Visualizations of marginal and conditional quantiles based on weighted mixed-effects models

Nicola Orsini

Department of Global Public Health
Karolinska Institutet

2022 Northern European Stata Conference, Oslo

October 12, 2022
Outline

• Examples and common practice
• Weighted mixed-effects models
• Marginal and conditional quantiles
• Post-estimation visualization tool in Stata/Python
• Applications to simulated and empirical data
• Final remarks
Blood Pressure Effects of Sodium Reduction: Dose–Response Meta-Analysis of Experimental Studies. *Circulation* 2021

![Graphs showing the effects of sodium reduction on blood pressure](image)

**Figure 2.** Dose–response meta-analysis of changes in SBP and DBP levels (mm Hg) according to achieved sodium excretion in the treatment and control groups at the end of the trials (all studies) and by type of intervention (supplementation or diet).

The average curve (solid line) with 95% confidence limits (dashed lines) was estimated with a 1-stage random-effects restricted cubic spline model, using 2 g/d as reference. DBP indicates diastolic blood pressure, and SBP systolic blood pressure.
Examination of Dosing of Antipsychotic Drugs for Relapse Prevention in Patients With Stable Schizophrenia. *JAMA Psychiatry* 2021

**Figure 1. Relapse**

The dose-response curve for the primary outcome relapse after pooling all drugs using the primary scientific dose-equivalence method (the maximum effective dose method). The marks on the x-axis indicate for which doses data from study arms were available. A total of 26 studies with 71 individual dose arms including 4749 patients were included (1 publication reported on 2 studies). The shaded areas indicate 95% CIs for the primary outcome.
Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *Lancet Public Health* 2022

*Figure 3: Dose-response association between steps per day and all-cause mortality, by age group*

Thick lines indicate hazard ratio estimates, with shaded areas showing 95% CIs. Reference set at the median of the medians in the lowest quartile group (age ≥60 years = 3000 steps per day and <60 years = 5000 steps per day). Model is adjusted for age, accelerometer wear time, race and ethnicity (if applicable), sex (if applicable), education or income, body-mass index, and study-specific variables for lifestyle, chronic conditions or risk factors, and general health status. Pinteraction=0.012 by age group. 14 studies included in spline analysis, excluded Baltimore Longitudinal Study of Aging. The y-axis is on a log scale.
What’s in common in these examples?

• There is a quantitative factor measured in either experimental or observational studies

• Effect measures can be of any type (mean difference, odds ratios, hazard ratios)

• Research questions are about the shape of the dose-response relationship or some specific less known aspects of it

• Design of the meta-analysis can be either retrospective (previously published) or prospective (pooling projects)

• A mixed model is used to learn from multiple tables of correlated empirical estimates

• Only 3 quantiles (0.025, 0.500, and 0.975) of the marginal dose-response relationship are shown graphically
Weighted Mixed Effects Models

A one-stage approach for meta-analysis of summarized dose-response data has been proposed in the general framework of linear mixed effects model (Stat Meth Med Res, 2019).

\[ \hat{\gamma}_i = X_i \beta + Z_i b_i + \epsilon_i \]

\[ b_i \sim N(0, \Psi) \]

\[ \epsilon_i \sim N(0, S_i) \]

\( \hat{\gamma}_i \) is the vector of empirical contrasts (i.e. mean differences, log odds ratios, log hazard ratios) estimated relative to a common referent in the \( i \)-th study.

ML/REML estimates can be obtained with the dosresmeta package in R (Crippa & Orsini J Stat Soft) and drmeta command (Orsini, Stata J) in Stata.
Flexible modelling using splines

More info on Chapter 18 *Handbook of Meta-Analysis.*
Quantiles of confidence

• **Marginal quantiles.** What is the degree of confidence that can be assigned to an *inequality* regarding the *unknown* effect of the dose on a *typical study* in light of the data and specified statistical model?

