Transformations of summary statistics as input in meta-analysis for linear dose-response models on a logarithmic scale: a methodology developed within EURRECA


Olga W Souverein (Olga.Souverein@wur.nl)
Carla Dullemeijer (Carla.Dullemeijer@wur.nl)
Pieter van ’t Veer (Pieter.vantveer@wur.nl)
Hilko van der Voet (Hilko.vandervoet@wur.nl)

ISSN 1471-2288

Article type Research article
Submission date 3 January 2012
Acceptance date 25 April 2012
Publication date 25 April 2012

Article URL http://www.biomedcentral.com/1471-2288/12/57

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to http://www.biomedcentral.com/info/authors/

© 2012 Souverein  et al.; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Transformations of summary statistics as input in meta-analysis for linear dose–response models on a logarithmic scale: a methodology developed within EURRECA

Olga W Souverein1*
*Corresponding author
Email: Olga.Souverein@wur.nl

Carla Dullemeijer1
Email: Carla.Dullemeijer@wur.nl

Pieter van’t Veer1
Email: Pieter.vantveer@wur.nl

Hilko van der Voet2
Email: Hilko.vandervoet@wur.nl

1 Division of Human Nutrition, Wageningen University and Research Centre,
P.O. Box 8129, 6700 EV Wageningen, the Netherlands

2 Biometris, Wageningen University and Research centre, P.O. Box 100, 6700 AC Wageningen, the Netherlands

Abstract

Background
To derive micronutrient recommendations in a scientifically sound way, it is important to obtain and analyse all published information on the association between micronutrient intake and biochemical proxies for micronutrient status using a systematic approach. Therefore, it is important to incorporate information from randomized controlled trials as well as observational studies as both of these provide information on the association. However, original research papers present their data in various ways.

Methods
This paper presents a methodology to obtain an estimate of the dose–response curve, assuming a bivariate normal linear model on the logarithmic scale, incorporating a range of transformations of the original reported data.

Results
The simulation study, conducted to validate the methodology, shows that there is no bias in the transformations. Furthermore, it is shown that when the original studies report the mean
and standard deviation or the geometric mean and confidence interval the results are less variable compared to when the median with IQR or range is reported in the original study.

Conclusions

The presented methodology with transformations for various reported data provides a valid way to estimate the dose–response curve for micronutrient intake and status using both randomized controlled trials and observational studies.

Keywords

Methodology, Dose–response, Meta-analysis, EURRECA

Background

Meta-analysis of the association between micronutrient intake and biochemical proxies for micronutrient status or function is needed when setting micronutrient recommendations. Information on this association may come from randomized controlled trials as well as from observational studies. In a randomized trial subjects are randomized to receive either the intervention treatment or the control treatment, and a meta-analysis of such studies will usually provide a mean difference in micronutrient status between placebo and intervention groups, answering the question whether the biochemical status marker responds to the dietary intake of a micronutrient [1-3]. However, this analysis does not provide an estimate of the slope of the dose–response relationship. On the other hand, a meta-analysis of observational studies provides an estimate of the slope of the dose–response relation, but observational studies are hampered by for instance measurement error in the intake estimates, which causes bias in the reported association [4-6].

Ideally, information from observational studies and randomized controlled trials should be compared or even combined in a single meta-analysis to ensure that all reported information is taken into account over a broad range of intake. This requires that the summary statistics reported in individual studies are transformed into estimates of the dose–response relation. Since both intake and status are continuous variables, this estimate is actually an estimate of the regression coefficient of the linear regression of micronutrient status on micronutrient intake. The individual estimates of the dose–response regression coefficient may then be combined in a meta-analysis.

The statistical combination of study results may be complicated by the variety of ways that individual studies report the summary statistics. The results from randomized controlled trials as well as the baseline summary statistics of micronutrient intake and status may be reported as means, medians or geometric means. Variability is often reported as standard deviations, standard errors, interquartile ranges (IQR), ranges or confidence intervals (CI). In observational studies the relation between intake and status can be reported as a Pearson correlation coefficient, a Spearman rank correlation coefficient or a regression coefficient. In addition, either the intake variable or the status variable or both could have been logarithmically transformed before the correlation or association was calculated. All these different ways of reporting need to be standardized before meta-analysis is even possible.
This paper gives an overview of transformation methods to algebraically derive an estimate from each study of the regression coefficient (slope, b) and its standard error (se(b)), for studies that do not directly report these. The methods are validated by comparing the calculated values with theoretical values in a small-scale simulation study.

