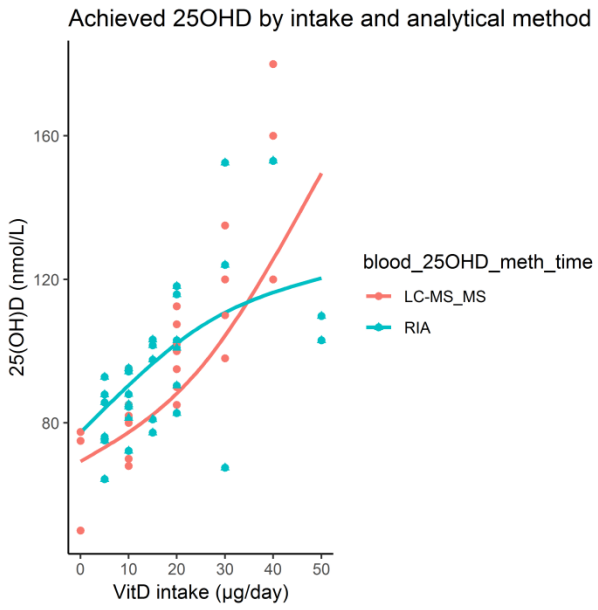

Figure 4.g: intake-response relationship by analytical method

Figure 4: Achieved serum 25(OH)D concentration on vitamin D intake by some potential moderators

*Categories for baseline serum 25(OH)D, latitude and duration class are reported in table 5 of the scientific opinion

1.1.4. Identification of the best predictive model

In order to explain the effect of moderator variables on the intake-response relationship to better explain heterogeneity and to account for the hierarchical structure in the data, a **mixed effect meta-regressive model** was used as suggested by van Houwelingen et al. (2002).

It includes both **fixed effects** and **random effects**:

- the fixed effects being variables that have an influence on the achieved concentration of the biomarker (serum 25(OH)D concentration) and have an impact on the value of the predictions,

- the random effects reflecting the correlation structure in the data.

As in any mixed effect model, the random components are assumed to have zero mean and therefore to contribute only to explain variability and heterogeneity in the data. They are intended to account for lack of independence in the data attributable to the fact that several arms (dose groups) are analysed for each individual trial and repeated measurements might be taken on the same arm. A compound symmetry structure is considered to formalise the hierarchy in the data. It assumes that the level of correlation is the same for all possible couples of observation (multiple arms in a study, multiple observations on the same arm). This approach conveniently allows reducing the number of parameters to be estimated as for the random components.

Four nested models were set up with **increasing number of fixed effects** (except the fourth).

- **Model 1** is the so-called '**null-model'** based on the assumption of no intake-response relationship (constant mean achieved serum 25(OH)D concentration across levels of vitamin D intake).
- **Model 2** includes only vitamin D intake and considers the baseline serum concentration of 25(OH)D instead of the intercept.

- **Model 3** adjusts the response of serum 25(OH)D concentration on the vitamin D intake for the baseline value and a series of **moderators that include: duration, analytical method, duration of gestation, latitude, age**.
- **Model 4** retains only **a subset** of the fixed factors in model 3.

Table 1: Goodness of fit indicators

Model	Model fixed effects	logLik	-2logLik	AIC	AICc	BIC
1	Intercept	-259.25	518.49	528.49	529.67	538.71
2	vitD intake, baseline serum 25(OH)D, no intercept	-250.45	500.90	512.90	514.61	525.05
3	vitD intake, baseline serum 25(OH)D, duration, analytical method, duration of gestation, age, no intercept	-214.51	429.02	455.02	465.42	479.62
4	vitD intake, baseline serum 25(OH)D, duration, age, no intercept	-223.49	446.97	468.97	475.74	490.22

LogLik: log-likelihood (the higher the better); -2LogLik: deviance; AIC: Akaike information criterion; AICc: Akaike information criterion corrected; BIC: Bayesian information criterion. For -2logLik, AIC, AICc and BIC the lower the better

A series of goodness of fit indicators were produced to identify the model better fitting the data (Table 1). The choice of the model was based on the goodness of fit and the interpretability of the results, in addition to the biological relevance and statistical significance of the fixed factors. The overall considerations and model results are provided in more detail only for models 3 and 4 since their goodness of fit largely exceeds that of models 1 and 2 based on all the indicators above. Statistical heterogeneity was tested using a χ^2 test (Cochrane's Q test – (Veroniki et al., 2016)).

1.1.4.1. Model 3

Results of the estimates of the residual heterogeneity and overall significance of the fixed effects, the parameters for fixed effects, structure and estimates of the random effects are reported in Tables 2–5 respectively. They are based on the six studies selected after the screening (Section 3 of the scientific opinion), corresponding to 17 arms and 58 time-points of measurements.

Table 2: Test for residual heterogeneity and overall significance of fixed effects

Test for	Value	p
Residual heterogeneity	QE(df = 49) = 7.6	1.0*
Fixed effects (overall)	QM(df = 9) = 681.2	< .0001**

QE: Q test for residual heterogeneity; QM: Q test for moderators; df: degrees of freedom

*hypothesis that residual heterogeneity is equal to 0 cannot be rejected with prob<0.05

**hypothesis that overall variability explained by fixed effects (moderators) is equal to 0 can be rejected with prob<0.05

Table 3: Fixed effects estimate

Factor	Effect estimate	Standard error	p-value	Prediction interval lower bound*	Prediction interval upper bound*
VitD intake	1.7348	0.4688	***0.0002***	0.8160	2.6537
Baseline serum 25(OH)D	0.4838	0.6642	0.4664	-0.8181	1.7856
Duration: 1	25.4606	37.4128	0.4962	-47.8672	98.7884
Duration: 3	43.4741	37.7349	0.2493	-30.4849	117.4331
Duration: 6	57.2148	39.0943	0.1433	-19.4087	133.8382
Duration: 12	65.4620	38.0235	***0.0851***	-9.0627	139.9867
Analytical method: RIA	2.3300	13.5712	0.8637	-24.2691	28.9291
Duration of gestation: mixed	-5.4677	12.0982	0.6513	-29.1797	18.2443
Age	-0.5935	0.3097	***0.0553***	-1.2006	0.0136

*the Prediction Interval of the estimated effects expresses both the sampling uncertainty and the uncertainty due to variability across studies. It provides the interval (lower and upper bound) that would contain a future true estimated effect (if extracting a new sample of studies) with a certain probability (usually 95%), given what has already been observed.
For categories of duration and analytical method; see Table 5 of the scientific opinion. RIA: radioimmunoassay.