• **Conditional quantiles.** What is the degree of confidence that can be assigned to an *inequality* regarding the *unknown study-specific effects* of the dose in light of the data and specified statistical model?
What is the point?

- If the confidence in an effect of the dose in a typical study below $k$ is 0.5, then it follows that the confidence in study effects below $k$ is also 0.5.

- Large discrepancies between marginal and conditional quantiles, eventually in opposite direction, would indicate a large uncertainty in dose effects.
Ideal visualization tool

- it works with a variety of study designs, dose transformations, and outcome measures
- it allows the investigator to derive any quantile (0.01 to 0.99) of the point-wise conditional and marginal dose-response relationship
- it allows the investigator to define a fine grid of dose values and to choose a common referent
- it shades quantiles differently according to the degree of confidence
- it allows the user to overlay the study-specific BLUPs
- easily provides both static (research article) and interactive visualizations (dissemination)

So I wrote `drmeta.het` using Plotly Python Graphing Library taking advantage of the recent Stata/Python integration.
A simple hypothetical example

The random-effect linear dose-response mechanism is $\beta_i \sim N(0.5, 0.2)$. Consider $I = 10$ studies of the same size $n = 1000$, equal dose distribution $X \sim \chi^2(5)$, and equal conditional outcome std deviation $\sigma_{Y_i} = 10$. Using a dose of 5 units as referent we have that

$$\beta_i(x - 5) \sim N(0.5(x - 5), 0.2(x - 5))$$

the typical standard error of the slope in any similar study would be

$$\hat{SE}(\hat{\beta}_i) = \frac{10}{\sqrt{5(2)}} \frac{1}{\sqrt{1000 - 1}} = 0.1$$

and the typical standard error of the slope for the average study would be

$$\hat{SE}(\hat{\beta}) = \frac{1}{\sqrt{1/(0.1^2 + 0.2^2)10}} = 0.07$$
Marginal vs Conditional Quantiles

The degree of confidence (C) in the inequality below is $p$

$$C(\beta(x - 5) \leq Q_p^M(\hat{\beta}(x - 5))) = p$$

where the **marginal** quantile would be

$$Q_p^M(\hat{\beta}(x - 5)) = 0.5(x - 5) + \phi^{-1}(p)\sqrt{(0.07^2)(x - 5)^2}$$

And similarly, the degree of confidence in the proposition below is $p$

$$C(\beta_i(x - 5) \leq Q_p^C(\hat{\beta}_i(x - 5))) = p$$

where the **conditional** quantile would be

$$Q_p^C(\hat{\beta}_i(x - 5)) = 0.5(x - 5) + \phi^{-1}(p)\sqrt{(0.07^2 + 0.2^2)(x - 5)^2}$$
Similar marginal but different conditional quantiles

In Scenario a), with smaller heterogeneity, the confidence in negative study-specific dose effects (15 vs 5 units) is about 0%

\[ C(\beta_i(15 - 5) \leq 0) = 0.009 \]

In Scenario b), with larger heterogeneity, the confidence in negative study-specific dose effects (15 vs 5 units) is about 20%

\[ C(\beta_i(15 - 5) \leq 0) = 0.214 \]
Let’s consider two transformations (i.e. splines, fractional polynomials), saying $f_1(x)$ and $f_2(x)$, of the original dose.

$$\beta_1(x) = \beta_1(f_1(x) - f_2(x)) + \beta_2(f_2(x) - f_2(x_0))$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_2 \end{bmatrix} \right)$$

The model is in terms of 2 fixed-effects plus 2 variances and 1 covariance of the 2 random-effects.