Methods

In order to derive transformations we assume a bivariate normal distribution on the log-scale for intake and status of an individual person. The log-scale was chosen because both intake and status values are always above zero, and the observed distributions of the micronutrient variables are often right-skewed. Moreover, as the true shape of the dose–response curve is usually unknown the linear relation between logarithmically transformed quantities provides the simplest approximation.

More in detail, for the dose–response meta-analysis of observational studies we assume that $\xi_0$ (intake of micronutrient) and $\eta_0$ (status or continuous health outcome) are log-normally distributed. The assumption of bivariate normality entails a linear association between $\xi = \ln(\xi_0)$ and $\eta = \ln(\eta_0)$, where $\ln$ denotes the natural logarithm. Note that we use the Greek letters $\xi$ and $\eta$ for the theoretical values of intake and status/response, and the Latin letters X and Y for the observed values of these variables. Furthermore, we reserve letters without subscript (e.g. X and Y) for values expressed on the ln-scale, and use letters with subscript 0 (e.g., $X_0$ and $Y_0$) for values expressed on the absolute (i.e., original) scale.

The process of data transformations to obtain the required statistics from what is reported in observational studies, consists of four steps (Figure 1). The first step is to obtain the mean of X ($m_X$) and Y ($m_Y$) and the standard deviation of X ($s_X$) and Y ($s_Y$). Secondly, the mean of $X_0$ ($m_{X_0}$) and $Y_0$ ($m_{Y_0}$) and the standard deviation of $X_0$ ($s_{X_0}$) and $Y_0$ ($s_{Y_0}$) are calculated when needed for the calculations in step 3. In this third step the correlation coefficient of the association between X and Y ($r_{XY}$) is calculated from the reported data. In the last step, the regression coefficient of the linear regression from Y on X ($b_{YX}$) is calculated from $r_{XY}$, and the $se(b_{YX})$ is calculated from $r_{XY}$, $s_Y$, $s_X$ and the sample size ($n$). For reports on randomized controlled trials, the process consists of three steps. In the first step, $m_Y$ and $s_Y$ are obtained for both intervention and placebo group. In the second step, $m_X$ is obtained, and in the last step, $b_{YX}$ and $se(b_{YX})$ are calculated. The equations for all these transformations are given below.

**Figure 1 Flowchart indicating the process of data transformations.** m, mean, s, standard deviation, b, regression coefficient, r, correlation coefficient. X indicates intake of micronutrient and Y indicates a proxy for micronutrient status or continuous health outcome. Capital letters without subscript are used for values expressed on the ln-scale and small letters with subscript 0 for values expressed on the absolute scale

**Univariate transformations**

First, we describe how the univariate statistics of the normal distributions at the ln-scale can be obtained from various reported statistics. We present formulas for $m_X$ and $s_X$, which of course can also be used similarly for $m_Y$ and $s_Y$ in observational studies. For randomized controlled trials the situation is different, because the variation in X is artificial and is not described by a normal distribution. Therefore, the transformations should be used only to
obtain mY and sY in the intervention and placebo groups separately. In most trials the within-

 group variation in X will be ignorable compared with the difference between the groups,
 consequently mX is calculated simply as mX_{con} = \ln(mX_{0,con}) for the placebo group and as
 mX_{int} = \ln(mX_{0,int}) the intervention group.

