RESEARCH ARTICLE　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

Dose-response meta-analysis of differences in means

Alessio Crippa[1*] and Nicola Orsini[1]

Abstract

Background: Meta-analytical methods are frequently used to combine dose-response findings expressed in terms of relative risks. However, no methodology has been established when results are summarized in terms of differences in means of quantitative outcomes.

Methods: We proposed a two-stage approach. A flexible dose-response model is estimated within each study (first stage) taking into account the covariance of the data points (mean differences, standardized mean differences). Parameters describing the study-specific curves are then combined using a multivariate random-effects model (second stage) to address heterogeneity across studies.

Results: The method is fairly general and can accommodate a variety of parametric functions. Compared to traditional non-linear models (e.g. E_{max}, logistic), spline models do not assume any pre-specified dose-response curve. Spline models allow inclusion of studies with a small number of dose levels, and almost any shape, even non monotonic ones, can be estimated using only two parameters. We illustrated the method using dose-response data arising from five clinical trials on an antipsychotic drug, aripiprazole, and improvement in symptoms in shizoaffective patients. Using the Positive and Negative Syndrome Scale (PANSS), pooled results indicated a non-linear association with the maximum change in mean PANSS score equal to 10.40 (95 % confidence interval 7.48, 13.30) observed for 19.32 mg/day of aripiprazole. No substantial change in PANSS score was observed above this value. An estimated dose of 10.43 mg/day was found to produce 80 % of the maximum predicted response.

Conclusion: The described approach should be adopted to combine correlated differences in means of quantitative outcomes arising from multiple studies. Sensitivity analysis can be a useful tool to assess the robustness of the overall dose-response curve to different modelling strategies. A user-friendly R package has been developed to facilitate applications by practitioners.

Keywords: Meta-analysis, Dose-response, Mean differences, Random-effects

Background

The identification and characterization of dose-response relationships is an essential part of the analysis in many scientific disciplines such as toxicology, pharmacology, and epidemiology. This is particularly important in the development and testing of new compounds (e.g. a new drug, pharmaceutical treatment) where trials at different stages aim to evaluate the efficacy of increasing levels of dosage (Phase II-III trials) or to derive a dose-response curve for selection of optimal doses (Phase IV trials) [1, 2].

Randomized clinical trials often investigate a continuous outcome variable, such as the efficacy or safety of a drug, reporting the change from baseline of a medical score, or the final value of a clinical measurement. The dose-response results are typically summarized by dose-specific means and standard deviations [3]. Measures of effect are expressed in terms of mean or standardized mean differences using a dose level, usually the placebo group, as referent [1]. Over the last few years methodological research focused on developing and improving methods for performing dose-response analysis in a single study [4, 5]. A conclusive result is hardly obtained by a single investigation and there is often the need to synthesize information collected from multiple studies. In such a

*Correspondence: alessio.crippa@ki.se
[1] Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden

case meta-analytic methods can be used to define an overall relation or to investigate heterogeneity across study findings.

A method for pooling aggregated dose-response data where the outcome is a log relative risk was originally presented by Greenland and Longnecker in 1992 [6]. Since then, several papers have refined and covered specific aspects of the methodology such as model specification [7, 8], modeling strategies [9, 10], and software implementation [11, 12]. Other methodological articles extended the approach for continuous outcome but in the case where individual patient data are available, mainly in the context of time-series environmental studies [13–15].

Only a few alternatives have been proposed to pool aggregated dose-response data where the findings are summarized by differences in means. Davis and Chen [16] in 2004 described a methodology for summarizing dose-response curves of first and second generation antipsychotics in schizoaffective patients. The authors reconstructed drug-specific dose-response curves and conducted a meta-analysis to compare the effectiveness of medium vs high dosages. A common alternative to analyze the drug effect consists of fitting a random-intercept E_{max} model, where the random component accounts for heterogeneity in placebo effect across trials [17]. Heterogeneity, however, may be related to other study characteristics rather than differences in placebo response such as implementation, participants, intervention, and outcome definition. Thomas et al. [18] adopted hierarchical Bayesian models to summarize and describe, independently, the distribution of study-specific model parameters derived from an E_{max} model.

The mentioned strategies assumed pre-specified models that do not allow for non-monotonic curves which may occur in practice [19], as in case of dose-response data of antipsychotics. In addition, fitting study-specific sigmoidal curves such as the E_{max} model requires that the single studies have assessed at least three dose levels in order to estimate model parameters. Discarding studies not providing enough data points represents a loss of information and may introduce bias.

The aim of this paper is to formalize and propose a general and flexible methodology to pool dose-response relations from aggregated data where the changes in the distribution of the quantitative outcome are expressed in terms of differences in means. We first present the data necessary for a dose-response meta-analysis and derive formulas for obtaining effect sizes and their variance/covariance structure. We describe flexible dose-response models with particular emphasis on regression splines. The method is then applied to dose-response data from clinical trials on use of aripiprazole and symptoms improvement in schizoaffective patients.

Methods

Dose-response data

The notation and data required for a dose-response meta-analysis for a generic study are displayed in Table 1. We consider I studies indexed by $i = 1, \dots, I$ reporting the results of a common treatment at different dose levels $x_{ij}, j = 1, \dots, J_i$, where $x_{0i} = 0$ indicates the control or placebo group in the i-th study. The study-specific results typically consist of dose-specific means of an outcome variable, Y_{ij}, that measures the efficacy of the j-th dose in the i-th study [3]. Additional information about the number of patients allocated in each treatment, n_{ij}, and the sample standard deviations of Y_{ij}, sd_{ij}, is generally reported or obtained from the study-specific results.