The fixed effects vitamin D intake, baseline concentration of the biomarker and age are expressed as continuous variables, whereas supplementation duration, analytical methods and duration of gestation are treated as categorical data. For the categorical fixed effects, one category is used as a reference and the parameters for the remaining classes indicate their additional effect with respect to the reference one (0 for the duration, LC-MS/MS for the analytical method, full-term for the category of duration of gestation).

Overall the model was able to **explain most of the heterogeneity in the data**, showing the residual component to be not statistically significant (Table 2). The fixed effects were **overall statistically significant** (Table 2), though most of the **individual main effects were not** (Table 3).

Table 4: Random effects - hierarchical structure in the data

Variance components	n. levels	
1st hierarchical level: study	6	
2nd hierarchical level: arm	17	
1st hierarchical level: arm		17
2nd hierarchical level: repeated measurement		58

Table 5: Random effect estimate – arms and repeated measurements

Parameters	Estimate	
$Var = \tau^2$	0.001	
$Corr = \rho$	0.50	
$Var = \gamma^2$		0.001
$Corr = \phi$		0.50

1.1.4.2. Model 4 – original scale

Results of the estimates of the residual heterogeneity and overall significance of the fixed effects, the parameters for fixed effects, structure and estimates of the random effects are reported in Tables 6–9 respectively. Fixed effects were taken forward from the previous model when they were statistically significant or marginally so ($p < 0.10$) (i.e. vitamin D intake, duration and age). Duration was still considered although only one category (12 months) was statistically significant. The baseline concentration of serum 25(OH)D was included based on biological considerations.

Table 6: Test for residual heterogeneity and overall significance of fixed effects

Test for	Value	p
Residual heterogeneity	QE(df = 51) = 7.9	1.0*
Fixed effects (overall)	QM(df = 7) = 680.8	< .0001**

QE: Q test for residual heterogeneity; QM: Q test for moderators; df: degrees of freedom
*hypothesis that residual heterogeneity is equal to 0 cannot be rejected with prob<0.05
**hypothesis that overall variability explained by fixed effects (moderators) is equal to 0 can be rejected with prob<0.05

Table 7: Fixed effects estimate

Factor	Effect estimate	Standard error	p-value	Prediction interval lower bound*	Prediction interval upper bound*
VitD intake	1.73	0.45	***0.0001***	0.85	2.61
Baseline serum 25(OH)D	0.39	0.41	0.3462	-0.42	1.20
Duration: 1	30.22	24.89	0.2247	-18.57	79.00
Duration: 3	48.21	19.42	***0.0131***	10.14	86.28
Duration: 6	62.85	20.01	***0.0017***	23.62	102.08
Duration: 12	71.11	19.78	***0.0003***	32.34	109.88
Age	-0.59	0.31	***0.0551***	-1.20	0.01

* the Prediction Interval of the estimated effects expresses both the sampling uncertainty and the uncertainty due to variability across studies. It provides the interval (lower and upper bound) that would contain a future true estimated effect (if extracting a new sample of studies) with a certain probability (usually 95%), given what has already been observed.
For categories of duration; see Table 5 of the scientific opinion.

As for model 3, the fixed effect vitamin D intake, baseline concentration of the biomarker and age are expressed as continuous variables, whereas supplementation duration, analytical methods and duration of gestation are treated as categorical data.

Table 8: Random effects – hierarchical structure in the data

Variance components	n. levels	
Outer factor: study	6	
Inner factor: arm	17	
Outer factor: arm		17
Inner factor: repeated measurement		58

Table 9: Random effect estimate – arms and repeated measurements

Parameters	Estimate	
$Var = \tau^2$	0.0010	
$Corr = \rho$	0.50	
$Var = \gamma^2$		0.0010
$Corr = \phi$		0.50

Also in this case, the model **explained most of the heterogeneity** in the data (Table 6). Overall the fixed effects were statistical significant (Table 6), with **all individual effects statistically significant** ($p < 0.05$) or marginally significant ($p < 0.10$) **except** for the baseline serum 25(OH)D concentration and one duration category (Table 7).

Based on these arguments, **model 4 was considered the most suitable** to predict mean study-arm value of serum 25(OH)D.

1.1.4.3. Model 4 – ln-scale

An additional model was fitted keeping the same fixed and random effects as in model 4 but using a ln-scale for the achieved serum concentration of 25(OH)D (response), its baseline value and the vitamin D intake. Fixed effects estimates for ln-transformed model are reported in Table 10.

Table 10: Fixed effect estimates under ln-scale model

Factor	Effect estimate	Standard error	p-value	Prediction interval lower bound*	Prediction interval upper bound*
Ln(vitD intake)	0.08	0.04	***0.03***	0.01	0.15
Ln(baseline serum 25(OH)D)	0.36	0.16	***0.02***	0.05	0.67
Duration: 1	2.84	0.64	***<.0001***	1.59	4.09
Duration: 3	3.03	0.58	***<.0001***	1.89	4.16
Duration: 6	3.04	0.58	***<.0001***	1.92	4.17
Duration: 12	3.06	0.54	***<.0001***	2.0	4.13
Age	-0.002	0.003	0.41	-0.01	0.004

* the Prediction Interval of the estimated effects expresses both the sampling uncertainty and the uncertainty due to variability across studies. It provides the interval (lower and upper bound) that would contain a future true estimated effect (if extracting a new sample of studies) with a certain probability (usually 95%), given what has already been observed.
For categories of duration; see Table 5 of the scientific opinion.

