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

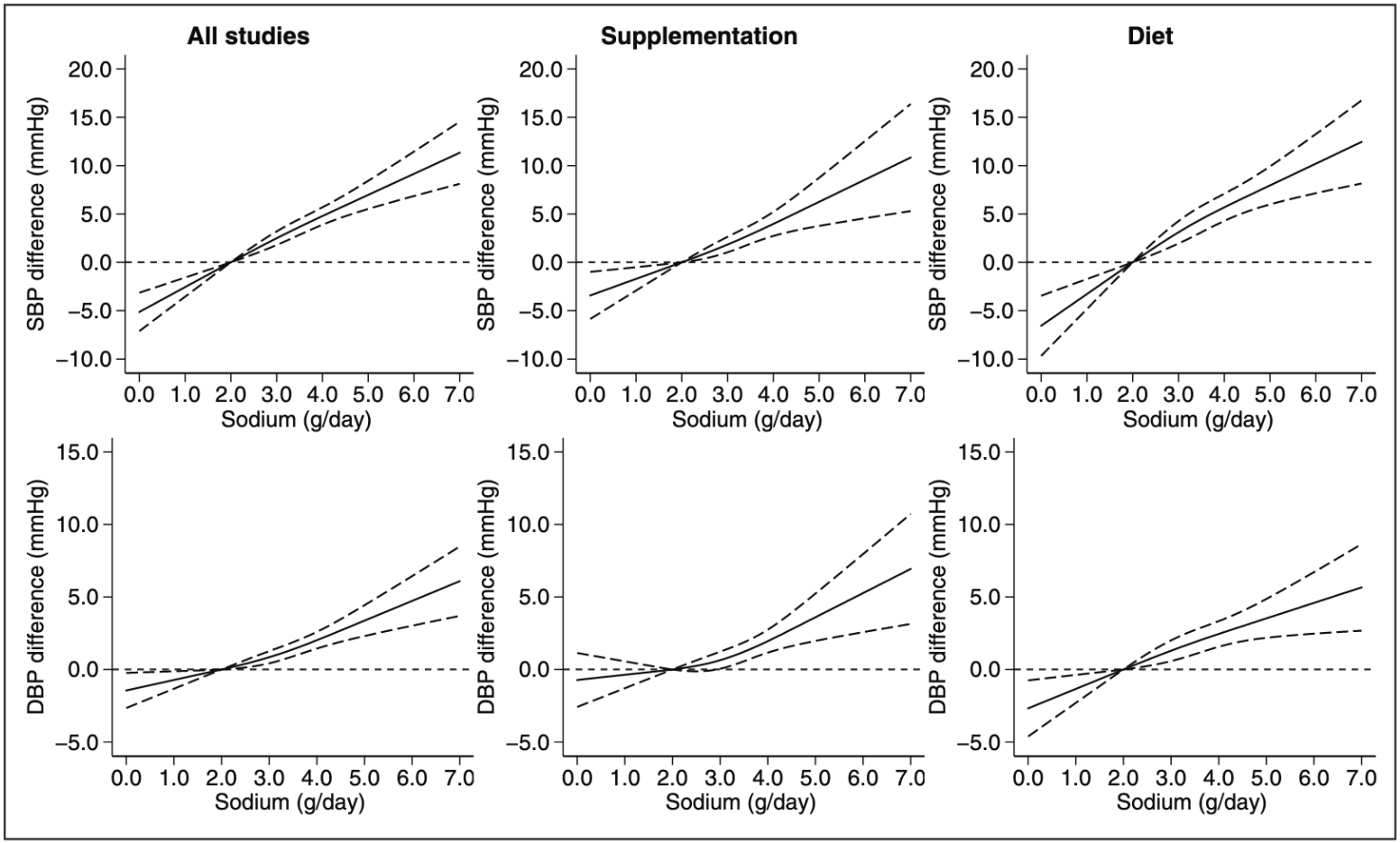
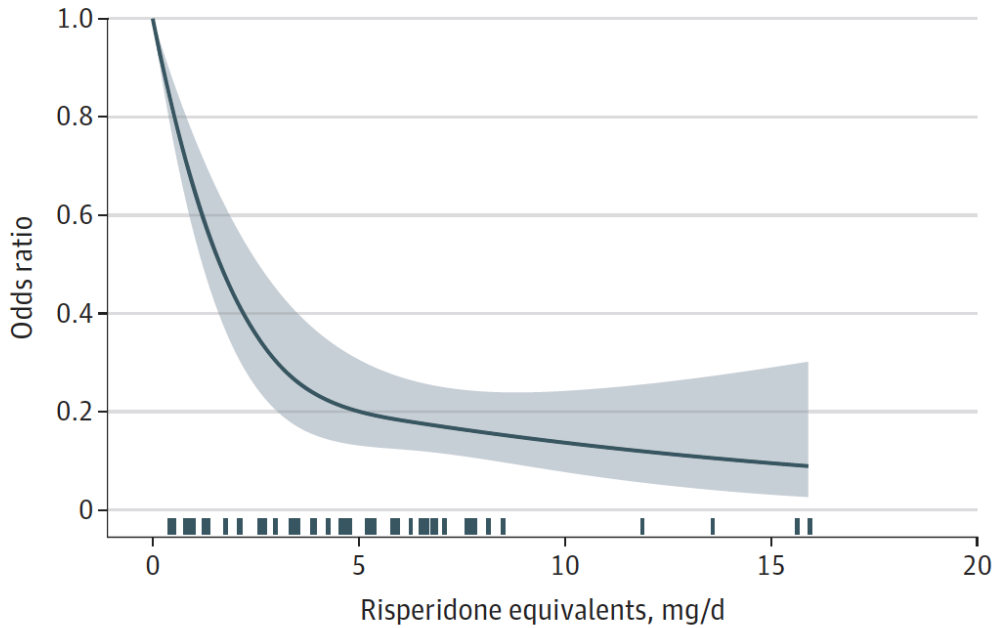


Figure 2. Dose–response meta-analysis of changes in SBP and DBP levels (mm Hg) according to achieved sodium excretion in the treatment and control groups at the end of the trials (all studies) and by type of intervention (supplementation or diet).
The average curve (solid line) with 95% confidence limits (dashed lines) was estimated with a 1-stage random-effects restricted cubic spline model, using 2 g/d as referent. 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).[27,32-55] 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