Effect sizes and their variance/covariance

A common way to reduce heterogeneity in placebo response is to compute the effect size (or treatment effect) as difference between dose-specific means and placebo mean. In case all studies measure the outcome on a common and interpretable scale, the difference can be based on the absolute scale

$$d_{ij} = \bar{Y}_{ij} - \bar{Y}_{i0}, \quad j = 1, \dots, J_i, \ i = 1, \dots, I \quad (1)$$

Assuming common study-specific population standard deviations, the variance of d_{ij} is defined as

$$\text{Var}\left(d_{ij}\right) = \frac{n_{ij} + n_{i0}}{n_{ij} n_{i0}} s_{p_i}^2, \quad j = 1, \dots, J_i, \ i = 1, \dots, I \quad (2)$$

where $s_{p_i}^2 = \sum_{j=0}^{J_i} \left(n_{ij} - 1\right) sd_{ij}^2 / \sum_{j=0}^{J_i} \left(n_{ij} - 1\right)$ is the square of the pooled standard deviation for the i-th study. Since the study-specific mean differences d_{ij} use the same referent values, $\bar{Y}_{i0}$, they cannot be regarded as independent. The covariance term is defined as

$$\text{Cov}\left(d_{ij}, d_{ij'}\right) = \text{Var}\left(\bar{Y}_{j0}\right) = \frac{s_{i0}^2}{n_{i0}}, \quad j \neq j', \ i = 1, \dots, I \quad (3)$$

Table 1 Notation for aggregated data in the i-th study used in dose-response meta-analysis of differences in meas

dose	mean(Y)[a]	sd(Y)	n[b]	d[c]	Var (d)	d*[d]	Var (d*)
0	$\bar{Y}_{i0}$	sd_{i0}	n_{i0}	0	–	0	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_{ij}	$\bar{Y}_{ij}$	sd_{ij}	n_{ij}	d_{ij}	Var $\left(d_{ij}\right)$	d_{ij}^*	Var $\left(d_{ij}^*\right)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_{iJ_i}	$\bar{Y}_{iJ_i}$	sd_{iJ_i}	n_{iJ_i}	d_{iJ_i}	Var $\left(d_{iJ_i}\right)$	$d_{iJ_i}^*$	Var $\left(d_{iJ_i}^*\right)$

[a] Y is the continuous outcome
[b] Number of patients
[c] Mean difference
[d] Standardized mean difference

In case the outcome is measured on different scales the effect sizes can be based on standardized mean differences

$$d_{ij}^* = \frac{\bar{Y}_{ij} - \bar{Y}_{i0}}{s_{p_i}}, \quad j = 1, \ldots, J_i, \ i = 1, \ldots, I \tag{4}$$

with

$$\text{Var}\left(d_{ij}^*\right) = \frac{1}{n_{ij}} + \frac{1}{n_{i0}} + \frac{d_{ij}^{*2}}{2\sum_{j=0}^{J_i} n_{ij}}, \quad j = 1, \ldots, J_i,$$

$$i = 1, \ldots, I$$

$$\text{Cov}\left(d_{ij}^*, d_{ij'}^*\right) = \frac{1}{n_{i0}} + \frac{d_{ij}^* d_{ij'}^*}{2\sum_{j=0}^{J_i} n_{ij}}, \quad j \neq j', \ i = 1, \ldots, I \tag{5}$$

Dose-response analysis

The chosen effect sizes and the corresponding (co)variances are used to estimate the study-specific dose-response curves. The dose-response curves characterize the relative efficacy of the dose under investigation using the placebo effect as referent (i.e. the relative efficacy for the placebo is zero by definition). The dose-response models are expressed through the parametric model f, which specifies how the effect size varies according to the dose values. The functional relationship f is parametrized in terms of $\boldsymbol{\theta}_i$, the $p \times 1$ vector of dose-response coefficients. We consider the case of mean differences, d_{ij}, but the same principles apply for standardized mean differences, d_{ij}^*. The study-specific curves can be written as

$$\boldsymbol{d}_i = f\left(\boldsymbol{x}_i, \boldsymbol{\theta}_i\right) + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N\left(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_i\right), \quad i = 1, \ldots, I \tag{6}$$

$\hat{\boldsymbol{\Sigma}}_i$ is the covariance matrix of the residual error term, with $\text{Var}\left(d_{ij}\right)$ along the diagonal and $\text{Cov}\left(d_{ij}, d_{ij'}\right)$ off-diagonal.

Several alternatives are available to model the dose-response curve (i.e. for the choice of f). Table 2 shows the definition of 4 types of models ordered according to the number of parameters, ranging from 1 for a linear model up to 3 for a logistic model. See Bretz et al. for a comprehensive description [2].

The most common choice in dose findings [2] is the use of the $E_{\max}$ model which is expressed in terms of three parameters: the maximum effect (θ_{1i}), the dose to produce half of the maximum effect (θ_{2i}) and the steepness of the curve (θ_{3i}) [20]. As other non-linear models, the $E_{\max}$ model assumes a specific shape that does not allow for non-monotonic curve and its estimation requires at least three non reference dose levels. Quadratic models are defined by only $p = 2$ coefficients but may poorly fit at extreme dose values [9]. Other non-linear models such as logistic and sigmoidal models, are commonly defined by $p \geq 3$ coefficients so that study-specific aggregated data may not be sufficient to estimate the parameters.

We propose the use of regression splines to flexibly model the dose of interest. Splines represent a family of smooth functions that can describe a wide range of curves (i.e. U-shaped, J-shaped, S-shaped, threshold) [21]. The curves consist of piecewise polynomials over consecutive intervals defined by $\boldsymbol{k}$ knots. Their use may facilitate curve fitting since many non-linear curves can be examined by estimating only a small number of coefficients. For instance, a restricted cubic spline model with three knots $\boldsymbol{k} = (k_1, k_2, k_3)$ is defined only in terms of $p = 2$ coefficients [22]

$$\text{E}\left[\boldsymbol{d}_i | \boldsymbol{x}_i\right] = \theta_{1i} \boldsymbol{x}_{1i} + \theta_{2i} \boldsymbol{x}_{2i} \tag{7}$$

with two transformations [23] defined as

$$x_1 = x$$

$$x_2 = \frac{(x - k_1)_+^3 - \frac{k_3 - k_1}{k_3 - k_2}(x - k_2)_+^3 + \frac{k_2 - k_1}{k_3 - k_2}(x - k_3)_+^3}{(k_3 - k_1)^2} \tag{8}$$

where the '+' notation, with $u_+ = u$ if $u \geq 0$ and $u_+ = 0$ otherwise, has been used.

An alternative flexible approach to model the dose-response association is represented by fractional polynomials. In particular, a dose-response model based on fractional polynomial of order two can be written as in Eq. 6 with the two transformations defined as

$$x_1 = x^{p_1} \text{ and } x_2 = x^{p_2} \text{ if } p_1 \neq p_2$$

$$x_1 = x^{p_1} \text{ and } x_2 = x^{p_1} \log(x) \text{ if } p_1 = p_2 \tag{9}$$

for each combination of p_1 and p_2 in the predefined set of values $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$; for $p = 0$, x^p becomes $\log(x)$. The best fitting fractional polynomial is typically chosen based on the Akaike's Information Criterion [24].