For these transformations, we assume that ξ is normally distributed with parameters \( \mu_\xi \) and
 \( \sigma_\xi \). For a lognormal distribution the mean on the absolute scale, \( \mu_\xi \), is given by
 \[ \mu_\xi = \exp(\mu_\xi + 0.5\sigma_\xi^2) \]
 and the standard deviation on the absolute scale, \( \sigma_\xi \), is given by
 \[ \sigma_\xi = \sqrt{\exp(\sigma_\xi^2 - 1)} \]. It follows that when the mean (mX_0) and the standard
 deviation (sX_0) are reported, mX can be calculated as:

\[
\text{mX} = \ln(\text{mX}_0) - 0.5sX^2
\]

where

\[
sX = \sqrt{\ln(1 + \left(\frac{sX_0}{mX_0}\right)^2)}
\]

The exponential function of the mean of the lognormal distribution is equal to the median on
 the absolute scale. Therefore, when the median (medX_0) has been reported on the absolute
 scale, mX is calculated as:

\[
\text{mX} = \ln(\text{medX}_0)
\]

As a measure of variability an IQRx or range (rangex) is often reported together with the
 median or mean. The IQR is the difference between the third quartile Q_3 and first quartile Q_1
 (the 75th percentile and the 25th percentile). Basically, there are two cases. If the lower and
 upper limits are reported as such, the difference between the ln-transformed limits may be
 equated to an appropriate multiple of the standard deviation sX. On the other hand, if only the
 IQR or range is reported as such, the derivation is more complex. When IQRX_0 is reported
 together with the median, the relation between these and sX is given by

\[
\text{IQRX}_0 = \text{medX}_0 \times [\exp(z \cdot sX) - \exp(-z \cdot sX)]
\]

where z represents the appropriate percentage point in the standard normal distribution (i.e., \( z_{0.75} = 0.6745 \)).

In this case sX may be calculated as

\[
sX = \sqrt{\ln\left(\frac{1}{2} \times \left(\frac{\text{IQRX}_0}{\text{medX}_0} + \sqrt{\left(\frac{\text{IQRX}_0}{\text{medX}_0}\right)^2 + 4}\right)\right)}
\]

(4)
When the IQR is reported together with the mean no explicit formula exists to derive $s_X$. Therefore, to obtain an estimate of $s_X$ from these quantities a nonlinear function optimization is employed to find the value of $s_X$ for which the following equation holds

$$\text{IQR}_X = m_X \times \exp\left(-0.5s_X^2\right) \times \left[\exp(z \times s_X) - \exp(-z \times s_X)\right]$$  \hspace{1cm} (5)$$

When the lower and upper bounds of the IQR (i.e., $Q_1(X_0)$ and $Q_3(X_0)$ respectively) are reported, rather than the difference, $s_X$ may be calculated as $s_X = \left[Q_3(X) - Q_1(X)\right]/2z$.

The range is the difference between the maximum and the minimum value of the data. Equations (4) and (5) may be similarly used when the range is reported, but here we consider that the minimum and the maximum represent the lower and upper $(1/n)$ fraction of the dataset of $n$ observations. Therefore we expect a fraction $p = 1 - 1/(2n)$ below the minimum and the same fraction above the maximum, and in the equations above we need to use $z_p$. For example, in a dataset with $n = 100$ we use $z_{0.995} = 2.576$.

The geometric mean (gm) of the lognormal distribution is equal to $\exp(m_X)$, and is most often reported in papers together with the 95% confidence limits. $m_X$ and $s_X$ are obtained for these quantities using:

$$m_X = \ln(gm_X)$$

$$s_X = \sqrt{n} \times \left[\ln(X_{0,\text{upp}}) - \ln(X_{0,\text{low}})\right] / 2 \times z_{0.975}$$ \hspace{1cm} (7)$$

where $X_{0,\text{upp}}$ is the upper limit, $X_{0,\text{low}}$ is the lower limit of the 95% confidence interval and $z_{0.975} = 1.96$ represents the 97.5th percentage point in the standard normal distribution.

Then in step 2 for observational studies, $m_X$ and $s_X$ are calculated in case these estimates were not already available. These statistics at the original scale may be needed in the bivariate transformations described below. The equations are:

$$m_X_0 = \exp(m_X + 0.5s_X^2)$$ \hspace{1cm} (8)$$

$$s_X_0 = m_X_0 \times \sqrt{\exp(s_X^2) - 1}$$ \hspace{1cm} (9)$$

**Bivariate transformations (to obtain regression or correlation coefficients)**

For observational studies, the next step is to obtain an estimate of the correlation between $X$ and $Y$ ($r_{XY}$). The equations below can be used to obtain $r_{XY}$ from reported correlation and regression coefficients taking into account the possibility that either $X_0$, $\log_{10}(X_0)$, $X$, $Y_0$, $\log_{10}(Y_0)$ or $Y$ was used for the originally reported statistic.