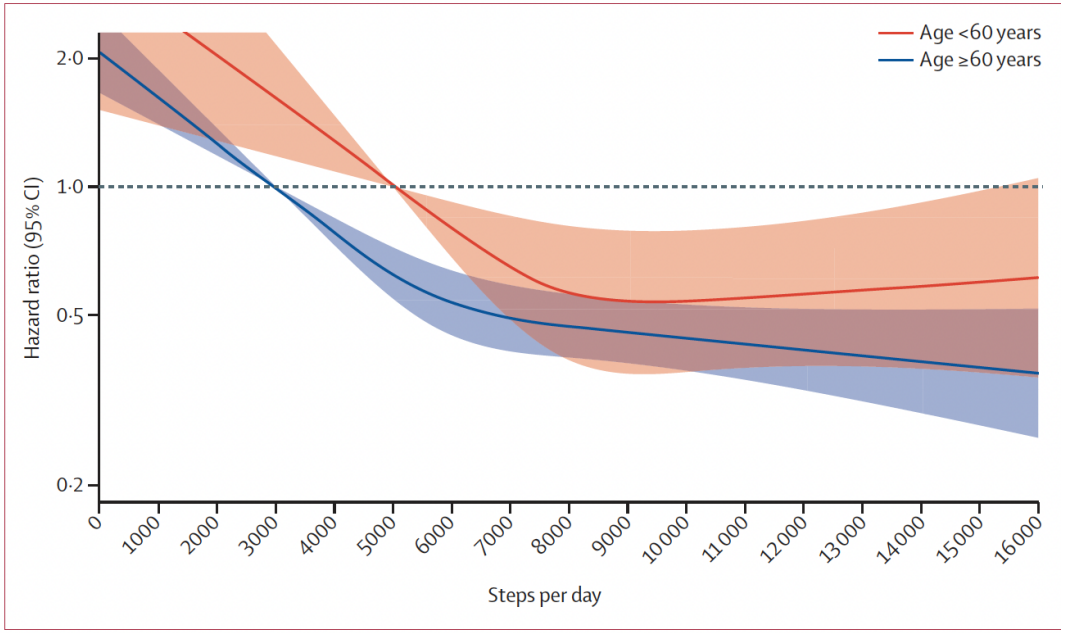


Figure 3: **Dose-response association between steps per day and all-cause mortality, by age group**
Thick lines indicate hazard ratio estimates, with shaded areas showing 95% CIs. Reference set at the median of the medians in the lowest quartile group (age ≥60 years = 3000 steps per day and <60 years = 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. $p_{interaction}$=0·012 by age group. 14 studies included in spline analysis, excluded Baltimore Longitudinal Study of Aging.[19] 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).

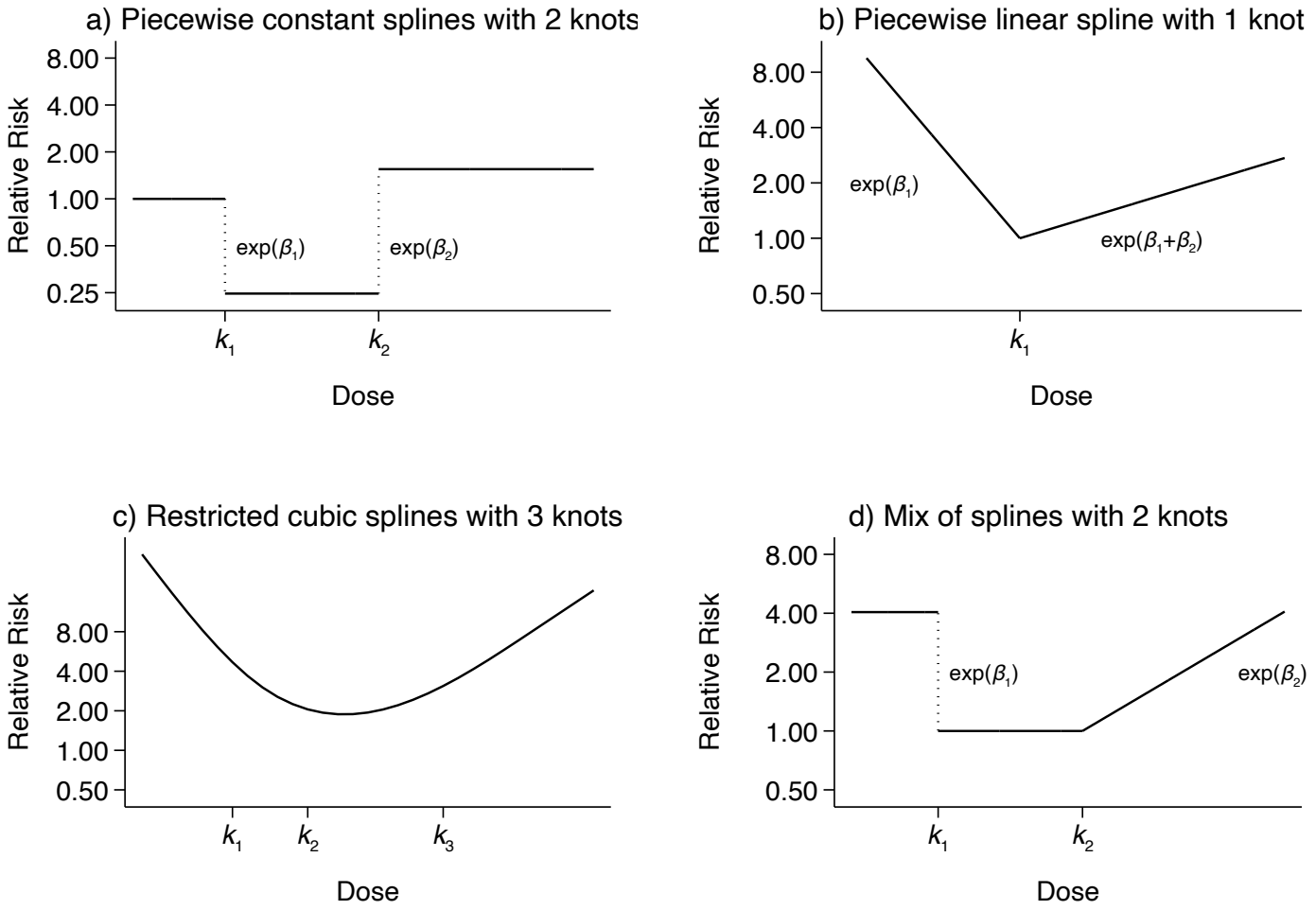
$$\hat{\boldsymbol{\gamma}}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$$
$$\epsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_i)$$