Once the functional relation f has been selected, generalized least square estimation can be performed to efficiently estimates the dose-response coefficients $\hat{\boldsymbol{\theta}}_j$ and the corresponding (co)variance matrix $\hat{\boldsymbol{V}}_j$, by minimizing

$$\left(\boldsymbol{d}_i - f\left(\boldsymbol{x}_i, \boldsymbol{\theta}_i\right)\right)^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left(\boldsymbol{d}_i - f\left(\boldsymbol{x}_i, \boldsymbol{\theta}_i\right)\right) \tag{10}$$

that generally requires numerical optimization algorithms. If f is a linear combination of the parameters

Table 2 Frequently used dose-response models

Model	Equation	No. of parameters
Linear	$\text{E}\left[\boldsymbol{d}_i | \boldsymbol{x}_i\right] = \theta_{1i} \boldsymbol{x}_i$	1
Quadratic	$\text{E}\left[\boldsymbol{d}_i | \boldsymbol{x}_i\right] = \theta_{1i} \boldsymbol{x}_i + \theta_{2i} \boldsymbol{x}_i^2$	2
$E_{\max}$	$\text{E}\left[\boldsymbol{d}_i | \boldsymbol{x}_i\right] = \theta_{1i} \boldsymbol{x}_i^{\theta_{3i}} / \left(\theta_{2i} + \boldsymbol{x}_i^{\theta_{3i}}\right)$	3
Logistic	$\text{E}\left[\boldsymbol{d}_i | \boldsymbol{x}_i\right] = \theta_{1i} / \{1 + \exp\left[(\theta_{2i} - \boldsymbol{x}_i)\right] / \theta_{3i}\}$	3

$\boldsymbol{\theta}_i$, as in the case of regression splines and fractional polynomials, the close solution can be written as

$$\hat{\boldsymbol{\theta}}_i = \left(X_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} X_i\right)^{-1} X_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{d}_i$$

$$\hat{\boldsymbol{V}}_i = \text{Var}\left(\hat{\boldsymbol{\theta}}_i\right) = \left(X_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} X_i\right)^{-1} \tag{11}$$

where X_i indicates the $J_i \times p$ design matrix in the i-th study.

Meta-analysis

The estimated study-specific dose-response coefficients $\hat{\boldsymbol{\theta}}_i$ and the accompanying (co)variance matrices $\hat{\boldsymbol{V}}_i$ are combined by means of multivariate meta-analysis

$$\hat{\boldsymbol{\theta}}_i \sim N\left(\boldsymbol{\theta}, \hat{\boldsymbol{V}}_i + \boldsymbol{\Psi}\right) \tag{12}$$

A fixed-effects model assumes no statistical heterogeneity among study results, i.e. differences in the dose-response coefficients are only related to sampling error. The assumption of homogeneity may not hold in practice, unless it is known that the studies are performed in a similar way and are sampled from the same population [25]. The Cochran's Q test [26] is typically used to test statistical heterogeneity across studies ($H_0 : \boldsymbol{\Psi} = \boldsymbol{0}$) [27]. Selected studies, however, will typically differ with respect to study design and implementation, selection of participants, and type of analyses. A certain degree of heterogeneity is expected and should be taken into account in the analysis. A random-effects model allows the dose-response coefficients, $\boldsymbol{\theta}_i$, to vary across studies. Statistical heterogeneity is captured by the between-studies variance $\boldsymbol{\Psi}$ while $\boldsymbol{\theta}$ represents the mean of the distribution of dose-response coefficients and an estimate, $\hat{\boldsymbol{\theta}}$, can be obtained using (restricted) maximum likelihood estimation [15].

As a final result, the pooled dose-response curve can be presented in either a graphical or tabular form by predicting the mean differences of the outcome for a set of x dose values

$$\text{E}[\,\hat{\boldsymbol{d}}|\boldsymbol{x}] = f\left(X, \hat{\boldsymbol{\theta}}\right) \tag{13}$$

with an approximate $(1 - \alpha/2)\%$ confidence interval (CI), that in case f is a linear combination of $\boldsymbol{\theta}$ can be expressed as

$$\text{E}[\,\hat{\boldsymbol{d}}|\boldsymbol{x}] \mp z_{\frac{\alpha}{2}} \sqrt{\text{diag}\left(f\left(X, \hat{\boldsymbol{\theta}}\right)^T \text{Cov}\left(\hat{\boldsymbol{\theta}}\right) f\left(X, \hat{\boldsymbol{\theta}}\right)\right)} \tag{14}$$

where $z_{\alpha/2}$ is the $\alpha/2$-th quantile of a standard normal distribution.

Dose findings

Once the pooled dose-response curve has been estimated, it may be of interest to determine a set of target doses, i.e. doses associated with prespecified outcome effects. In development of new compounds it is often important to select an optimal dose which is almost as effective as the maximum effective dose but has less undesired side effects, which often occur at high dosages. Suppose one wants to determine which is the lowest dose (ED_γ) to produce an almost complete effect, e.g. $\gamma\%$ of the observed maximum predicted response.

The ED_γ can be determined as

$$\widehat{\text{ED}}_\gamma = \underset{x \in (0, x_{\max})}{\arg\max} \left\{ \frac{\text{E}[\,\hat{d}|x]}{\text{E}[\,\hat{d}|x_{\max}]} \geq \gamma \right\} \tag{15}$$

where $x_{\max}$ is the dose corresponding to the maximum predicted outcome.

An important step when presenting results from dose findings analysis is to accompany the previous estimates with a measure of precision, typically confidence intervals. Pinheiro et al. [3] proposed the use a parametric bootstrap approach based on the asymptotic normal distribution of $\hat{\boldsymbol{\theta}}$, the pooled estimate of the dose-response coefficients. The approach consists in re-sampling the dose-response coefficients $\boldsymbol{\theta}$ from its approximate normal distribution and derive the distribution of $\widehat{\text{ED}}_\gamma$ based on the samples. Approximated confidence intervals for $\widehat{\text{ED}}_\gamma$ can be constructed using percentiles of the sampling distribution.