When a study reports the association as a Spearman rank correlation coefficient ($r_S$), $r_{XY}$ is calculated as

$$r_{XY} = r_S$$ \hspace{1cm} (10)$$
Another option is that the association between $X_0$ and $Y_0$ is reported as a regression coefficient ($bY_0X_0$). In that case the correlation coefficient, $rX_0Y_0$, is calculated first using

$$r_{X_0Y_0} = bY_0X_0 \times \frac{sX_0}{sY_0}$$

(11)

and then $r_{XY}$ is calculated using the following equation which was derived from Johnson & Kotz [7]:

$$r_{XY} = \frac{\ln\left(1 + r_{X_0Y_0} \times \sqrt{\left[\exp\left(sX^2\right) - 1\right] \times \left[\exp\left(sY^2\right) - 1\right]}\right)}{sX \times sY}$$

(12)

This formula (12) is also used when the Pearson product–moment correlation coefficient $r_{X_0Y_0}$ is directly reported in a paper.

For observational studies that report the regression coefficient between $Y_0$ and $X$, the correlation coefficient, $r_{XY0}$, is calculated using

$$r_{XY0} = bY_0X \times \frac{sX}{sY_0}$$

(13)

When $\log_{10}(X_0)$ is used instead of $X$, $sX$ is replaced by $sX/\ln(10)$ in formula (13).

Then $r_{XY}$ is calculated using the following equation [8,9]:

$$r_{XY} = r_{XY_0} \times \frac{\sqrt{\exp(sY^2) - 1}}{sY}$$

(14)

This formula (14) is also used when $r_{XY_0}$ is reported directly or when the Pearson product–moment correlation coefficient is reported between $\log_{10}(X_0)$ and $Y_0$.

When the regression coefficient between $Y$ and $X_0$ is reported in an observational study, the regression coefficient, $r_{X_0Y}$, is calculated using

$$r_{X_0Y} = bYX_0 \times \frac{sX_0}{sY}$$

(15)

When $\log_{10}(Y_0)$ is used instead of $Y$, $sY$ is replaced by $sY/\ln(10)$ in formula (15).

Using $r_{X_0Y}$ or the directly reported Pearson product–moment correlation coefficient between $X_0$ and $\log_{10}(Y_0)$ or $Y$ in an observational study, $r_{XY}$ is calculated using [8,9]:

$$r_{XY} = r_{X_0Y} \times \frac{\sqrt{\exp(sX^2) - 1}}{sX}$$

(16)
When the regression coefficient between X and Y is reported, \( r_{XY} \) is calculated as

\[
r_{XY} = b_{YX} \frac{s_X}{s_Y}
\]

(17)

**Calculation of dose–response regression coefficient**

In the last step, for both observational studies and randomized controlled trials, we need to obtain \( b_{YX} \) and \( \text{se}(b_{YX}) \). For observational studies, the required regression coefficient \( b_{YX} \) is calculated from the correlation coefficient:

\[
b_{YX} = r_{XY} \frac{s_Y}{s_X}
\]

(18)

and the corresponding standard error (\( \text{se}(b_{YX}) \)) is calculated as

\[
\text{se}(b_{YX}) = \sqrt{\frac{s_Y^2 \times (1 - r_{XY}^2)}{(N - 2) \times s_X^2}}
\]

(19)

For randomized controlled trials, the required regression coefficient \( b_{YX} \) is calculated as:

\[
b_{YX} = \frac{m_{Y_{\text{int}}} - m_{Y_{\text{con}}}}{m_{X_{\text{int}}} - m_{X_{\text{con}}}}
\]