$\hat{\boldsymbol{\gamma}}_i$ is the vector of empirical constrasts (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

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

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

$$\widehat{SE}(\hat{\beta}) = 1/\sqrt{1/(0.1^2+0.2^2)10} = 0.07$$

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

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

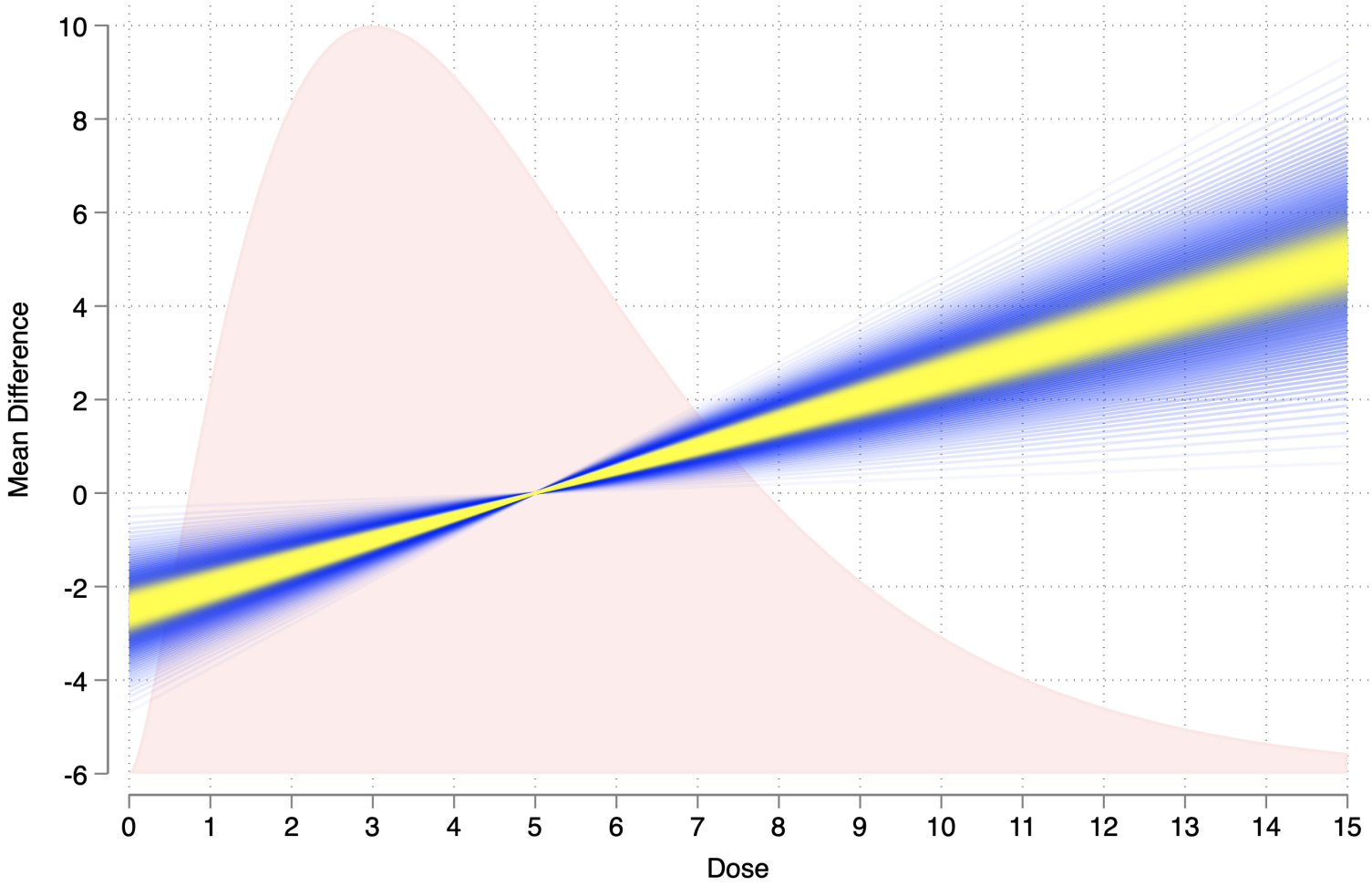
And similarly, the degree of confidence in the proposition below is p

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

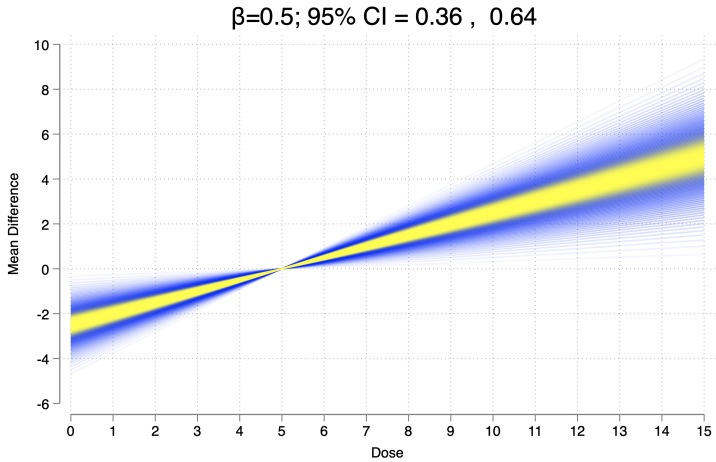
where the **conditional** quantile would be