Results

To illustrate the methodology we examined the dose-response relation between aripiprazole, a second-generation antipsychotic, and symptoms improvement in schizoaffective patients. We updated the search strategy presented in a previous review by Davis and Chen [16] by searching the Medline, International Pharmaceutical Abstracts, CINAHL, and the Cochrane Database of Systematic Reviews. To reduce the exclusion of unpublished papers, additional sources including Food and Drug administration website, data from Cochrane reviews, poster presentations and conference abstracts were also searched. All random-assignment, double-blind, controlled clinical trials of schizoaffective patients providing dose-response results for at least two non-zero dosages of aripiprazole were eligible.

Five studies [28–32] met the inclusion criteria and were included in the analysis. All the studies reported mean changes from baseline as main outcome variable, using the Positive and Negative Syndrome Scale (PANSS). The PANSS scale is an ordinal score derived from 30 items ranging from 1 to 7. Computations of ratios such as percentage changes are not directly applicable and may lead to erroneous results [33, 34]. To address this issue, the theoretical minimum (i.e. 30) needs to be subtracted from the original score [35]. Information about the means, the number of patients assigned to each treatment, and the standard deviation was available from the published data. Because all the studies measured the outcome variable on

the same scale, we computed PANSS mean differences as effect sizes. Data are reported in Table 3.

We used the trial by Cutler et al. [28] to illustrate the steps required for estimating the dose-response curve for a single study. For example, the difference in mean PANSS comparing the dose of 2 mg/day relative to 0 mg/day is $d_{11} = 8.23 - 5.3 = 2.93$ mg/day. Its variance is $\mathrm{Var}(d_{11}) = (85 + 92)/(85 \times 92) \times s_{p_1}^2 = 7.59$, where $s_{p_1}^2 = 119,419/356 = 335.4$. The covariance of this difference in means PANSS is $18.31^2/85 = 3.94$. The variance/covariance structure associated with the vector of differences in means for this trial $\boldsymbol{d}_1$ can be presented in a matrix form

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 7.59 \\ 3.94 & 7.72 \\ 3.94 & 3.94 & 7.52 \end{bmatrix}$$

To estimate the dose-response curve we need first to specify the model f. We characterized the dose-response relation using a restricted cubic spline model with three knots located at the 10th, 50th, and 90th percentiles (0, 10, and 30 mg/day) of the overall dose distribution ($p = 2$). The restricted cubic spline dose-response model is defined as in Eq. 7. Efficient estimates of the dose-response coefficients and (co)variance matrix were obtained by generalized least square estimation

$$\hat{\boldsymbol{\theta}}_1 = (1.215, -5.738)^T$$

$$\hat{\boldsymbol{V}}_1 = \mathrm{Var}\left(\hat{\boldsymbol{\theta}}_1\right) = \begin{bmatrix} 0.49 \\ -3.65 & 31.64 \end{bmatrix} \quad (16)$$

We applied the same procedure to the other studies included in the meta-analysis in order to obtain the study-specific $\hat{\boldsymbol{\theta}}_i$ and $\hat{\boldsymbol{V}}_i, i = 1, \ldots, 5$ (Table 4). The study-specific predicted curves are presented in Fig. 1. Under a random-effects model, restricted maximum likelihood estimates were

$$\hat{\boldsymbol{\theta}} = (0.937, -1.156)^T$$

$$\widehat{\mathrm{Cov}}\left(\hat{\boldsymbol{\theta}}\right) = \begin{bmatrix} 0.03 \\ -0.05 & 0.10 \end{bmatrix} \quad (17)$$

A p-value < 0.001 for the multivariate Wald-type test $H_0 : \boldsymbol{\theta} = \boldsymbol{0}$ provided strong evidence against the null hypothesis of no relation between different doses of aripiprazole and mean change PANSS score. The Q test ($Q = 3.5$, p-value = 0.899) did not detect substantial statistical heterogeneity across studies.

To communicate results of the pooled dose-response analysis, we can estimate the pooled mean differences in PANSS scores using 0 mg/day as referent as $0.937x_1 - 1.156x_2$, together with the corresponding 95 % confidence interval for a generic dose x of interest as following

$$(0.937x_1 - 1.156x_2) \mp 1.96\sqrt{0.03x_1^2 + 0.1x_2^2 - 0.1x_1x_2}$$

Table 3 Aggregated dose-response data of five clinical trials investigating effectiveness of different dosages of aripiprazole in schizoaffective patients. The continuous outcome is measured on the Positive and Negative Syndrome Scale and summarized by mean values (mean(Y)) and standard deviations (sd(Y))

ID	Author, Year	dose	mean(Y)	sd(Y)	n	d	Var(d)
1	Cutler, 2006 [28]	0	5.300	18.310	85	0.000	0.000
		2	8.230	18.320	92	2.930	7.593
		5	10.600	18.310	89	5.300	7.715
		10	11.300	18.320	94	6.000	7.515
2	McEvoy, 2007 [29]	0	2.330	26.100	107	0.000	0.000
		10	15.040	27.600	103	12.710	13.344
		15	11.730	26.200	103	9.400	13.344
		20	14.440	25.900	97	12.110	13.764
3	Kane, 2002 [30]	0	2.900	24.280	102	0.000	0.000
		15	15.500	26.490	99	12.600	12.038
		30	11.400	22.900	100	8.500	11.977
4	Potkin, 2003 [31]	0	5.000	21.140	103	0.000	0.000
		20	14.500	20.160	98	9.500	8.563
		30	13.900	20.880	96	8.900	8.654
5	Study 94202 [32]	0	1.400	25.730	57	0.000	0.000
		2	11.000	25.000	51	9.600	25.447
		10	11.500	25.200	51	10.100	25.447
		30	15.800	28.510	54	14.400	24.701

Table 4 Study-specific dose-response coefficients and corresponding covariances for different dose-response models considered in the analysis