(20)

where ‘int’ indicates the intervention group and ‘con’ indicates the control or placebo group. The corresponding standard error is calculated as:

\[
\text{se}(b_{YX}) = \sqrt{\left(\frac{(N_{\text{con}} - 1) \times s_{Y_{\text{con}}}^2 + (N_{\text{int}} - 1) \times s_{Y_{\text{int}}}^2}{(N_{\text{con}} + N_{\text{int}} - 2)}\right) \times \left(\frac{1}{N_{\text{con}}} + \frac{1}{N_{\text{int}}}\right) \times \left(\frac{1}{m_{X_{\text{int}}} - m_{X_{\text{con}}}^2}\right)}
\]

(21)

**Simulation study**

A simulation study was conducted to validate the performance of the transformations given in this paper. Bivariate lognormal data (X,Y) were simulated where X \~\text{Normal}(1.60,0.85^2) and Y \~\text{Normal}(5.70,0.45^2). Parameter values were based on values of vitamin B12 intake (X) and serum/plasma vitamin B12 (Y) [10-13]. Different strengths of the correlation between X and Y were simulated, namely 0.1, 0.5 and 0.9.

A sample of individuals (with sample size 100, 200 or 500) was randomly drawn, and values that represent different often used reporting options were calculated from this sample, namely the mean and SD, the median and IQR, the median and range and the geometric mean and 95% CI (all summary statistics on the absolute scale). Also, the correlation and regression coefficients of X and Y expressed in different scales were calculated. These ‘reported’ values were rounded to two decimal places. From these ‘reported’ values, the parameter estimates
mX, mY, sX, sY and rXY were calculated using the transformations described in this paper. This process was repeated 1000 times.

Results

Table 1 shows the simulation results for the univariate statistics. On average the calculated values of mX and mY are almost the same as the true values, indicating that no important bias is present in these calculations. As expected, the 95% CI of the simulations is smaller for the simulations with a sample size of 500 than for the simulations with a sample size of 200 or 100. For sX and sY, the estimates are most precise when a geometric mean with a 95% CI is reported, and least precise when a median with a range is reported.

Table 1 Simulation results for mX, sX, mY and sY

<table>
<thead>
<tr>
<th>n</th>
<th>mX</th>
<th>sX</th>
<th>mY</th>
<th>sY</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>1.6</td>
<td>0.85</td>
<td>5.7</td>
<td>0.45</td>
</tr>
<tr>
<td>Mean, SD</td>
<td>100</td>
<td>1.6 (1.4-1.8)</td>
<td>0.82 (0.65-1.06)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.6 (1.4-1.7)</td>
<td>0.83 (0.70-1.03)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.6 (1.5-1.7)</td>
<td>0.84 (0.75-0.98)</td>
<td>5.7 (5.7-5.7)</td>
</tr>
<tr>
<td>Median, IQR</td>
<td>100</td>
<td>1.6 (1.4-1.8)</td>
<td>0.84 (0.63-1.10)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.6 (1.5-1.8)</td>
<td>0.85 (0.70-1.02)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.6 (1.5-1.7)</td>
<td>0.85 (0.76-0.95)</td>
<td>5.7 (5.7-5.7)</td>
</tr>
<tr>
<td>Median, range</td>
<td>100</td>
<td>1.6 (1.4-1.8)</td>
<td>0.83 (0.58-1.14)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.6 (1.5-1.8)</td>
<td>0.83 (0.63-1.12)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.6 (1.5-1.7)</td>
<td>0.83 (0.68-1.06)</td>
<td>5.7 (5.7-5.7)</td>
</tr>
<tr>
<td>Gm, 95% CI</td>
<td>100</td>
<td>1.6 (1.4-1.8)</td>
<td>0.85 (0.73-0.97)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.6 (1.5-1.7)</td>
<td>0.85 (0.77-0.94)</td>
<td>5.7 (5.6-5.8)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.6 (1.5-1.7)</td>
<td>0.85 (0.80-0.90)</td>
<td>5.7 (5.7-5.7)</td>
</tr>
</tbody>
</table>

Figure 2 shows the simulation results when a correlation coefficient is reported, and Figure 3 shows the simulation results when a linear regression coefficient is reported. Both these figures show the simulation results with true rXY = 0.5. Results are similar for true rXY = 0.9 and true rXY = 0.1 (data not shown). For the situation in which a correlation coefficient is the reported bivariate statistic, there is no difference for the four univariate reporting options. Therefore, these results are pooled in Figure 2.