Also in this case the model **explained most of the heterogeneity** in the data. Overall the fixed effects were statistical significant, with **all individual effects statistically significant** ($p < 0.05$) **except** for the age that was anyhow kept in the model in order to being able to produce separated estimates by age categories (Table 10).

Compared to the model based on variables expressed in the original scale, the model with ln-transformed scale for response and some fixed effects is considered to better meet assumptions of normality and homoscedasticity. Adherence to linearity is better achieved in the model expressed in the original scale.

1.1.5. Model formal description

Based on the discussion above, model 4 with variables expressed in the original and ln-transformed scale was retained for further analysis. The formal structure of model 4 (original scale and ln-transformed scale) is described below:

$$Y_{ijk} = \beta_0 \mathrm{X}_{0ij} + \beta_1 \mathrm{X}_{1ij} + \beta_2 \mathrm{X}_{2ijk} + \beta_3 \mathrm{X}_{3ijk} + \mathrm{s}_{ij} + \mathrm{r}_{jk} \qquad \text{Model in the original scale}$$

$$ln(Y_{ijk}) = \beta_0 \ln(\mathrm{X}_{0ij}) + \beta_1 \ln(\mathrm{X}_{1ij}) + \beta_2 \mathrm{X}_{2ijk} + \beta_3 \mathrm{X}_{3ijk} + \mathrm{s}_{ij} + \mathrm{r}_{jk} \qquad \text{Model in the ln-scale}$$

Where $\mathrm{X}_{0ij}, \mathrm{X}_{1ij,} \mathrm{X}_{2ijk}, \mathrm{X}_{3ijk}$ are the **fixed effects**:

- ✓ Baseline value of the serum 25(OH)D prior to start vitamin D supplementation
- ✓ Vitamin D supplemental intake (in µg/day)
- ✓ Age of the infants (in weeks)
- ✓ Supplementation duration in categories (in weeks)

And $\mathrm{s}_{ij}, \mathrm{r}_{jk}$ are the **random factors** with variability denoting the amount of heterogeneity explained by the correlation structure among arms (*j*) within a study (*i*) and repeated measures (*k*) within each study-arm (*j*). The random factors were assumed to be normally distributed with a compound symmetry structure for the variance/covariance matrix with component τ^2 (across study variability) and ρ (between arms correlation) and γ^2 (within arm variability) and ϕ (between repeated observations correlation) respectively. The compound symmetry is a correlation structure that

assumes a constant correlation between each couple of arms from the same studies and each couple of repeated observations on the same arm. A Restricted Maximum Likelihood method was used to estimate heterogeneity components (Viechtbauer, 2005; Raudenbush, 2009)

1.1.6. Model diagnostics

Diagnostics were performed for each of the two models (original and ln-scale) in order to identify possible deviations from main assumptions, outliers (if any) and more influential observations (if any).

1.1.6.1. Diagnostic for deviation from linearity and outliers detection

From a biological viewpoint, it would have been more realistic to expect that the achieved serum 25(OH)D concentrations levels off at high doses of vitamin D. In addition there could be some uncertainty in the shape of the association for doses larger than 40 µg/day due to scarcity of data at higher intake. However **inspection of the standardised residuals versus the fitted study-arm mean serum 25(OH)D values** (Figure 5) has not highlighted major patterns that might raise concerns **for any of the two models** on the linearity of the relationship since overall study arms (dots) are evenly spread around the zero line.

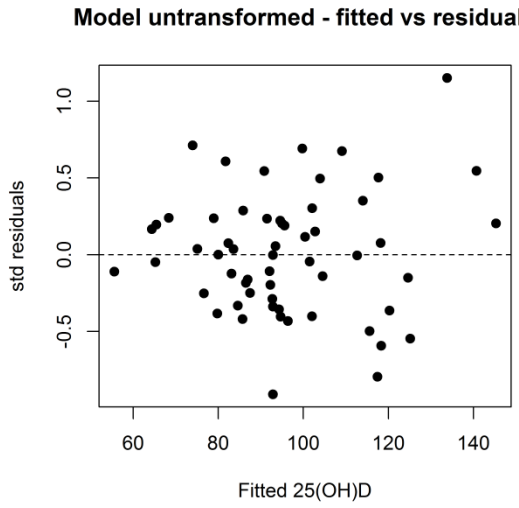

Figure 5a

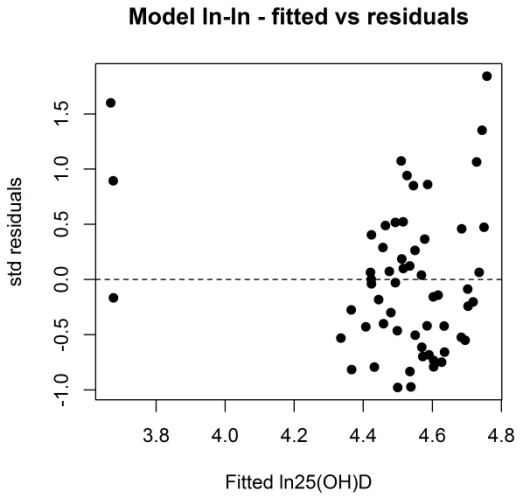

Figure 5b

Figure 5: Standardised residuals – original scale and ln-transformed scale model

One dot represents a repeated study arm or a repeated observation on a study arm. Residuals obtained from adjusted models (i.e. including moderators).

It was considered that a non-linear model reaching a plateau at high doses would be expected to lead to lower predictions of the study-arm mean values in the upper tail of the distribution (lower responses at higher doses) as compared to the ones estimated with the linear model. Therefore, the Panel concluded that the estimates obtained with a linear model are conservative.