Model	id	$\hat{\theta}_1$	$\hat{\theta}_2$	Var $\left(\hat{\theta}_1\right)$	Cov $\left(\hat{\theta}_1, \hat{\theta}_2\right)$	Var $\left(\hat{\theta}_2\right)$
Restricted cubic splines	1	0.55	0.55	0.065	0.065	0.065
	2	0.59	0.59	0.031	0.031	0.031
	3	0.84	−1.69	0.054	−0.12	0.38
	4	0.47	−0.54	0.021	−0.043	0.15
	5	0.78	−1.25	0.23	−0.62	1.8
Fractional Polynomials	1	17.47	−7.84	2.9e+02	−2.2e+02	1.8e+02
	2	29.59	−12.42	2.5e+02	−1.6e+02	1e+02
	3	32.12	−13.47	1.4e+02	−70	37
	4	18.48	−6.42	1.8e+02	−93	49
	5	21.97	−7.00	2.3e+02	−1.1e+02	55
Emax	1	8.13	3.13	27	24	36
	2	11.39	0.00	38	35	42
	3	10.54	0.00	37	53	1e+02
	4	9.20	0.00	60	1.4e+02	3.6e+02
	5	13.28	0.94	23	2.7	2.7
Quadratic	1	1.54	−0.09	0.96	−0.086	0.0083
	2	1.54	−0.05	0.35	−0.017	0.00089
	3	1.40	−0.04	0.18	−0.0055	0.00018
	4	0.83	−0.02	0.14	−0.0047	0.00017
	5	1.08	−0.02	0.51	−0.016	0.00051
Piecewise linear	1	0.55		0.065		
	2	0.59		0.031		
	3	0.84	−1.69	0.054	−0.12	0.38
	4	0.47	−0.54	0.021	−0.043	0.15
	5	0.78	−1.25	0.23	−0.62	1.8

where x_1 and x_2 are defined as in Eq. 8. For instance, the model-based predicted mean changes in PANSS score compared to placebo were 4.52 (95 % CI: 2.96, 6.08) for 5 mg/day, 8.08 (95 % CI: 5.43, 10.73) for 10 mg/day, 9.95 (95 % CI: 6.97, 12.94) for 15 mg/day, 10.38 (95 % CI: 7.49, 13.27) for 20 mg/day, 9.84 (95 % CI: 6.86, 12.83) for 25 mg/day, and 8.83 (95 % CI: 5.11, 12.54) for 30 mg/day. The pooled predicted dose-response curve together with the confidence intervals and the model mean differences is provided in Fig. 2.

The results indicated a statistically significant positive association between increasing doses of aripiprazole and the mean change in PANSS score with the maximum value of 10.39 (95 % CI: 7.48, 13.30) observed at $x_{\max}$ = 19.32 mg/day. The model suggested a slight decrease in the predicted mean PANSS score for dosages greater than 20 mg/day. The estimated dose to produce 50 % and 80 % of the predicted maximum effect were $\widehat{ED}_{50}$ = 5.82 mg/day (95 % CI: 5.10, 8.58) and $\widehat{ED}_{80}$ = 10.43 mg/day (95 % CI: 9.02, 16.73).
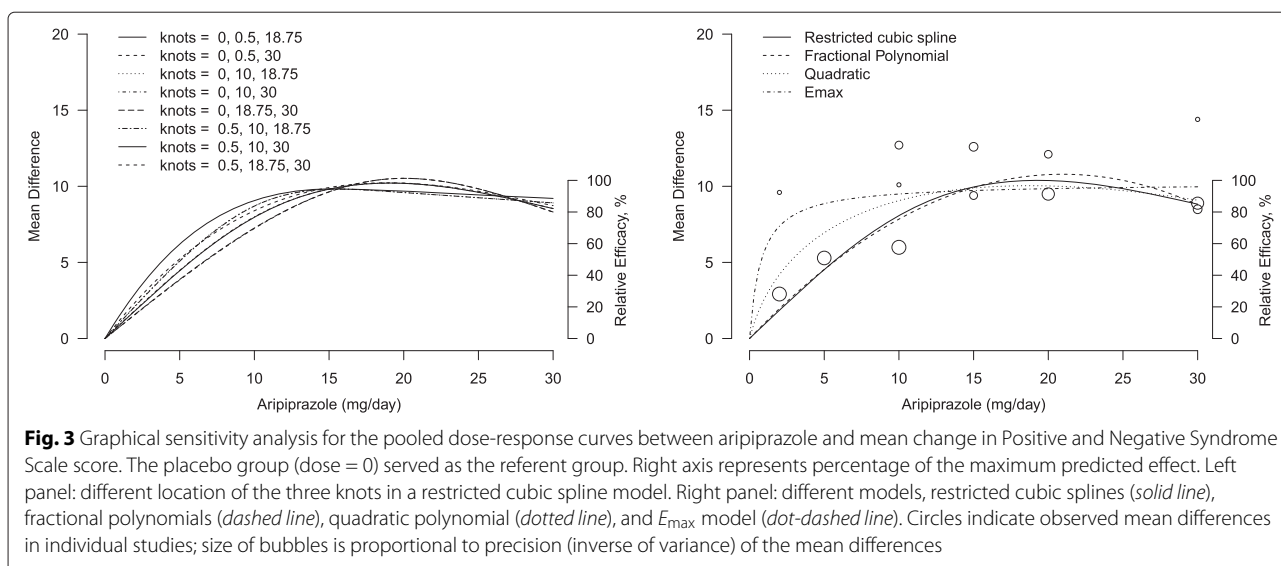
Sensitivity analysis