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

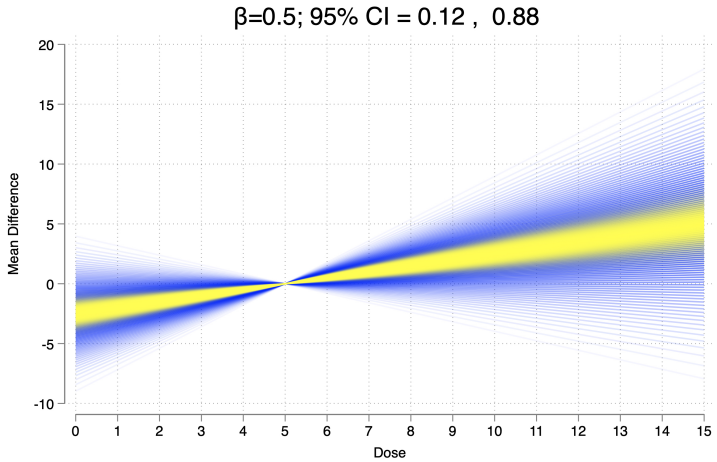
Marginal and conditional quantiles



Similar marginal but different conditional quantiles



(a) $\tau = 0.2$



(b) $\tau = 0.6$

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%

$$C(\beta_i(15-5) \leq 0) = 0.214$$

Extend the reasoning

Let's consider two transformations (i.e. splines, fractional polynomials), saying $f_1(x)$ and $f_2(x)$, of the original dose.

$$\beta_{1i}(f_1(x) - f_2(x_0)) + \beta_{2i}(f_2(x) - f_2(x_0))$$

$$\begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} \xi_1 & \\ \xi_3 & \xi_2 \end{bmatrix} \right)$$

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

$$\boldsymbol{\beta_i} \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Psi})$$

Quantiles for the marginal and conditional dose-response relationship

Marginal

$$Q_p^M = (\boldsymbol{X}^* - \boldsymbol{x}_0^*)\hat{\boldsymbol{\beta}} + \phi^{-1}(p)\mathrm{diag}[(\boldsymbol{X}^* - \boldsymbol{x}_0^*)V(\hat{\boldsymbol{\beta}})(\boldsymbol{X}^* - \boldsymbol{x}_0^*)']^{1/2}$$

Conditional

$$Q_p^C = (\boldsymbol{X}^* - \boldsymbol{x}_0^*)\hat{\boldsymbol{\beta}} + \phi^{-1}(p)\mathrm{diag}[(\boldsymbol{X}^* - \boldsymbol{x}_0^*)(V(\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\Psi}})(\boldsymbol{X}^* - \boldsymbol{x}_0^*)']^{1/2}$$

where

$\boldsymbol{X}^*$ indicates a matrix of user specified transformations

$\boldsymbol{x}_0^*$ indicates a matrix of reference values

Snapshot of the aggregated data

```
. use http://www.stats4life.se/data/md_10_studies, clear
. list id md dose semd n sd if id <= 3, sepby(id) noobs
```

id	md	dose	semd	n	sd
1	-.4142542	2.043162	.762961	334	9.785315
1	0	4.466444	0	333	9.918493
1	2.280955	8.789105	.8060276	333	10.8613
2	0	2.071578	0	334	10.41234
2	1.425228	4.396238	.787194	333	9.912573
2	3.382661	8.83116	.7879598	333	9.932815
3	-3.110585	2.120871	.7938005	334	10.62627
3	0	4.425666	0	333	9.861535
3	1.759263	8.376437	.7596498	333	9.742396

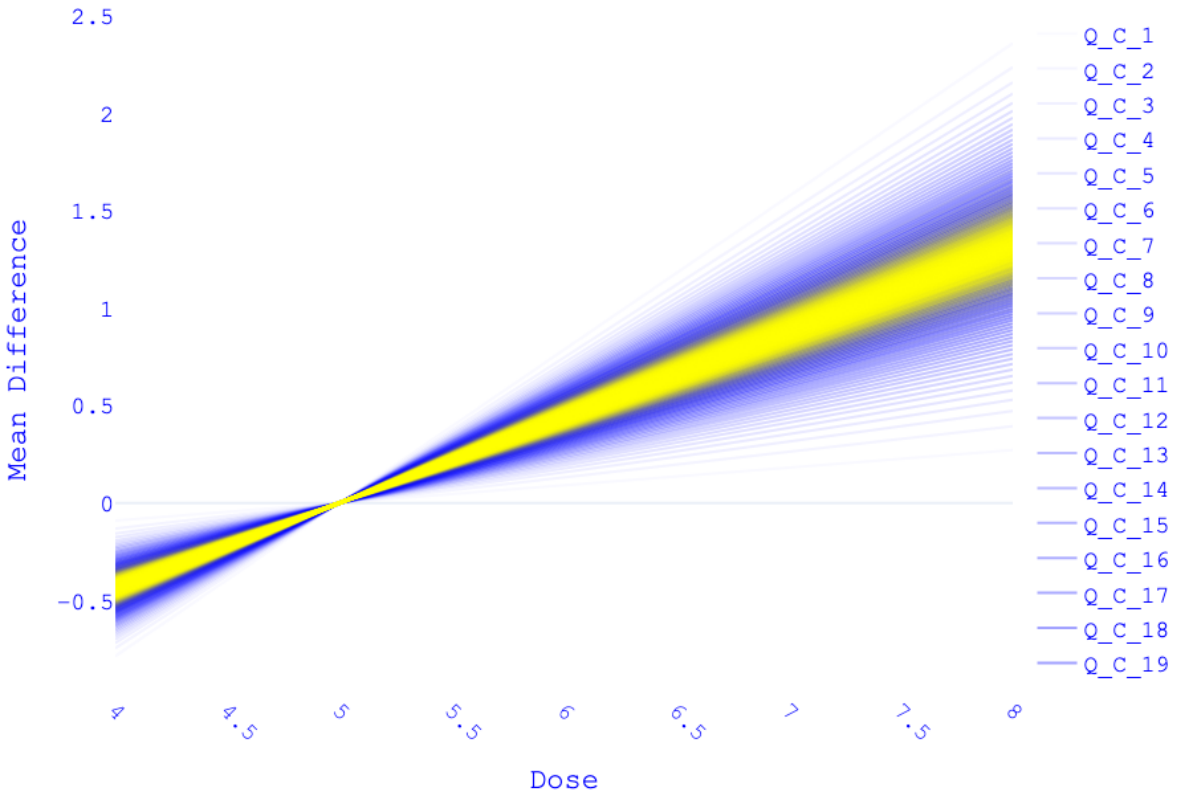
Obtain estimates of the weighted mixed model using a linear function

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

```
. 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
------------------------------------------------------------------------------
          md | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        dose |   .5375663   .0572842     9.38   0.000     .4252913    .6498414
------------------------------------------------------------------------------
---------------------------------------------
  Random-effects parameters  |   Estimate
-----------------------------+---------------
std(dose,dose)               |    .1384381
---------------------------------------------
LR test vs. no random-effects model = 4.9471603       Prob >= chi2(1) = 0.0261
```