1.1.6.2. Outliers and influential case diagnostic

Conventionally, any observation with standardised residual greater than 3 (positive or negative) is considered as an outlier. The inspection of the standardised residuals plot (Figure 5) has not highlighted evident outliers since no standardised residuals exceed 2 (positive or negative). A variety of **influential case diagnostics** can be computed when conducting a meta-analysis (Viechtbauer and Cheung, 2010). Figure 6 shows a plot of the Hat values the Cook's distances computed **on the**

58 mean observations that form the body of evidence. The Hat value is an indicator of the **distance between predicted and observed value**. Cook's distance can be interpreted as the distance between the entire set of predicted values once with the i^{th} study included and once with the i^{th} study excluded from the model fitting.

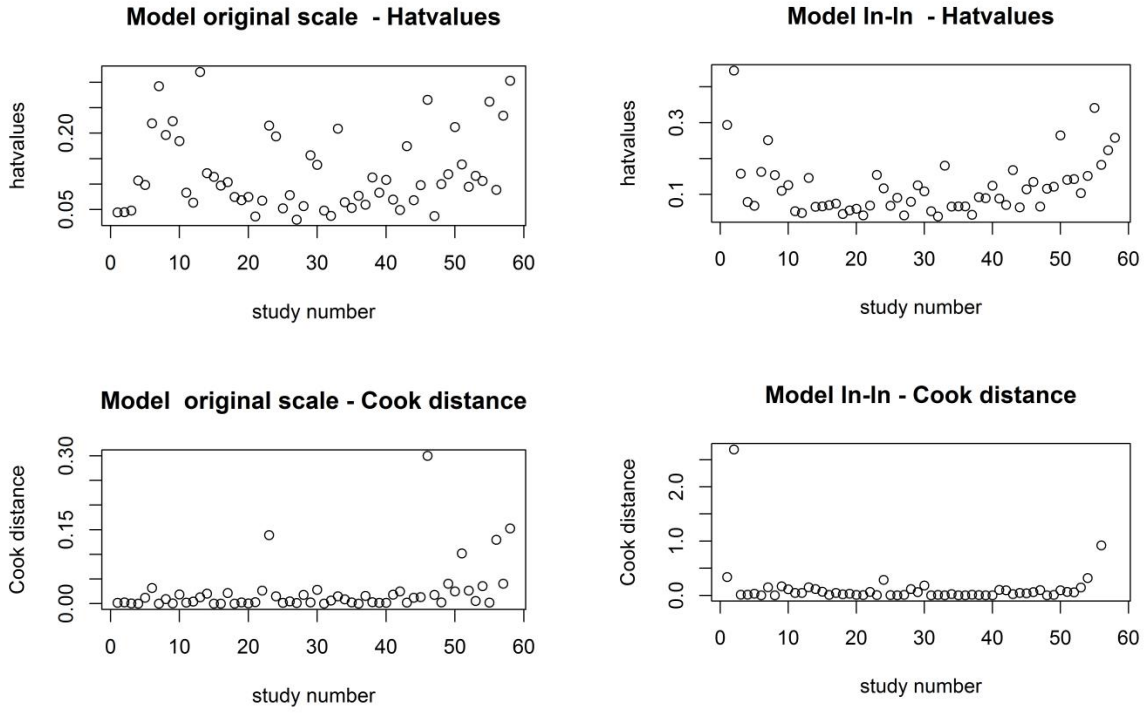

Figure 6a

Figure 6b

Figure 6: Original scale (left) and ln-transformed scale (right) model - Hat values and Cook's distance for the 58 arms-measurement occasions

X-axis: the progressive number identifies arms-measurement occasions; Y-axis: Hat values (above) and the Cook's distance (below) calculated for each arms-measurement occasions.

The analysis of the Hat values shows there is **no study-arm that largely overcomes the others** for both models (original and ln-transformed scales). All Hat values fall within a range of 3 times the overall Hat values mean (0.12 for both models) except one for the ln-transformed scale (observation number 2). As for the Cook's distance, when using the original scale model, **5 arms-repeated measurement occasions** appear to be **more influential,** since their value exceeded 3 times the overall average (0.024), whereas 3 study-arms-measurement occasions are identified as more influential in the ln-transformed model (values above 3 times the mean of 0.18). Tables 11 and 12 list the highly influential observations for the original and the ln-transformed scale models respectively.

Table 11: Original scale model: study arm and measurement occasions with Cook's distance exceeding three times the mean

Study ID	Paper	arm	time of observation (weeks)	Cook
3897	(Ziegler et al., 2014)	2	48	0.1392671
2921	(Holst-Gemeiner et al., 1978)	1	2	0.2999646
3687	(Gallo et al., 2013)	3	48	0.1018319
3687	(Gallo et al., 2013)	4	8	0.1294815
3779	(Gordon et al., 2008)	2	6	0.1527853

ID: automatic identification number.

Table 12: Ln-transformed scale model: study arm and measurement occasions with Cook's distance exceeding three times the mean

Study ID	Paper	arm	time of observation (weeks)	Cook
3792	(Grant et al., 2014)	1	17	0.34
3792	(Grant et al., 2014)	1	26	2.69
3687	(Gallo et al., 2013)	4	8	0.92

ID: automatic identification number.

After further investigation of the most influencial data, it was concluded that there are no major concerns, since no unusual patterns or possible anomalities were identified for these observations. Both models in the original and ln-transformed scales are used in the following analysis to give a sense of the uncertainty associated to the several choices made at the methodological level and their influence on the results.

1.2. Predicted study mean achieved serum 25(OH)D concentration

The distribution of the **mean** achieved serum concentration of 25(OH)D was **simulated using the two predictive models** (original and ln-transformed scale) described above. The hypothesis here is that study-arm means in the body of evidence represent a **random sample** from a theoretical population of study-arm mean values whose distribution can be described conditionally to the value of the explanatory variables 'vitamin D intake', 'baseline concentration', 'duration' and 'age'. Since few observations are available in the body of evidence, an empirical distribution was generated with a large number of simulations (random draws) to approximate better the true distribution.