A sensitivity analysis is often required to evaluate the robustness of the pooled dose-response curve. In the spline model, for example, the location of the knots may affect the shape of the dose-response curve. Therefore we considered alternative knots locations including different combinations of the 10th, 25th, 50th, 75th and 90th percentiles of the overall dose distribution (0, 0.5, 10, 18.75, and 30 mg/day). A graphical comparison is presented in the left panel of Fig. 3. The alternative curves roughly described the same dose-response shape with no substantial variation, all indicating an increase in the mean change PANSS score up to 20 mg/day of aripiprazole. We can assess whether there is an increasing trend above 20 mg/day by simply re-defining x_2 equal to $(x - 20)_+$ in Eq. 8; this approach is known as piecewise linear model. The rate of change in the PANSS mean differences was negative and not statistically significant ($\theta_1 + \theta_2$=-0.284, p = 0.18) after 20 mg/day of aripiprazole.
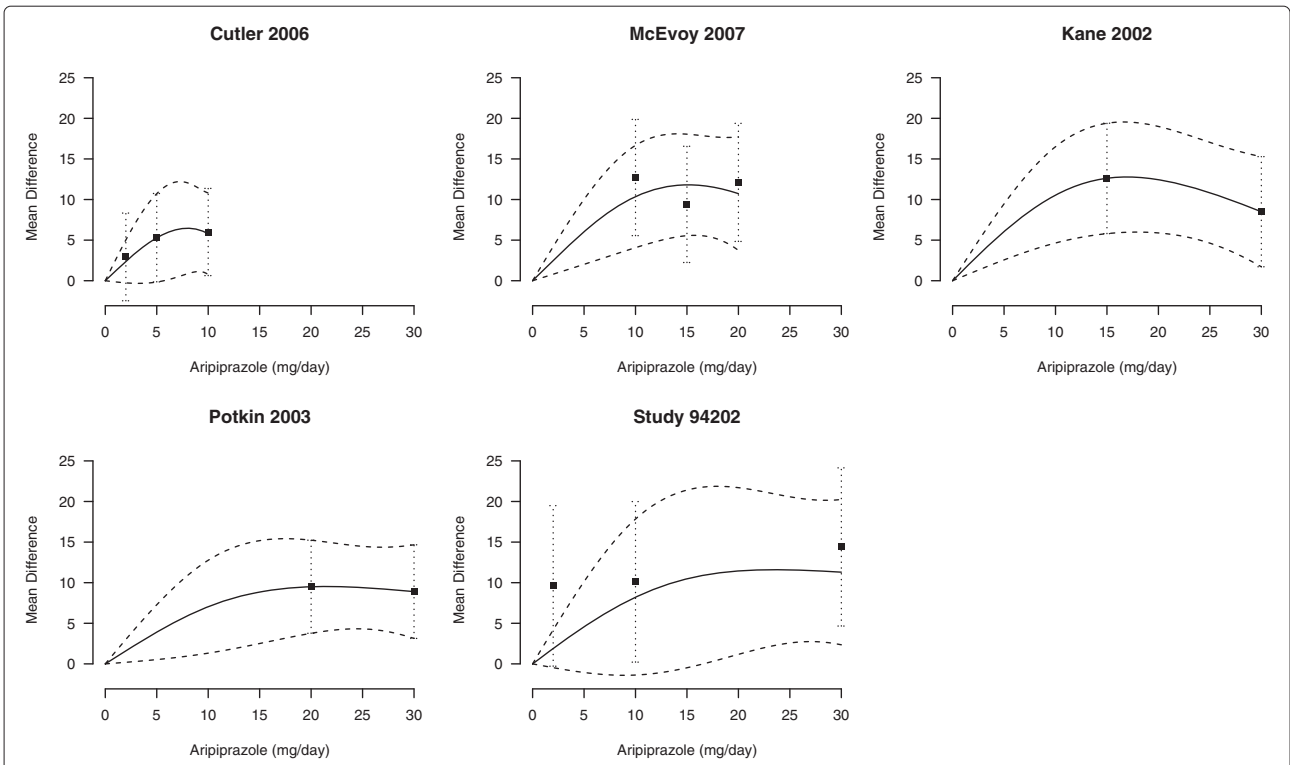
Fig. 1 Study-specific mean differences in Positive and Negative Syndrome Scale score for increasing dosages of aripiprazole. The first author and year of publication of the subjects included in the original analyses are reported. Black squares indicate the mean differences and whiskers their 95 % confidence interval. Aripiprazole dosage was modeled with restricted cubic splines. Solid lines represent the estimated dose-response curves, dashed lines the corresponding 95 % confidence intervals. The placebo group (dose = 0) served as the referent group
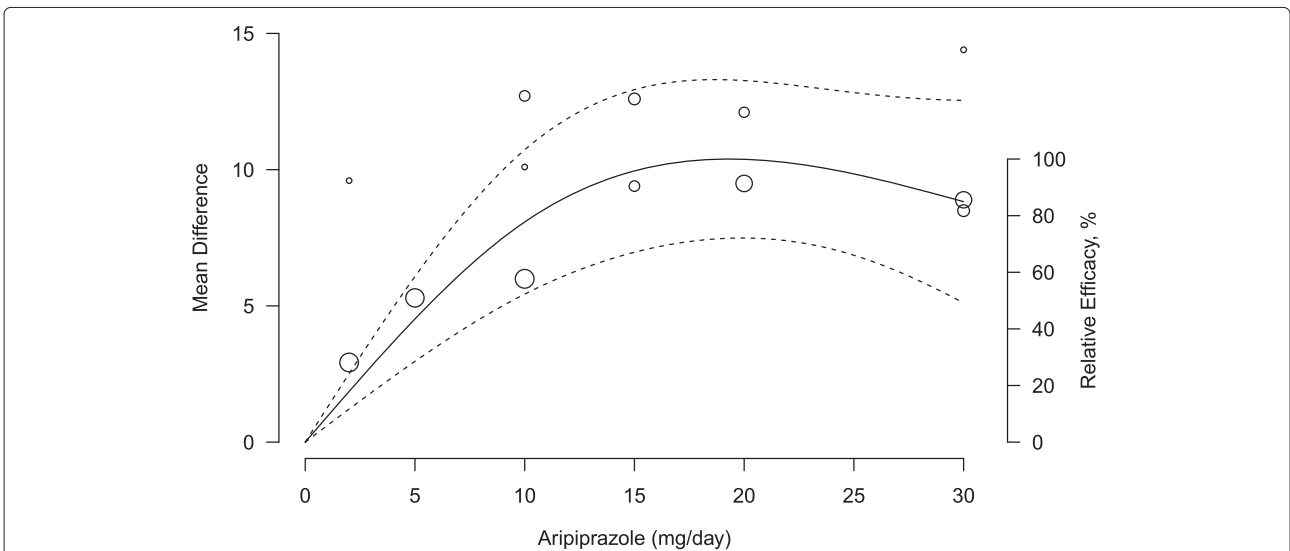
Fig. 2 Pooled dose-response association between aripiprazole and mean change in Positive and Negative Syndrome Scale score (*solid line*). Aripiprazole dosage was modeled with restricted cubic splines in a random-effects model. Dash lines represent the 95 % confidence intervals for the spline model. The placebo group (dose = 0) served as the referent group. Circles indicate observed mean differences in individual studies; size of bubbles is proportional to precision (inverse of variance) of the mean differences. Right axis represents percentage of the maximum predicted effect

Fig. 3 Graphical sensitivity analysis for the pooled dose-response curves between aripiprazole and mean change in Positive and Negative Syndrome Scale score. The placebo group (dose = 0) served as the referent group. Right axis represents percentage of the maximum predicted effect. Left panel: different location of the three knots in a restricted cubic spline model. Right panel: different models, restricted cubic splines (*solid line*), fractional polynomials (*dashed line*), quadratic polynomial (*dotted line*), and E_{max} model (*dot-dashed line*). Circles indicate observed mean differences in individual studies; size of bubbles is proportional to precision (inverse of variance) of the mean differences

To evaluate the sensitivity of the dose-response curve to the choice of the parametric model f adopted, instead, we considered three alternatives: fractional polynomials; quadratic; and E_{max}. Since two studies only had two non-referent doses, the study-specific (sigmoidal) E_{max} models as described in Table 2 cannot be estimated. A common solution is to fix the steepness of the curve θ_3 to be 1, also referred to as hyperbolic E_{max} [20].