Present marginal + conditional quantiles

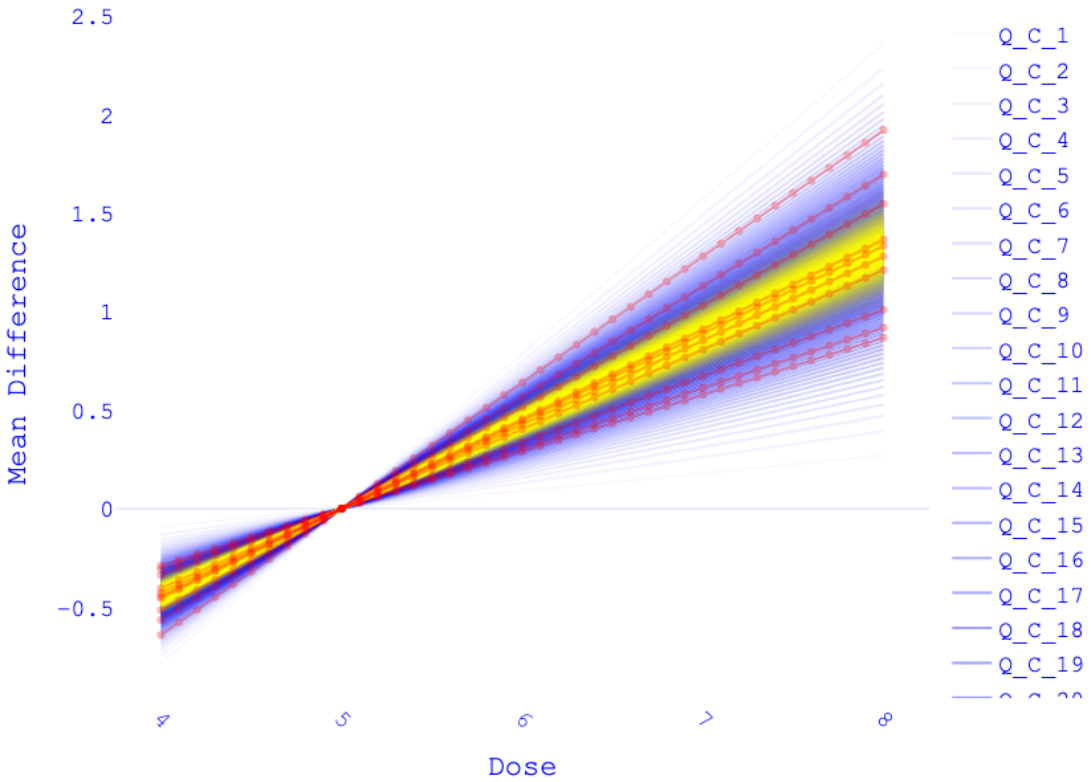
```
drmeta_het , dose(4(.1)8) ref(5) eq(d) iqc iqm
```



Click Here

Present marginal + conditional quantiles + BLUPs

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



Click Here

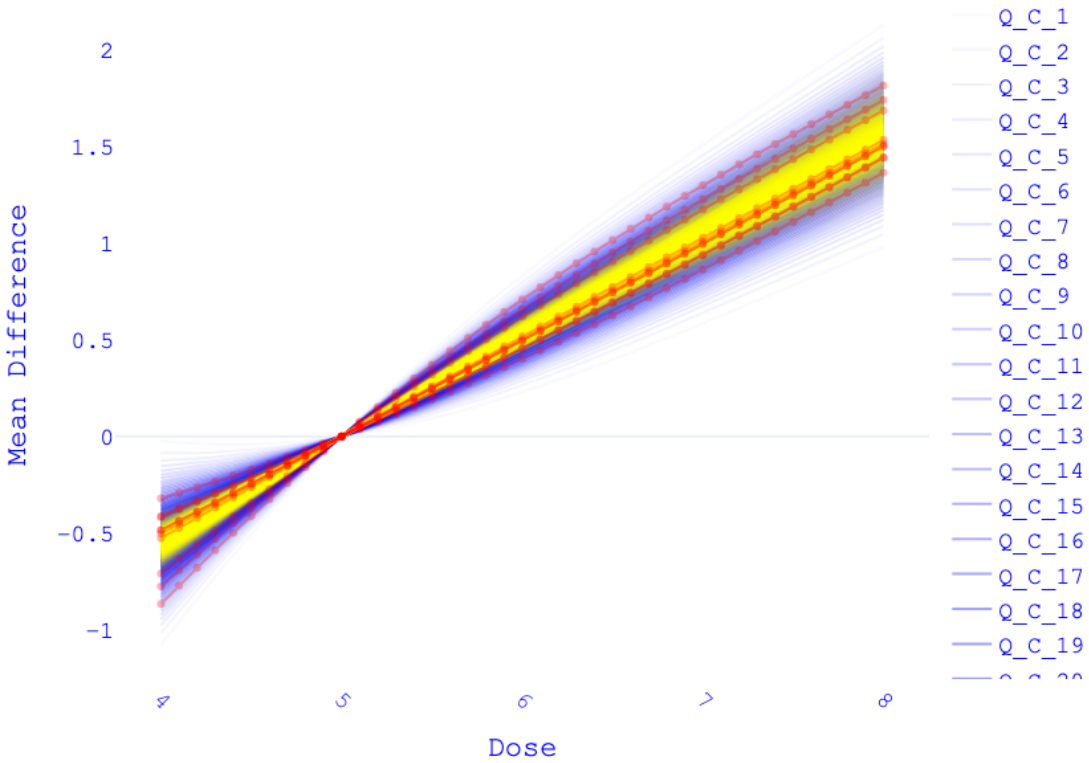
Obtain estimates of the weighted mixed model using a restricted cubic spline function

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

```
. mkspline doses = dose, nk(3) cubic displayknots

             |     knot1      knot2      knot3
-------------+------------------------------
        dose |  1.994129   4.410952    8.83116

. 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                                  Number of obs =        26
          AIC = 82.47                                Model chi2(2) =    101.94
Log likelihood = -36.233006                            Prob > chi2 =    0.0000
------------------------------------------------------------------------------
          md | 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  |   Estimate
-----------------------------+------------
std(doses1,doses1)           |    .3478605
std(doses2,doses2)           |    .3439548
corr(doses1,doses1)          |         -1
```