To make the model predictions **more realistic**, **a probability distribution was used for the baseline value of serum 25(OH)D to reflect the variability that is expected for this factor in a theoretical population of studies**. The distribution was elicited based on expert's knowledge. A truncated normal distribution with a mean of 50 nmol/L, standard deviation of 20 nmol/L and range between 10 and 100 nmol/L was considered realistic. A truncated normal was preferred to avoid values biologically unrealistic (i.e. baseline study mean concentrations below 10 and above 100 nmol/L). Figure 7 shows the simulated empirical distribution of the study mean baseline serum 25(OH)D concentration, obtained generating 500 random draws, in the original (left side) and ln-transformed scale (right side) respectively.

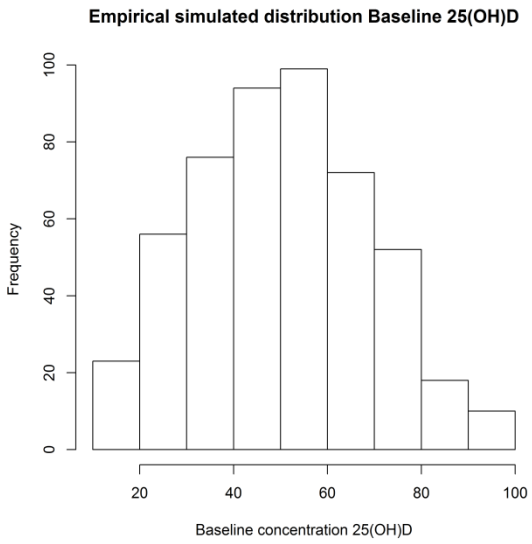

Figure 7a

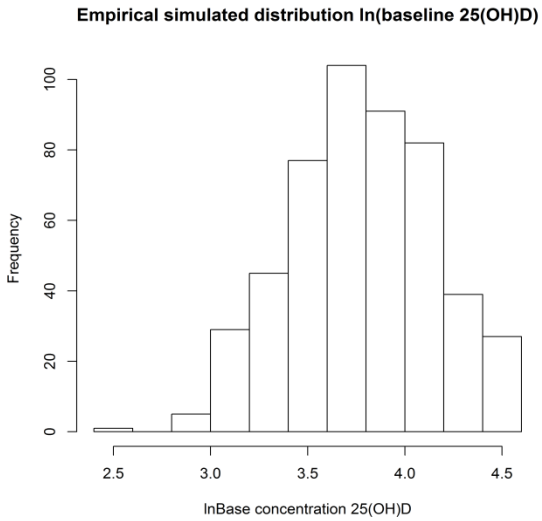

Figure 7b

Figure 7: Empirical distribution of Baseline serum 25(OH)D (nmol/L). Absolute frequency out of 500 random drawings – original scale (left) and ln-transformed scale (right)

As for **vitamin D intake** and **age** of the infants, a range of 5–50 µg/day and 1–52 weeks was considered appropriate. For the latter factor, the assumption was that the mean age of the infants in a random sample of studies has approximately a **uniform distribution** over the range 1 to 52 weeks. For simplicity, the range 1 to 52 was used. For the vitamin D intake, the range **observed in the body of evidence** was considered (5 to 50 µg/day). Table 13 summarises the distributions/values considered realistic for the fixed factors.

Table 13: Distributions/values of the fixed factors in the predictive model (2 lines correspond to model in the original and ln-transformed scale)

Fixed factor	Distribution or range of values	Scale
Baseline serum 25(OH)D (in nmol/L)	$\sim \text{TruncNorm}(\min = 10, \max = 100, \quad m = 50, sd = 20)$ $\sim \text{TruncNorm}(\min = 2.302, \max = 4.605, \quad m = 3.837, sd = 0.385)$	Original Natural log
Vitamin D intake (in µg/day)	Range 5-50 all integer values in the range Range 1.609–3.912 all integer values in the range	Original Natural log
Age (in weeks)	Range 1–52 all integer values in the range Range 1–52 all integer values in the range	Original Natural log

Max: maximum; min: minimum; natural log: natural logarithm i.e. ln; sd: standard deviation, TruncNorm: truncated normal

The supplementation duration was **always set at 6 months,** since this was the duration with the second highest study-arm mean achieved serum 25(OH)D concentration, and with the highest achieved concentration for the highest vitamin D intake levels (Figure 4.e, Section 1.1.3. of this Annex).

Monte Carlo simulations (Burmaster and Anderson, 1994; Robert and Casella, 2004) were used to **generate the empirical distribution** of the baseline serum 25(OH)D concentration and the **predicted study-arm mean** of the achieved serum 25(OH)D concentration. A total of 1,196,000 predictions were generated as the result of all possible combinations of the fixed effects. The empirical distributions and the related quartiles of the predicted mean achieved serum 25(OH)D concentration obtained using the model in the original and ln-transformed scales are provided in Figure 8 and Table 14.

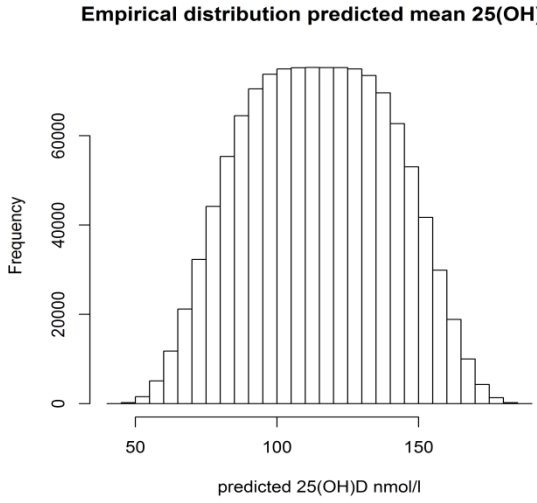

Figure 8a

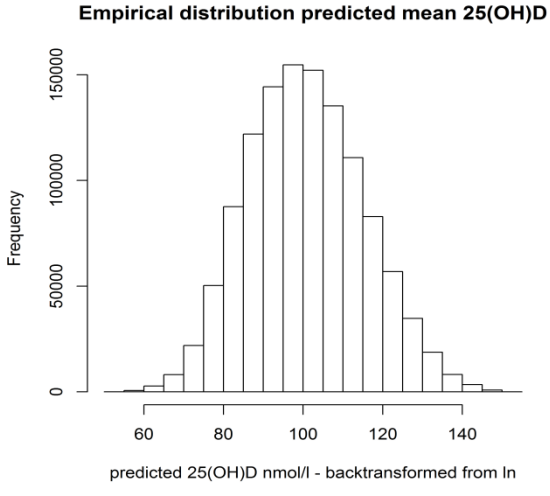

Figure 8b

Figure 8: Model in the original (left side) and ln-transformed (right side) scale - Empirical distribution of the predicted study-arm mean serum 25(OH)D concentration (nmol/L)

Note: y-axis: Absolute frequency out of 1,196,000 simulations.