The "best" fractional polynomials ($p_1 = 0.5$, $p_2 = 1$) provided overall a similar dose-response curve when compared to the spline model, with slightly higher value for the maximum predicted response (right panel of Fig. 3). The hyperbolic E_{max} had substantially higher predicted mean differences for low values of the dose. The non-linear model assumes a specific hyperbolic dose-response curve that did not seem to fit the data and may be dependent from the choice of fixing θ_3 to be 1. The dose-response curve described by the quadratic model fall in between the spline and the hyperbolic E_{max} curves.

Discussion

In this paper we proposed a statistical method to combine differences in means of quantitative outcomes. The method consists of dose-response models estimated within each study (first stage) and an overall curve obtained by pooling study-specific dose-response coefficients (second stage). The covariance among study-specific mean differences is taken into account in the first stage analysis using generalized least square estimators, while statistical heterogeneity across studies is allowed by multivariate random-effects model in the second stage.

One major strength of the proposed method is that it is fairly general and can accommodate different modeling strategies, including non-linear ones described by

Pinheiro et al. [3]. Non-linear models, however, are defined by at least three or four parameters, and hence require an equal number of dose levels for each single study included in the analysis. Given that some studies may have investigated a lower number of dose levels, exclusion of these studies may result in substantial loss of information. In addition, many non-linear models assume a specific behaviour (e.g. monotonicity) requiring a strong a priori information about the dose-response curve. The choice of the parametric model is critical, since it highly influences the final results [3]. Indeed, the selection of the dose-response model should be informed by subject-matter knowledge as well as understanding of the research questions at hand. We presented the use of regression splines as a flexible tool for modeling any quantitative exposure. The major advantage is that a variety of curves, even non monotonic ones, can be estimated using only two parameters. It is considered to be closed to non-parametric regression, since no major assumptions about the shape of the curve are needed [9]. A possible alternative is the use of fractional polynomials. In comparing the two strategies, we did not find important differences between the two strategies and concluded that both are useful tools to characterize a (non-linear) dose-response curve. Nonetheless a sensitivity analysis is generally required to evaluate the robustness of the combined results.

A possible limitation of the proposed methodology is that it requires information about dose-specific means and standard deviations. Studies providing other summary measures, such as dose-specific medians, would not be included the analysis. The dose-response analysis presented in Eq. 6 is based on the asymptotic normal distribution of the conditional mean effect size. Extension of

the introduced methodology to percentiles is not straight-forward and may represent an interesting topic of future research.

An additional limitation of aggregated dose-response data is that supplementary information for approximating the covariance terms may not be available. Articles may report directly mean or standardized mean differences and standard errors for non-referent dose groups. Whenever the standard deviation for the outcome variable in the control group (s_{i0}^2) cannot be obtained, it may be approximated using the pooled standard deviation based on the non-referent dose levels ($s_{p_{ij}}^2$). Alternatively a specific value may be imputed and a sensitivity analysis can be performed to evaluate how the results of the meta-analysis vary for different values of s_{0j}^2. Further limitations relate to the general application of meta-analysis based on aggregated data. These include restrictions in subgroup analyses, the impossibility of assessing the appropriateness of individual analyses, and to harmonize variable definitions and analyses for reducing the extent of heterogeneity, as well as specific biases such as aggregation (or ecological) bias in meta-regression models. Meta-analysis of individual patient data, however, are often difficult to undertake especially for the availability of individual data, so that usage of aggregated data may represent the only alternative [36]. Specific to aggregated dose-response data, different dose references and exposure range may complicate the analysis. The presented methodology assume that all the selected studies share a common dose-response model. Important departure from this assumption may limit and/or impact the pooling of individual dose-response coefficients. An alternative methods has been proposed based on a series of univariate meta-analyses of effect sizes for a pre-specified grid of dose-levels [37]. Further work is needed to analyze this possibility and potential advantages. Depending on the extent of heterogeneity of the dose-response curves, however, it may not be opportune to pool study-specific results, and meta-regression or stratified analyses should be performed [38].

In our application, we considered the effectiveness of increasing dosages of aripiprazole in shizoaffective patients. We described the steps needed to obtain the overall dose-response curve and to present it in a graphical form. We observed a non-linear association with the maximum efficacy corresponding to aripiprazole 19.32 mg/day. An estimated dose of 10.46 mg/day, however, may be sufficient to obtain 80 % of the maximum effect, which may be relevant for avoiding possible undesired side effects. Sensitivity analysis showed similar results as compared to fractional polynomials. The $E_{\max}$ model presented higher drug efficacy for low dosages. Compared to the previous models, the $E_{\max}$ model did not seem to fit properly the data at low dosages.

Conclusions

We described an approach to combine differences in means of a quantitative outcome contrasting different dose levels relative to a placebo in randomized trials. The general framework of the proposed methodology can include a variety of flexible models. Sensitivity analysis can be a useful tool to assess the stability of the overall dose-response curve to different modelling strategies. Although the method was presented for the analysis of randomized trials, it may be extended to observational studies where mean differences are further adjusted for potential confounders. Future work is needed to evaluate the properties of the statistical model and validity of the underlying assumptions. A user friendly procedure is implemented in the `dosresmeta` R package [39] with worked examples available on GitHub.

Authors' contributions
AC developed the methods and prepared a draft. NO provided critical reviews, corrections and revisions. Both authors read and approved the final version of the manuscript.

Authors' information
AC is a PhD student in Epidemiology and Biostatistics. Dr. NO is Associate Professor of Medical Statistics.