At this point, it helps to use a compact matrix notation

$$\beta_i \sim \mathcal{N}(\beta, \Psi)$$
Quantiles for the marginal and conditional dose-response relationship

Marginal

\[ Q_p^M = (X^* - x_0^*)\hat{\beta} + \phi^{-1}(p)\text{diag}[(X^* - x_0^*)V(\hat{\beta})(X^* - x_0^*)']^{1/2} \]

Conditional

\[ Q_p^C = (X^* - x_0^*)\hat{\beta} + \phi^{-1}(p)\text{diag}[(X^* - x_0^*)(V(\hat{\beta}) + \hat{\Psi})(X^* - x_0^*)']^{1/2} \]

where

- \( X^* \) indicates a matrix of user specified transformations
- \( x_0^* \) indicates a matrix of reference values
. use http://www.stats4life.se/data/md_10_studies, clear
. list id md dose semd n sd if id <= 3, sepby(id) noobs

<table>
<thead>
<tr>
<th>id</th>
<th>md</th>
<th>dose</th>
<th>semd</th>
<th>n</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.4142542</td>
<td>2.043162</td>
<td>0.762961</td>
<td>334</td>
<td>9.785315</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>4.466444</td>
<td>0</td>
<td>333</td>
<td>9.918493</td>
</tr>
<tr>
<td>1</td>
<td>2.280955</td>
<td>8.789105</td>
<td>0.8060276</td>
<td>333</td>
<td>10.8613</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2.071578</td>
<td>0</td>
<td>334</td>
<td>10.41234</td>
</tr>
<tr>
<td>2</td>
<td>1.425228</td>
<td>4.396238</td>
<td>0.787194</td>
<td>333</td>
<td>9.912573</td>
</tr>
<tr>
<td>2</td>
<td>3.382661</td>
<td>8.83116</td>
<td>0.7879598</td>
<td>333</td>
<td>9.932815</td>
</tr>
<tr>
<td>3</td>
<td>-3.110585</td>
<td>2.120871</td>
<td>0.7938005</td>
<td>334</td>
<td>10.62627</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>4.425666</td>
<td>0</td>
<td>333</td>
<td>9.861535</td>
</tr>
<tr>
<td>3</td>
<td>1.759263</td>
<td>8.376437</td>
<td>0.7596498</td>
<td>333</td>
<td>9.742396</td>
</tr>
</tbody>
</table>
Obtain estimates of the weighted mixed model using a linear function

\[ \hat{\gamma}_{ij} = (\beta_1 + b_{1i})x_{ij} + \epsilon_{ij} \]

```
. drmeta md dose, se(semd) data(n sd) id(id) type(type_md) ml stddev

One-stage random-effects dose-response model
Number of studies = 10
Optimization = ml
Number of obs = 26
AIC = 78.78
Model chi2(1) = 88.06
Log likelihood = -37.388447 Prob > chi2 = 0.0000

------------------------------------------------------------------
 | Coefficient Std. err. z P>|z| [95% conf. interval]
|---------------+--------------------------------------------------------
| dose | .5375663 .0572842 9.38 0.000 .4252913 .6498414
------------------------------------------------------------------

Random-effects parameters

<table>
<thead>
<tr>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>std(dose,dose)</td>
</tr>
</tbody>
</table>

LR test vs. no random-effects model = 4.9471603 Prob >= chi2(1) = 0.0261
```
Present marginal + conditional quantiles

drmeta_het, dose(4(.1)8) ref(5) eq(d) iqc iqm
drmeta_het, dose(4(.1)8) ref(5) eq(d) iqc iqm iqcb

Click Here
Obtain estimates of the weighted mixed model using a restricted cubic spline function

\[
\hat{\gamma}_{ij} = (\beta_1 + b_{1i})s_1(x)_{ij} + (\beta_2 + b_{2i})s_2(x)_{ij} + \epsilon_{ij}
\]

mkspline doses = dose, nk(3) cubic displayknots

<table>
<thead>
<tr>
<th></th>
<th>knot1</th>
<th>knot2</th>
<th>knot3</th>
</tr>
</thead>
<tbody>
<tr>
<td>dose</td>
<td>1.994129</td>
<td>4.410952</td>
<td>8.83116</td>
</tr>
</tbody>
</table>