Table 14: Model in the original and ln-transformed scales: quartiles of the empirical distribution of the predicted study-arm mean achieved serum 25(OH)D concentration

Distribution value	Predicted study-arm mean achieved serum 25(OH)D	
	Original scale model	Ln-transformed model*
Minimum	44.87	54.08
1st quartile	94.48	90.22
Median	114.50	100.20
Mean	114.50	100.80
3rd quartile	134.50	110.70
Maximum	187.50	150.10

* Values are back-transformed to the original scale

1.3. Distribution of simulated individual responses

The Tolerable Upper Intake Level (UL) is defined as the 'maximum level of total chronic intake of a nutrient from all sources judged to be unlikely to pose a risk of adverse health effects in humans' (EFSA NDA Panel, 2010).

In order to identify such an intake of vitamin D for infants, **individual responses were simulated** from each study-arm mean response, similarly to what was done for establishing individual population coverage of the adequate intake recently established for vitamin D (EFSA NDA Panel, 2016). For each of the two models (original and ln-transformed scale) the following steps were followed:

- A **truncated normal distribution** was assumed to describe the variability of the individually achieved serum 25(OH)D concentration around the study population mean.
- In the absence of reliable information on the possible variability at the individual level, an **average coefficient of variation (CV) was derived** from the study-arm-measurement occasions available in the body of evidence, averaging across all the within-study sampling variability with weights given by sample size. Implicit assumption was that within-study variability provides an unbiased estimate of the inter-individual variability in the population from which individuals have been selected.
- **The mean CV has been used to compute the standard deviation** of each individual distribution by multiplying by the study-arm predicted mean value.
- The **distribution of individually achieved serum 25(OH)D concentration** was obtained for each of the two models, in the original and ln-transformed scale, with 100 random draws simulating the hypothetical population of individual around the study-arm mean.

1.4. Percentage of infants exceeding defined serum 25(OH)D concentrations

For each class of vitamin D intake, categories of the biomarker at baseline (below 30 nmol/L; 30–60 nmol/L; 60–100 nmol/L) and age classes (below and above 6 months of age), the percentages of infants expected to exceed a specific concentration of the biomarker were computed on the basis of the original and ln-transformed scale models. Due to the uncertainty surrounding the concentration of serum 25(OH)D associated with an increased risk of adverse outcomes (Section 3.3.6. of the scientific opinion), two concentrations (150 and 200 nmol/L) were considered in order to investigate how this would change the results of the meta-regression analysis.

1.4.1. Results

Results are reported in Table 15 and 16 for infants up to the age of 6 months and between 6 months and 12 months respectively for the model in the original scale, for concentration equal to 150 and 200 nmol/L. Results are given in Tables 17–18 for the model in the ln-transformed scale.

Table 15: Model in the original scale - Percentage of infants exceeding the serum 25(OH)D concentrations of 150 and 200 nmol/L – infants up to 6 months of age (26 weeks included)

VitD intake	% infants with achieved serum 25(OH)D concentration exceeding 150 nmol/L				% infants with achieved serum 25(OH)D concentration exceeding 200 nmol/L			
	Baseline serum 25(OH)D				Baseline serum 25(OH)D			
	[10, 30)	[30,60)	[60,100]	Any	[10, 30)	[30,60)	[60,100]	Any
[5-10)	1	2	6	3	0	0	0	0
[10-15)	2	5	11	7	0	0	1	0
[15-20)	5	10	18	12	0	1	2	1
[20-25)	10	17	25	19	0	2	4	2
[25-30)	16	24	33	25	1	3	7	4
[30-35)	23	32	40	34	3	6	11	7
[35-40)	32	39	48	41	6	10	16	11
[40-45)	38	47	54	48	10	15	21	17
[45-50)	46	53	60	54	15	20	27	22

Table 16: Model in the original scale - Percentage of infants exceeding the serum 25(OH)D concentrations of 150 and 200 nmol/L – infants between 6 and 12 months of age (26 weeks excluded)

VitD intake	% infants with achieved serum 25(OH)D concentration exceeding 150 nmol/L				% infants with achieved serum 25(OH)D concentration exceeding 200 nmol/L			
	Baseline serum 25(OH)D				Baseline serum 25(OH)D			
	[10, 30)	[30,60)	[60,100]	Any	[10, 30)	[30,60)	[60,100]	Any
[5-10)	0	0	1	0	0	0	0	0
[10-15)	0	1	3	1	0	0	0	0
[15-20)	1	2	7	4	0	0	0	0
[20-25)	2	6	12	8	0	0	1	0
[25-30)	6	11	19	13	0	1	2	1
[30-35)	11	18	27	20	1	2	4	2
[35-40)	18	25	35	27	2	4	8	5
[40-45)	25	34	42	36	4	7	12	8
[45-50)	33	41	49	42	7	11	17	12

Table 17: Model in the ln-transformed scale - Percentage of infants exceeding the serum 25(OH)D concentrations of 150 and 200 nmol/L – infants up to 6 months of age (26 weeks included)