Present marginal + conditional quantiles + BLUPs

```
drmeta_het ,  matk(knots) dose(4(.1)8) ref(5) iqm iqc iqcb
```



Click Here

Simulated Example: Walking and mortality

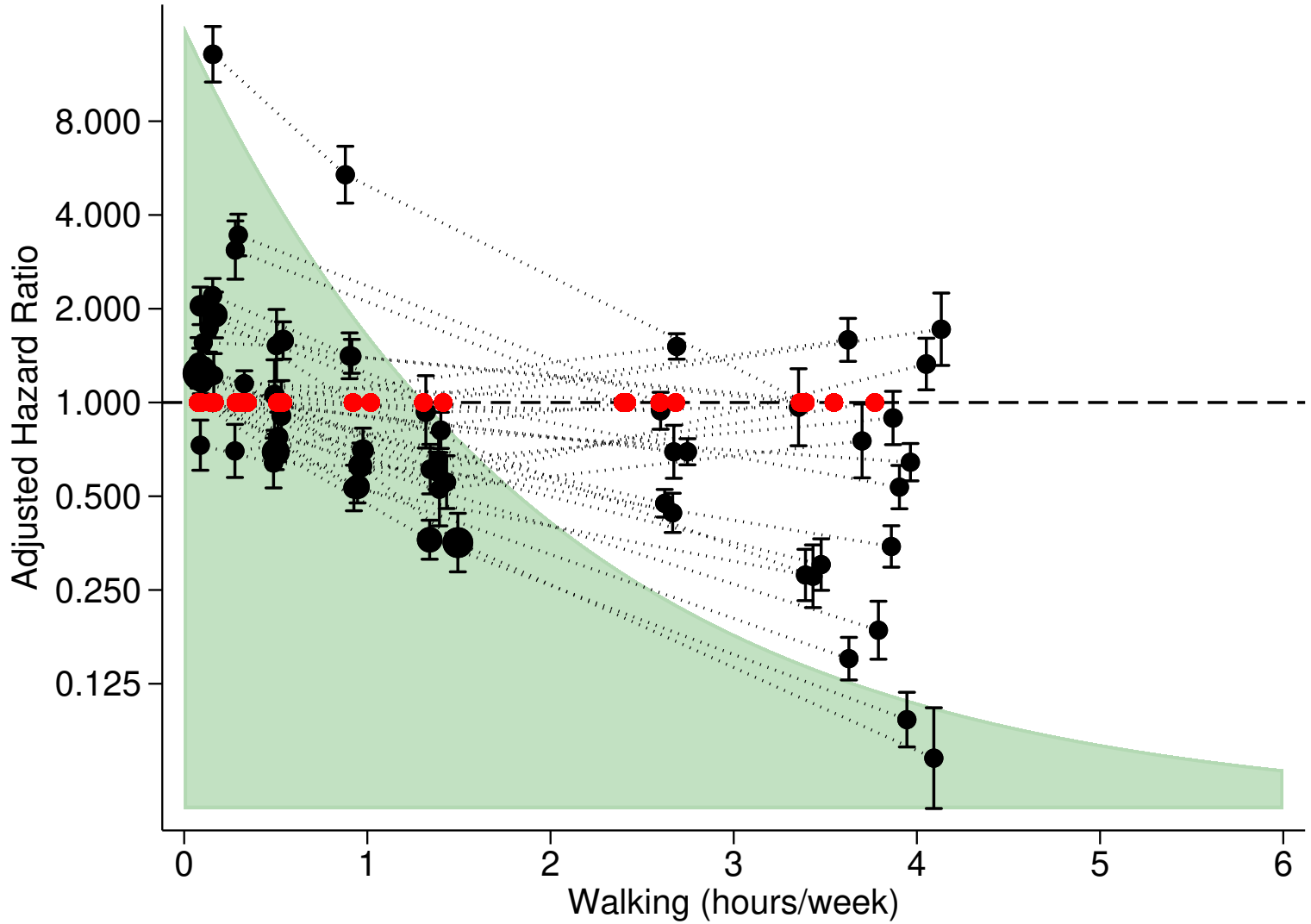
- Consider 30 prospective cohort studies investigating the association between baseline walking, measured in hours/week, and time until death, or end of follow-up (10 years), whichever came first (*Stata J*, 2021).
- Age is inversely associated with walking levels and positively associated with higher 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

$$HR = e^{-0.5(x-2)+0.5(x>2)(x-2)}$$

Snapshot of the aggregated data

```
. use http://www.stats4life.se/data/hr_drm, clear
. list id walk b seb case py if inlist(id, 1, 20, 23)
  +-----------------------------------------+
  | id   walk       b    seb   case      py |
  |-----------------------------------------|
  |  1    0.3    1.13   0.11    229     777 |
  |  1    2.4    0.00   0.00    137    1704 |
  |-----------------------------------------|
  | 20    0.1    0.21   0.10    239     674 |
  | 20    0.5    0.00   0.00    216     946 |
  | 20    1.5   -1.04   0.11    133    1773 |
  | 20    4.1   -2.63   0.19     32    2318 |
  |-----------------------------------------|
  | 23    0.2    0.65   0.09    311     973 |
  | 23    0.9    0.00   0.00    247    1765 |
  | 23    3.4   -1.28   0.12    101    2752 |
  +-----------------------------------------+
```