matrix knots = r(knots)

drmeta md doses1 doses2, se(semd) data(n sd) id(id) type(type_md) ml stddev

One-stage random-effects dose-response model

Number of studies = 10
Optimization = ml
AIC = 82.47
Model chi2(2) = 101.94
Log likelihood = -36.233006
Prob > chi2 = 0.0000

|   | Coefficient | Std. err. | z   | P>|z| | [95% conf. interval] |
|---|-------------|-----------|-----|-----|----------------------|
| doses1 | .5798272   | .1629389  | 3.56 | 0.000 | .2604728 .8991816    |
| doses2 | -.0728658  | .2150756  | -0.34 | 0.735 | -.4944062 .3486746  |

Random-effects parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>std(doses1,doses1)</td>
<td>.3478605</td>
</tr>
<tr>
<td>std(doses2,doses2)</td>
<td>.3439548</td>
</tr>
<tr>
<td>corr(doses1,doses1)</td>
<td>-1</td>
</tr>
</tbody>
</table>
Present marginal + conditional quantiles + BLUPs

drmeta_het, matk(knots) dose(4(.1)8) ref(5) iqm iq c iqcb

Click Here
Simulated Example: Walking and mortality

• Consider 30 prospective cohort studies investigating the association between baseline walking, measured in hours/week, and time until death, or end of follow-up (10 years), whichever came first (Stata J, 2021).

• Age is inversely associated with walking levels and positively associated with higher mortality rates independently of walking levels.

• The true summary age-adjusted mortality hazard ratio is decreasing with higher walking levels with a threshold effect at 2 hours per week

\[ HR = e^{-0.5(x-2)+0.5(x>2)(x-2)} \]
. use http://www.stats4life.se/data/hr_drm, clear
. list id walk b seb case py if inlist(id, 1, 20, 23)

<table>
<thead>
<tr>
<th>id</th>
<th>walk</th>
<th>b</th>
<th>seb</th>
<th>case</th>
<th>py</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3</td>
<td>1.13</td>
<td>0.11</td>
<td>229</td>
<td>777</td>
</tr>
<tr>
<td>1</td>
<td>2.4</td>
<td>0.00</td>
<td>0.00</td>
<td>137</td>
<td>1704</td>
</tr>
<tr>
<td>20</td>
<td>0.1</td>
<td>0.21</td>
<td>0.10</td>
<td>239</td>
<td>674</td>
</tr>
<tr>
<td>20</td>
<td>0.5</td>
<td>0.00</td>
<td>0.00</td>
<td>216</td>
<td>946</td>
</tr>
<tr>
<td>20</td>
<td>1.5</td>
<td>-1.04</td>
<td>0.11</td>
<td>133</td>
<td>1773</td>
</tr>
<tr>
<td>20</td>
<td>4.1</td>
<td>-2.63</td>
<td>0.19</td>
<td>32</td>
<td>2318</td>
</tr>
<tr>
<td>23</td>
<td>0.2</td>
<td>0.65</td>
<td>0.09</td>
<td>311</td>
<td>973</td>
</tr>
<tr>
<td>23</td>
<td>0.9</td>
<td>0.00</td>
<td>0.00</td>
<td>247</td>
<td>1765</td>
</tr>
<tr>
<td>23</td>
<td>3.4</td>
<td>-1.28</td>
<td>0.12</td>
<td>101</td>
<td>2752</td>
</tr>
</tbody>
</table>
Plotting the empirical contrasts

-or-

Walking (hours/week)

Adjusted Hazard Ratio

Orsini N (GPH, KI)

M and C Quantiles

October 12, 2022
We specify a dose-response model with constant change for the age-adjusted log mortality hazard ratio associated with every 1 hour per week increase in walking before and after the knot at 2 hours per week.