VitD intake	% infants with achieved serum 25(OH)D concentration exceeding 150 nmol/L				% infants with achieved serum 25(OH)D concentration exceeding 200 nmol/L			
	Baseline serum 25(OH)D				Baseline serum 25(OH)D			
	[10, 30)	[30,60)	[60,100]	Any	[10, 30)	[30,60)	[60,100]	Any
[5-10)	0.3	4.2	15. 6	7.1	0	0.1	1.4	0.5
[10-15)	0.6	6.0	19.5	9.4	0	0.2	2.2	0.8
[15-20)	0.9	7.4	22.4	11.0	0	0.3	2.9	1.0
[20-25)	1.2	8.6	24.5	12.4	0	0.4	3.5	1.3
[25-30)	1.5	9.6	26.1	13.5	0	0.5	4.1	1.5
[30-35)	1.7	10.5	27.6	14.6	0	0.6	4.6	1.7
[35-40)	1.9	11.3	28.9	15.4	0	0.7	5.2	1.9
[40-45)	2.2	12.1	29.9	16.2	0	0.8	5.6	2.1
[45-50)	2.4	12.7	31.0	16.9	0	0.9	6.0	2.3

Table 18: Model in ln-transformed scale - Percentage of infants exceeding the serum 25(OH)D concentrations of 150 and 200 nmol/L – infants between 6 and 12 months of age (26 weeks excluded)

VitD intake	% infants with achieved serum 25(OH)D concentration exceeding 150 nmol/L				% infants with achieved serum 25(OH)D concentration exceeding 200 nmol/L			
	Baseline serum 25(OH)D				Baseline serum 25(OH)D			
	[10, 30)	[30,60)	[60,100]	Any	[10, 30)	[30,60)	[60,100]	Any
[5-10)	0.1	2.1	10.2	4.2	0	0	0. 6	0.2
[10-15)	0.2	3.2	13.3	5.9	0	0.1	1	0.3
[15-20)	0.4	4.2	15.7	7.1	0	0.1	1.4	0.5
[20-25)	0.5	5.0	17.5	8.2	0	0.1	1.7	0.6
[25-30)	0.6	5.7	19.2	9.1	0	0.2	2.1	0.7
[30-35)	0.7	6.9	20.3	9.8	0	0.2	2.4	0.8
[35-40)	0.8	7.0	21.5	10.5	0	0.3	2.7	1.0
[40-45)	0.9	7.5	22.5	11.2	0	0.3	3.0	1.1
[45-50)	1.1	8.0	23.5	11.8	0	0.3	3.3	1.2

NB: ranges expressed as [a-b) mean including a but excluding b.

1.4.2. Results interpretation

These results have to be read with caution. They represent predictions obtained from modelling, simulations and related assumptions (previously specified). The exceedance percentages should **not be interpreted as precise estimates, rather as informed quantitative judgements** about the expected prevalence of infants that might exceed the serum 25(OH)D concentration at the various vitamin D intakes, given baseline values of the biomarker and age groups.

For infants **younger than 6 months,** based on the results of the prediction model in the original scale, at a vitamin D intake of up to **25 µg/day**, which is the UL previously set by EFSA NDA Panel (2012), depending on the baseline serum 25(OH)D concentration, 10 to 25% of individuals **younger than 6 months** would achieve serum 25(OH)D concentrations above 150 nmol/L, and **0 to 4%** of infants would achieve serum 25(OH)D concentrations above 200 nmol/L (Table 15). The results of the ln-transformed scale up to a dose of 25 µg/day are consistent with the results in the original scale (Table 17).

For infants **between 6 and 12 months of age,** based on the results of the prediction model in the original scale (Table 16), the predicted percentage of individuals exceeding serum 25(OH)D concentrations of 150 nmol/L or 200 nmol/L would be:

- 2 to 12% or **0 to 1%** at a supplemental vitamin D intake of up to 25 µg/day,
- 6 to 19% or 0 to 2% at intakes of up to 30 µg/day,
- 11 to 27% or 1 to 4% at intakes at of up to **35** µg/day

The predicted percenteges based on the results of the prediction model in the ln-transformed scale for doses up to **35** µg/day range from 1% to 20% for the concentration of 150 nmol/L, and from **0 to around 2%** for the concentration of 200 nmol/L (Table 18).

1.5. Unaddressed sources of uncertainty

This section intends to provide a list of the uncertainties that have not been addressed in the statistical analysis, neither quantitatively nor qualitatively.

1.5.1. Uncertainties related to modelling

Some limitations have to be acknowledged in the set-up of the models.

The intake-response relationship is estimated using aggregated data (study-arm mean value). The relationship observed averaging across trials might not be the same as the one observed within a trial.

This issue is known as **ecological fallacy or aggregation bias** (difficult to investigate if no individual data are available).

Some potential confoundings or moderators have not been measured in the studies and **could not be included in the model.**

Sampling uncertainty was not accounted for in the predicted study-arm mean values. The mean prediction has been considered instead of the upper bound of the credible interval of the prediction.

A compound-symmetry structure was used to describe the correlation structure in the data. Other structures could have been considered.

Non-linearity could have better met the expected dose-response relationship from a biological viewpoint. Although no significant deviations from linearity were identified, the lack of a large body of evidence covering high doses makes the true shape of the relationship at doses higher than 40 µg/day somehow uncertain. Of note, a non-linear model showing mean serum 25(OH)D concentration reaching a plateau at high doses would have probably led to study-arm mean values lower in the upper tail of the distribution as compared to the ones estimated with the linear model. Consequently, a non-linear model would have probably led to a lower percentages of individuals exceeding a certain serum 25(OH)D concentration. Therefore, the current estimates obtained on the basis of the linearity assumption are considered conservative.

The **inter-individual variability** necessary to estimate distribution of individually achieved serum 25(OH)D concentration was unknown. It was estimated on the basis of the mean coefficient of variation (CV) of the within study variability. The same CV has been applied to all study means to derive an inter-individual variance. Therefore, implicitly, studies with larger mean were assumed to have a larger dispersion of the individual values around it and vice-versa. This assumption, based on observation of the real world and therefore realistic, has contributed to amplify the difference in the predicted percentage of infants exceeding a serum 25(OH)D concentration between original and ln-transformed scale models.