Plotting the empirical contrasts



Piecewise linear weighted mixed-effects model

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     Number of studies =        30
Optimization   = ml                                  Number of obs =        61
           AIC = 37.55                               Model chi2(2) =    110.27
Log likelihood = -13.773298                            Prob > chi2 =    0.0000
```

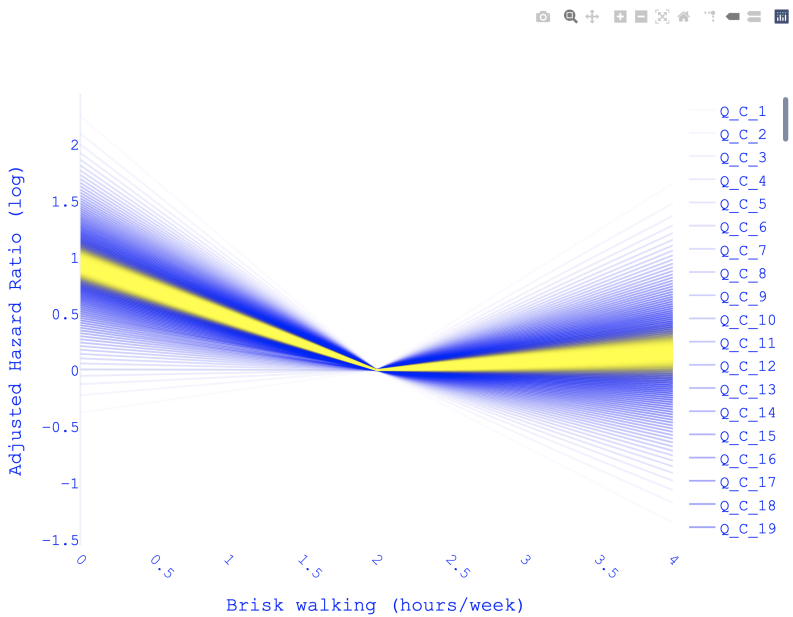
b	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
walk	-.4678671	.0536744	-8.72	0.000	-.573067	-.3626673
walkplus	.5432787	.0626324	8.67	0.000	.4205213	.666036

Random-effects parameters	Estimate
var(walk,walk)	.0766958
var(walkplus,walkplus)	.0507463
cov(walk,walkplus)	-.0136841

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
```



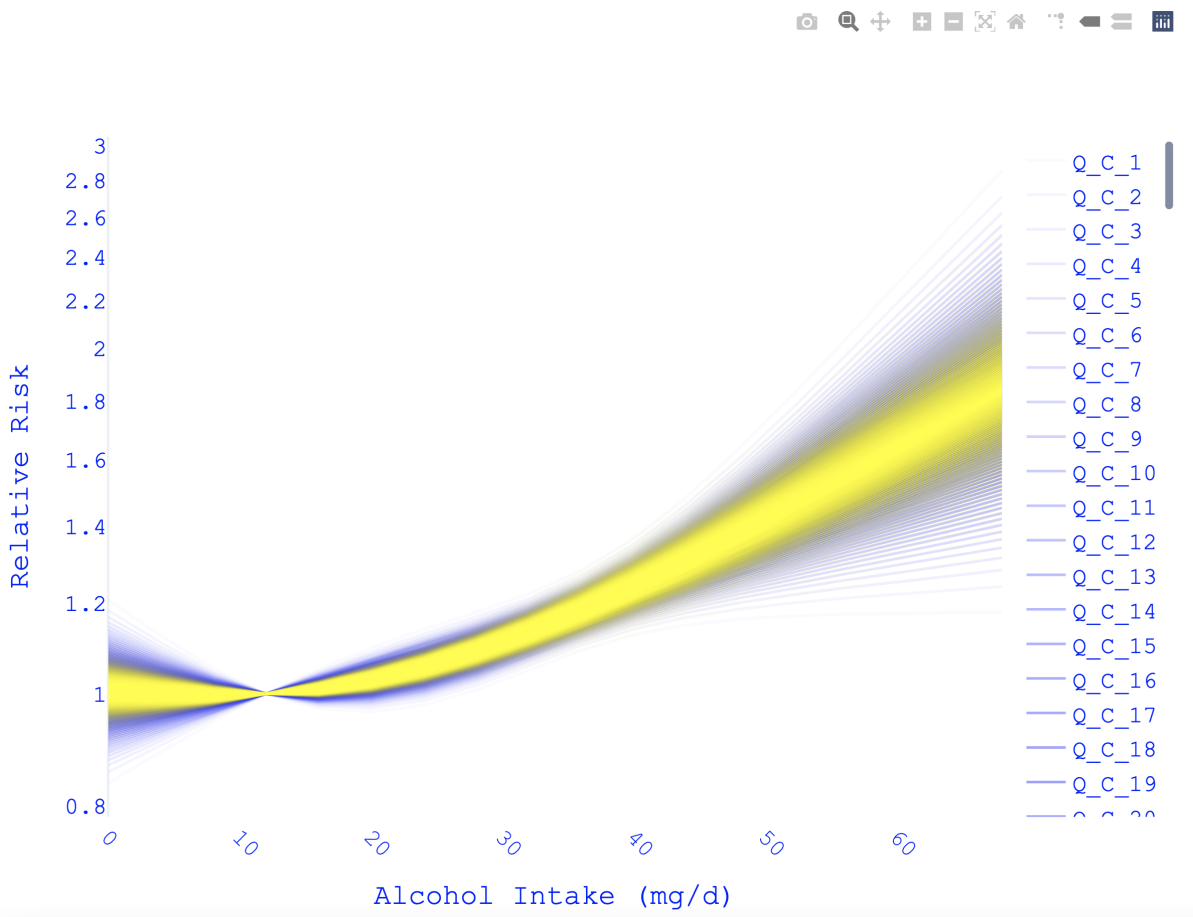
Click Here

Alcohol intake and colorectal cancer risk (cubic splines)