Figure 2 Simulation results for rXY where the true rXY was 0.5. Circles indicate that the reported bivariate statistic was rX0Y0, squares indicate rXY0 and diamonds indicate rX0Y. Bars represent 95% confidence intervals. In each group the three bars from left to right are for sample sizes of 100, 200 and 500 individuals, respectively.

Figure 3 Simulation results for rXY from different reporting options where the true rXY was 0.5. A) b from linear regression of Y0 on X0; B) b from linear regression of Y on X0; C) b from linear regression of Y0 on X; D) b from linear regression of log10(Y0) on log10(X0). Circles indicate that the reported univariate statistic was mean ± SD, squares indicate median and IQR, diamonds indicate median and range, and triangles indicate geometric mean. Bars represent 95% confidence intervals. In each group the three bars from left to right are for sample sizes of 100, 200 and 500 individuals, respectively.
None of the combinations of univariate and bivariate reporting options shows evidence of bias with the average of the simulations almost equal to the true value. The width of the confidence interval indicates the variability of the simulations. Because there is no appreciable bias, a smaller CI width indicates that the individual simulations are closer to the true correlation. The accuracy is best when $rxY$ is reported and worst when $rxy$ is reported. As expected, the accuracy is also better when the sample size is larger. Figure 3 shows that the CI is wider when the reported univariate statistics are the median and IQR or median and range. The larger variation in the results for the transformation from $bYx$ (Figure 3B) compared with the variation in the results from $byX$ (Figure 3C) is caused by the fact the $X$ was simulated with larger standard deviation than $Y$.

**Example**

To illustrate the methodology some examples of its use on real data for vitamin B12 are reported in Table 2 (observational studies [14,15]) and Table 3 (randomized controlled trials [16,17]). The tables show the statistics as reported in the studies and the statistics that are calculated using the different equations presented in this paper (which are entitled ‘required statistics’ in the tables).
### Table 2 Example statistics for observational studies on vitamin B12 intake (X) and vitamin B12 status (Y)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Observed univariate statistics</th>
<th>Observed bivariate statistic</th>
<th>Required statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type X₀ and Y₀</td>
<td>X₀</td>
<td>Y₀</td>
</tr>
<tr>
<td>[14]</td>
<td>Mean, SD</td>
<td>9.3, 9.3</td>
<td>330, 140</td>
</tr>
<tr>
<td>[15]</td>
<td>gm, 95% CI</td>
<td>7.3, 7.1-7.5</td>
<td>354, 348-360</td>
</tr>
</tbody>
</table>

### Table 3 Example statistics for randomized controlled trials on vitamin B12 intake and vitamin B12 status

<table>
<thead>
<tr>
<th>Reference</th>
<th>Observed univariate statistics</th>
<th>Required statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X₀*</td>
<td>Type Y₀</td>
</tr>
<tr>
<td>[16]</td>
<td>intervention</td>
<td>405</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>5</td>
</tr>
<tr>
<td>[17]</td>
<td>intervention</td>
<td>505</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>5</td>
</tr>
</tbody>
</table>

*) X₀ represents the dose provided plus the dietary intake. When dietary intake of vitamin B₁₂ was not reported 5 µg/day was added to the provided dose. The 5 µg/day was calculated as the average dietary intake from several studies.
Discussion

The investigated means, standard deviations, correlation coefficients and sample sizes were based on real-life values. The univariate statistics that are investigated in this paper were limited to mean and SD, median and IQR or range and geometric mean and 95% CI. These do not represent all reporting options that can be encountered in the literature, but cover most published papers. Other combinations of univariate statistics that were seen are for example mean with IQR, mean with range, and geometric mean with standard deviation. Also, the investigated regression and correlation coefficients are limited in this paper to those on the absolute or logarithmic scale, whereas sometimes other transformations to normality have been used in reports, such as a square root transformation. However, as the logarithmic transformation is by far the most often used transformation in papers in the medical research area, the equations in this paper will cover most published papers in this field.