1.5.2. Additional sources of uncertainty

Additional sources of uncertainties have not been addressed.

The dose-response has been computed considering **only vitamin D intake from supplementation**. Background intake from food has not been considered. Assumption was that **bio-availability** is the same for vitamin D naturally contained in food or added in fortified food and provided as supplements to the infants (Section 5.6.3. of the scientific opinion). The same assumption (in terms of bio-availability or compliance/risk of overdosage) applies to the **form of supplementation** whether provided in drops, pills or other forms.

Some **analytical methods** may overestimate the 'true' value of the serum 25(OH)D concentration, especially in infants. Impact of the concentration of C3-epimer of 25(OH)D, particularly in the youngest infants, has also not be considered for this analysis (Section 1.8.6. of the scientific opinion).

Compliance to the planned administration is one source of uncertainty that could have equally led to an overestimation or underestimation of the vitamin D doses administered to the infants. Parents might equally forget to give supplements to infants or inadvertently provide a higher dose to them.

1.6. Software

Data editing and cleaning was performed using SAS version 9.3. Statistical analyses were carried out with R version 3.3.2 (R Core Team, 2013) and Rstudio version 1.0.136. The meta-regression was performed using the 'metafor' package (Viechtbauer, 2010b).

References

Burmaster DE and Anderson PD, 1994. Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. Risk Analysis, 14, 477-481.

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2010. Scientific Opinion on principles for deriving and applying Dietary Reference Values. EFSA Journal 2010; 8(3):1458, 30 pp. doi:10.2903/j.efsa.2010.1458

EFSA NDA Panel (EFSA Panel on Dietetic Products Nutrition and Allergies), 2012. Scientific Opinion on the Tolerable Upper Intake Level of vitamin D. EFSA Journal 2012;10(7):2813, 45 pp. doi:10.2903/j.efsa.2012.2813 doi:10.2903/j.efsa.2012.2813

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2016. Scientific Opinion on Dietary Reference Values for vitamin D. EFSA Journal 2016;14(10):4547, 179 pp. doi:10.2903/j.efsa.2016.4547

EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen KH, More S, Mortensen A, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Silano V, Solecki R, Turck D, Aerts M, Bodin L, Davis A, Edler L, Gundert-Remy U, Sand S, Slob W, Bottex B, Abrahantes JC, Marques DC, Kass G and Schlatter JR, 2017. Update: Guidance on the use of the benchmark dose approach in risk assessment. EFSA Journal 2017;15(1):4658, 41 pp. doi:10.2903/j.efsa.2017.4658

Gallo S, Comeau K, Vanstone C, Agellon S, Sharma A, Jones G, L'Abbe M, Khamessan A, Rodd C and Weiler H, 2013. Effect of different dosages of oral vitamin D supplementation on vitamin D status in healthy, breastfed infants: a randomized trial. JAMA, 309, 1785-1792.

Gordon CM, Williams AL, Feldman HA, May J, Sinclair L, Vasquez A and Cox JE, 2008. Treatment of hypovitaminosis D in infants and toddlers. Journal of Clinical Endocrinology and Metabolism, 93, 2716-2721.

Grant CC, Stewart AW, Scragg R, Milne T, Rowden J, Ekeroma A, Wall C, Mitchell EA, Crengle S, Trenholme A, Crane J and Camargo CA, Jr., 2014. Vitamin D during pregnancy and infancy and infant serum 25-hydroxyvitamin D concentration. Pediatrics, 133, e143-153.

Higgins JP, White IR and Anzures-Cabrera J, 2008. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. Statistics in Medicine, 27, 6072-6092.

Holst-Gemeiner D, Gemeiner M, Pilz I and Swoboda W, 1978. [Plasma 25-hydroxycholecalciferol after daily vitamin D administration in comparison with massive single-dose prophylaxis (author's transl)]. Wiener Klinische Wochenschrift, 90, 509-512.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available online: http://www.R-project.org/

Raudenbush SW, 2009. Analyzing effect sizes: Random effects models. In: The handbook of research synthesis and meta-analysis. Ed H. Cooper LVH, & J. C. Valentine Russell Sage Foundation, New York, 295–315.

Robert C and Casella R, 2004. Monte Carlo Statistical Methods. Springer, New York.

Shapiro SS and Wilk MB, 1965. An analysis of variance test for normality (complete samples). Biometrika, 52, 591–611.

van Houwelingen HC, Arends LR and Stijnen T, 2002. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine, 21, 589-624.

Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JP, Langan D and Salanti G, 2016. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods, 7, 55-79.

Viechtbauer W, 2005. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. Journal of Educational and Behavioral Statistics, 30, 261-293.

Viechtbauer W, 2010a. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software, 36.

Viechtbauer W and Cheung MW, 2010. Outlier and influence diagnostics for meta-analysis. Res Synth Methods, 1, 112-125.

Viechtbauer W, 2010b. "metafor : Meta-Analysis Package for R." R package version 1.4-0. Available online: http://CRAN.R-project.org/package=metafor

Ziegler EE, Nelson SE and Jeter JM, 2014. Vitamin D supplementation of breastfed infants: a randomized dose-response trial. Pediatric Research, 76, 177-183.

Abbreviations

AIC	Akaike information criterion
AICc	Akaike information criterion corrected
BIC	Bayesian information criterion
CV	coefficient variation
df	degrees of freedom
DRV	dietary reference value
ln-25(OH)D	natural logarithmic transformed concentration of serum 25(OH)D
ln-scale	natural logarithmic transformed scale
LogLik	log-likelihood
-2LogLik	deviance
LC-MS/MS	liquid chromatography-tandem mass spectrometry
NDA	Panel on Dietetic Products, Nutrition and Allergies
p-value	statistical significance level
prob	probability
Q	quantile
QE	Q test for residual heterogeneity
QM	Q test for moderators
RIA	radioimmunoassay
UL	tolerable upper intake level
serum 25(OH)D	25-hydroxy-vitamin D in serum
Vitamin D_2	ergocalciferol
Vitamin D_3	cholecalciferol