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

Age and and breast cancer mortality

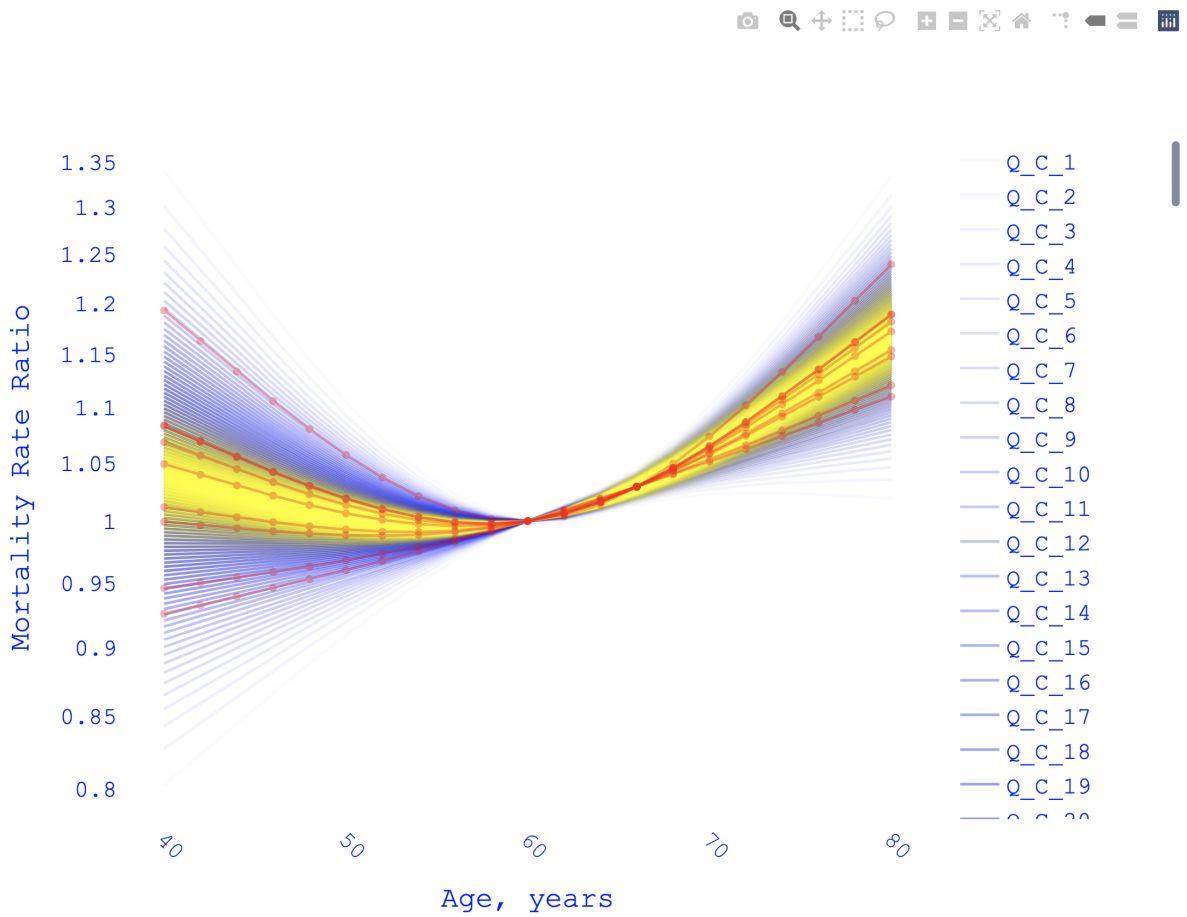
We use data from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. The SEER program provides data about cancer statistics from several population-based registries in the USA (http://seer.cancer.gov) from San Francisco- Oakland, Connecticut, Metropolitan 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*.

Age and and breast cancer mortality

```
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 and breast cancer mortality

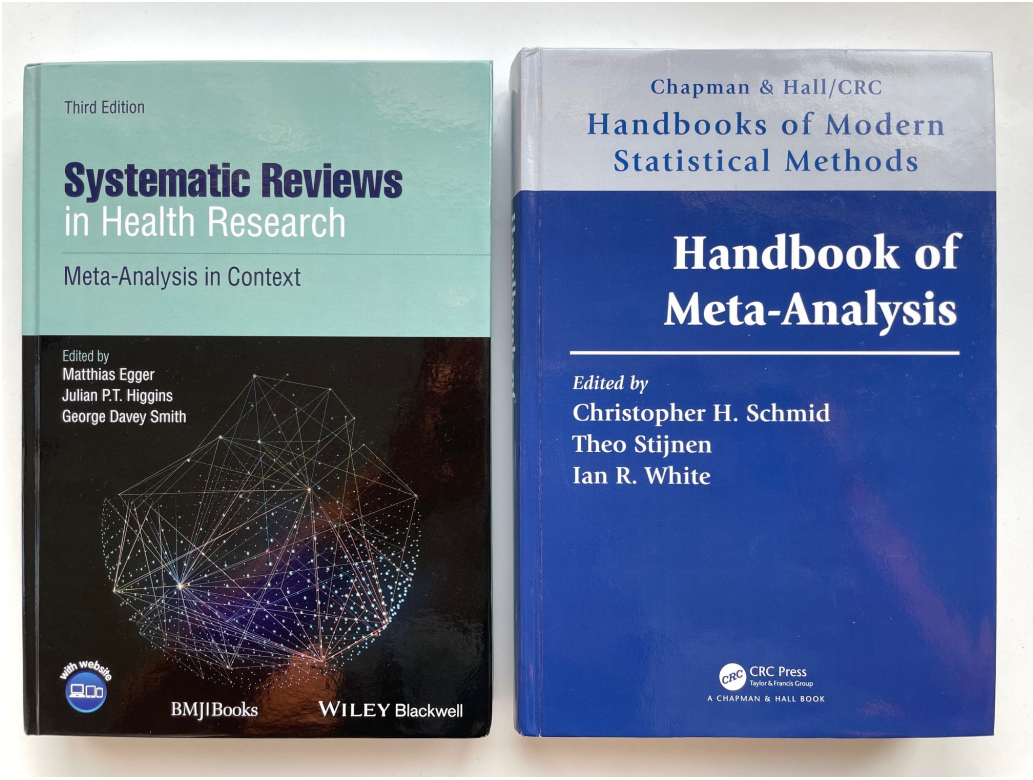


Click Here

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



- **Orsini N**, Larsson, SC, Salanti, G (2022). Dose–Response Meta-Analysis. Chapter 14. *Systematic Reviews in Health Research: Meta-Analysis in Context*, 258-269. John Wiley Sons Ltd.
- **Orsini N**, and Spiegelman D. (2020) Meta-Analysis of Dose-Response Relationships. Chapter 18. *Handbook of Meta-Analysis*, 395-428. Chapman and Hall/CRC.

Selected references

- **Orsini N**. Weighted mixed-effects dose-response models for tables of correlated contrasts. *Stata Journal*. 2021 (2), 320-347.
- Sera, F., Armstrong, B., Blangiardo, M., Gasparrini, A. (2019). An extended mixed-effects framework for meta-analysis. *Statistics in medicine*, 38(29), 5429-5444.
- Crippa A, Discacciati A, Bottai M, Spiegelman D, **Orsini N**. One-stage dose-response meta-analysis for aggregated data. *Stat Methods Med Res*. 2019 May;28(5):1579-1596.
- Schweder, T., Hjort, N. L. (2016). *Confidence, likelihood, probability*. Cambridge University Press.