\[ \hat{\gamma}_{ij} = (\beta_1 + b_{1i})x_{ij} + (\beta_2 + b_{2i})I(x_{ij} > 2)(x_{ij} - 2) + \epsilon_{ij} \]
Obtain the estimates of the model

```
.drmeta b walk walkplus, se(seb) data(py case) type(type) id(id) ml
```

One-stage random-effects dose-response model

<table>
<thead>
<tr>
<th>Optimization</th>
<th>Number of studies =</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>ml</td>
<td>Number of obs =</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>AIC = 37.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Model chi2(2) =</td>
<td>110.27</td>
</tr>
<tr>
<td></td>
<td>Log likelihood =</td>
<td>-13.773298</td>
</tr>
<tr>
<td></td>
<td>Prob &gt; chi2 =</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| b          | Coefficient | Std. err. | z       | P>|z|   | [95% conf. interval] |
|------------|-------------|-----------|---------|-------|---------------------|
| walk       | -.4678671   | .0536744  | -8.72   | 0.000 | -.573067            | -.3626673          |
| walkplus   | .5432787    | .0626324  | 8.67    | 0.000 | .4205213            | .666036            |

Random-effects parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>var(walk,walk)</td>
<td>.0766958</td>
</tr>
<tr>
<td>var(walkplus,walkplus)</td>
<td>.0507463</td>
</tr>
<tr>
<td>cov(walk,walkplus)</td>
<td>-.0136841</td>
</tr>
</tbody>
</table>

LR test vs. no random-effects model = 2713.6

Prob >= chi2(3) = 0.0000
Graph marginal and conditional quantiles

drmeta_het, eq(d (d>2)*(d-2) ) dose(0(.1)4) ///
ref(2) ///
ytitle("Adjusted Hazard Ratio (log)") ///
xtitle("Brisk walking (hours/week)") ///
iqm iqc iqcbm

Click Here
We combine the dose-response relation between alcohol intake and colorectal cancer rate arising from 8 prospective cohort studies including 489,979 women and men participating in the Pooling Project of Prospective Studies of Diet and Cancer. A total of 3,646 cases and 2,511,424 person-years are included in this analysis.

```
use http://www.stats4life.se/data/ex_alcohol_crc.dta, clear
mkspline doses = dose, nk(3) cubic
mat knots = r(knots)

drmeta logrr doses1 doses2 , data(peryears cases) id(study) type(type) se(se) ml

drmeta_het , dose(0(4)70) ref(12) matk(knots) eform ytitle("Relative Risk") xtitle("Alcohol Intake (mg/d)") iqc iqm
```
Alcohol intake and colorectal cancer risk

Click Here
We use data from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. The SEER program provides data about cancer statistics from several population-based registries in the USA (http://seer.cancer.gov) from San Francisco-Oakland, Connecticut, Metropolitan Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Metropolitan Atlanta that here are considered as different studies. Analysis are based on 9 studies on prognostic factors for breast cancer survival including a total of 84,404 women. During 554,812 person-years, 8,520 women died from breast cancer.

More info on Chapter 18 *Handbook of Meta-Analysis*. 
use http://www.stats4life.se/data/seer_sd_drm, clear

mkspline ages = age, knot(42 61 78) cubic displayknots
mat knots = r(knots)
drmeta logrr ages1 ages2, se(se) data(py case) id(regID) type(type)

drmeta_het, list dose(40(2)80) ref(60) matk(knots) ///
   ytitle("Mortality Rate Ratio") ///
   xtitle("Age, years") ///
eform iqm iqc iqcb
Age and breast cancer mortality

Click Here
Final remarks

- Based on data and statistical model, quantiles can help expressing a degree on confidence in inequalities regarding unknown quantities.
- The post-estimation command `drmeta_het` allows the user to explore and compare marginal and conditional quantiles (not just 3 of them) of the dose-response relationship.
- The visualization tool is widely applicable to different study designs and effect measures.
- Quantiles were derived from a standard normal distribution but it can be extended.
A couple of book chapters on dose-response meta-analysis