Competing interests
The authors declare that they have no competing interest.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

References
1. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics. 2005;61(3): 738–48. doi:10.1111/j.1541-0420.2005.00344.x. Accessed 24 Feb 2015.
2. Bretz F, Hsu J, Pinheiro J, Liu Y. Dose finding - a challenge in statistics. Biometrical Journal. Biometrische Zeitschrift. 2008;50(4):480–504. doi:10.1002/bimj.200810438.
3. Pinheiro JC, Bretz F, Branson M. Analysis of dose–response studies. modeling approaches. In: Dose Finding in Drug Development. New York City: Springer; 2006. p. 146–71.
4. Ting N. Dose finding in drug development. New York City: Springer; 2006.

5. Chevret S. Statistical methods for dose-finding experiments. New York City: Wiley; 2006.

6. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am J Epidemiol. 1992;135(11):1301–1309.

7. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. Epidemiology. 1993;4(3):218–28.

8. Dumouchel W. Meta-analysis for dose–response models. Stat Med. 1995;14(5-7):679–85.

9. Bagnardi V, Zambon A, Quatto P, Corrao G. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. Am J Epidemiol. 2004;159(11): 1077–1086. doi:10.1093/aje/kwh142. Accessed 25 Nov 2013.

10. Liu Q, Cook NR, Bergström A, Hsieh CC. A two-stage hierarchical regression model for meta-analysis of epidemiologic nonlinear dose–response data. Comput Stat Data Anal. 2009;53(12):4157–167.

11. Orsini N, Bellocco R, Greenland S, et al. Generalized least squares for trend estimation of summarized dose-response data. Stata J. 2006;6(1):40.

12. Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. Am J Epidemiol. 2012;175(1):66–73.

13. Dominici F, Samet JM, Zeger SL. Combining evidence on air pollution and daily mortality from the 20 largest us cities: a hierarchical modelling strategy. J Royal Stat Soc Ser A (Stat Soc). 2000;163(3):263–302.

14. Simmonds MC, Higginsa JP, Stewartb LA, Tierneyb JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials. 2005;2(3):209–17.

15. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med. 2012;31(29): 3821–839. doi:10.1002/sim.5471. Accessed 11 Dec 2013.

16. Davis JM, Chen N. Dose response and dose equivalence of antipsychotics. J Clin Psychopharmacol. 2004;24(2):192–208.

17. Holford NH, Sheiner LB. Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models. Clin Pharmacokinet. 1981;6(6):429–53.

18. Thomas N, Sweeney K, Somayaji V. Meta-analysis of clinical dose-response in a large drug development portfolio. Stat Biopharm Res. 2014;6(4):302–17. doi:10.1080/19466315.2014.924876. Accessed 11 Feb 2015.

19. Geddes J, Freemantle N, Harrison P, Bebbington P. Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. BMJ : British Medical Journal. 2000;321(7273):1371–1376. Accessed 24 Mar 2015.

20. Macdougall J. Analysis of dose–response studies. emax model. In: Dose Finding in Drug Development. New York City: Springer; 2006. p. 127–45.

21. Durrleman S, Simon R. Flexible regression models with cubic splines. Stat Med. 1989;8(5):551–61. doi:10.1002/sim.4780080504. Accessed 30 Mar 2015.

22. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. Stat Med. 2010;29(9): 1037–1057. doi:10.1002/sim.3841. Accessed 25 Nov 2013.

23. Harrell F.E. Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst. 1988;80(15):1198–1202.

24. Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med. 2000;19(14):1831–1847.

25. Jackson D, White IR, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Stat Med. 2012;31(29): 3805–820.

26. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954;10(1):101–29. doi:10.2307/3001666. Accessed 09 Jan 2015.

27. Ritz J, Demidenko E, Spiegelman D. Multivariate meta-analysis for data consortia, individual patient meta-analysis, and pooling projects. J Stat Plann Infer. 2008;138(7):1919–1933.

28. Cutler AJ, Marcus RN, Hardy SA, O'Donnell A, Carson WH, McQuade RD. The efficacy and safety of lower soses of aripiprazole for the treatment of patients with acute exacerbation of schizophrenia. CNS Spectrums. 2006;11(09):691–702. doi:10.1017/S1092852900014784. Accessed 06 Feb 2015.

29. McEvoy JP, Daniel DG, Carson WH, McQuade RD, Marcus RN. A randomized, double-blind, placebo-controlled, study of the efficacy and safety of aripiprazole 10, 15 or 20 mg/day for the treatment of patients with acute exacerbations of schizophrenia. J Psychiatr Res. 2007;41(11): 895–905. doi:10.1016/j.jpsychires.2007.05.002.

30. Kane JM, Carson WH, Saha AR, McQuade RD, Ingenito GG, Zimbroff DL, Ali MW. Efficacy and safety of aripiprazole and haloperidol versus placebo in patients with schizophrenia and schizoaffective disorder. J Clin Psychiatry. 2002;63(9):763–71.

31. Potkin SG, Saha AR, Kujawa MJ, Carson WH, Ali M, Stock E, Stringfellow J, Ingenito G, Marder SR. Aripiprazole, an antipsychotic with a novel mechanism of action, and risperidone vs placebo in patients with schizophrenia and schizoaffective disorder. Arch Gen Psychiatr. 2003;60(7):681–90. doi:10.1001/archpsyc.60.7.681.

32. Turner EH, Knoepflmacher D, Shapley L. Publication bias in antipsychotic trials: an analysis of efficacy comparing the published literature to the US Food and Drug Administration Database. PLoS Med. 2012;9(3). doi:10.1371/journal.pmed.1001189. Accessed 30 Mar 2015.

33. Obermeier M, Schennach-Wolff R, Meyer S, Möller HJ, Riedel M, Krause D, Seemüller F. Is the panss used correctly? a systematic review. BMC psychiatry. 2011;11(1):1.

34. Leucht S, Kissling W, Davis JM. The PANSS should be rescaled. Schizophrenia bulletin. 2010;36:461–462.

35. Leucht S, Davis JM, Engel RR, Kane JM, Wagenpfeil S. Defining response in antipsychotic drug trials: recommendations for the use of scale-derived cutoffs. Neuropsychopharmacology. 2007;32(9):1903–1910.

36. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. BMC Med Res Methodol. 2005;5(1):14.

37. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. Stat Med. 2011;30(28):3341–360.

38. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. Am J Epidemiol. 1994;140(3):290–6.

39. Crippa A, Orsini N. Multivariate dose-response meta-analysis: the dosresmeta r package. J Stat Softw. 2016. In Press.