The bivariate normal linear model on the logarithmic scale is an approximation that is used here because the data are positive data. Note that it allows the relationship between $X_0$ and $Y_0$ to be a linear, monotonic convex or monotonic concave function (i.e., for a slope equal, higher or lower than one, respectively). Even though some randomized controlled trials may investigate the dose–response relationship by providing multiple dosages in their study, most of these studies include only one intervention and one control group and consequently it is often unknown what the true relationship is. Therefore, this approximation provides a practical methodology to estimate the dose–response relationship and to combine the results from randomized controlled trials and observational studies. It was outside the scope of the simulation study to investigate other shapes of the dose–response relation.

The transformations in this paper consider reported regression and correlation coefficients that are unadjusted for other variables. It is possible to adjust the equations for adjusted regression or correlation coefficients, if these adjustments were done on the log-scale. However, most often adjustment has been done on another scale, and moreover studies do not report all required statistics. Therefore, we did not consider adjusted coefficients.

In this paper we presented a methodology that allows for information from RCTs and observational studies to be summarised in comparable statistics. One possible application is to combine results of both types of study in a single meta-analysis. In general, a meta-analysis should include as much information as possible. However, there may be systematic differences between observational studies and randomized controlled trials. Therefore, it is advisable to check whether the size of the estimated regression coefficient differs between this different study designs. This may be done by stratified analysis or by using meta-regression techniques.

Conclusions

The presented methodology provides calculations to use results from published literature to estimate the slope of the dose–response relation incorporating information from both randomized controlled trials and observational studies. The simulations clearly show that there is no observable bias associated with the transformations. Also, it can be seen that when a regression coefficient is reported, it is preferable to report the univariate statistics as mean and SD or geometric mean and 95% CI rather than as median with IQR or range.
Abbreviations

B, Regression coefficient; CI, Confidence interval; Gm, Geometric mean; IQR, Interquartile range; M, Mean; Med, Median; R, Correlation coefficient; S, Standard deviation; Se, Standard error

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

OS participated in the design of the simulation study, performed the statistical analysis and drafted the manuscript. CD helped to draft the manuscript and participated in the design of the simulation study. PvtV participated in the coordination of the study and revised the manuscript critically. HvdV conceived of the study, helped with the statistical analysis and interpretation of the data and revised the manuscript critically. All authors read and approved the final manuscript.

Authors’ information

OS and CD are both postdoctoral research fellows at the Division of Human Nutrition of Wageningen University, the Netherlands. PvtV is professor of Nutrition and Epidemiology at the Division of Human Nutrition, the Netherlands. HvdV is statistician at Biometris, Wageningen University and Research centre, the Netherlands.

Acknowledgements

We would like to thank the reviewer, Wolfgang Viechtbauer, for his valuable comments to the manuscript.

The work reported herein has been carried out within the EURRECA Network of Excellence (www.eurreca.org) which is financially supported by the Commission of the European Communities, specific Research, Technology and Development (RTD) Programme Quality of Life and Management of Living Resources, within the Sixth Framework Programme, contract no. 036196. This report does not necessarily reflect the Commission's views or its future policy in this area.

References


OBSERVATIONAL STUDIES

Step 1
Calculate mX, sX and mY, sY from reported univariate statistic using equations (1)-(7)

Step 2
Calculate mX₀, sX₀ and mY₀, sY₀ from mX, sX and mY, sY: equations (8), (9)

Step 3
Calculate rXY from reported bivariate statistic using equations (10)-(17)

Step 4
Calculate bYX from rXY: equation (18)
Calculate se(bYX): equation (19)

RANDOMIZED CONTROLLED TRIALS

Step 1
Calculate mY and sY from reported values after intervention for placebo and intervention group using equations (1)-(7)

Step 2
Calculate mX as ln(mX₀) for both placebo and intervention group

Step 3
Calculate bYX: equation (20)
Calculate se(bYX): equation (21)
Figure 2
Figure 3

A. from $bY_0X_0$

B. from $bYX_0$

C. from $bY_0X$

D. from $bYX$