

# Meta-analysis for psychiatric research using free software R

Ding-Geng CHEN

**Summary:** This paper provides a brief overview of meta-analysis (MA) with emphasis on classical fixed-effects and random-effects MA models. It illustrates the application of MA models with the open-source software R using publicly available data from five studies on lamotrigine to treat bipolar depression and finds that meta-analysis identifies a statistically significant advantage of lamotrigine over placebo that was not evident in the individual studies.

**Key words:** meta-analysis; fixed-effects model; random-effects model; bipolar disorder; lamotrigine

[Shanghai Arch Psychiatry. 2015; 27(3): 195-199. doi: <http://dx.doi.org/10.11919/j.issn.1002-0829.215063>]

## 1. Introduction

In all scientific fields there is an increasing importance placed on the synthesis of information and data from diverse studies to draw more reliable inferences and, thus, to arrive at more robust conclusions. This process of scientifically integrating diverse information is called 'systematic review' and, when it involves generating overall results based on pooled data, is usually referred to as 'meta-analysis' (MA). The widespread use of MA – over 500 have been published in the *New England Journal of Medicine* over the last decade – has led to numerous important discoveries. Interested readers can refer to several recent books about MA: Whitehead,<sup>[1]</sup> Hartung,<sup>[2]</sup> Borenstein,<sup>[3]</sup> Pigott,<sup>[4]</sup> and Chen and Peace.<sup>[5]</sup>

MA has been widely used in psychiatry. This article introduces the classical fixed-effects and random-effects models in MA and uses a meta-analysis about the use of lamotrigine in bipolar disorder<sup>[6]</sup> as an example to illustrate the step-by-step implementation of MA using R, an open source statistical software package that can be freely accessed from <http://www.r-project.org>.

## 2. Meta-analysis models

### 2.1 Fixed-effects MA model

A typical MA combines  $K$  independent studies in which the population effect-size (ES)  $\delta_k$  ( $k = 1, 2, \dots, K$ ) is estimated using the observed ES,  $\hat{\delta}_k$ . These studies can be single-arm studies or multiple-arm studies, randomized controlled studies or observational studies. The MA for the lamotrigine studies used as an example are two-arm studies, where the  $\delta_k$  are the underlying population effect size between the treatment and control groups.

In fixed-effects meta-analysis (FE-MA), a strong assumption is that there is no between-study variation among the  $K$  studies regardless of where, when, with whom, and how the studies were conducted. This is the 'homogeneity assumption', which assumes that the underlying population effect sizes  $\delta_k$  are constant across all studies (i.e.,  $\delta_1 = \dots = \delta_k = \delta$ ) and that the observed study effect sizes  $\hat{\delta}_k$  are a simple random sample from the population with a known sampling error. Therefore a typical FE-MA model can be described as

$$\hat{\delta}_k = \delta + \epsilon_k \quad (2.1)$$

where  $\epsilon_k$  represent the within-study variations and are assumed to be a normal distribution with mean 0 and a known variance  $\sigma_k^2$ , that is,  $\epsilon_k \sim N(0, \sigma_k^2)$ . Note that  $\sigma_k^2$  for all included studies are assumed to be known in MA, which is very different from the assumptions in analysis of variance or regression models where the error variance is estimated. These variances from  $K$  studies can be calculated for dichotomized outcome variables (to be illustrated for the lamotrigine example) but for continuous target variables or other types of variables the variance usually needs to be estimated from previously reported literature. The meta-estimate for the global population ES (combining data from all the independent studies) can then be estimated using a weighted-mean method where the weights are the inverse of these known variances. That is:

$$\hat{\delta} = \frac{\sum_{k=1}^K W_k \hat{\delta}_k}{\sum_{k=1}^K W_k} \quad (2.2)$$

School of Social Work, University of North Carolina at Chapel Hill, Chapel Hill, United States  
correspondence: [DrDG.Chen@gmail.com](mailto:DrDG.Chen@gmail.com)

A full-text Chinese translation of this article will be available at <http://dx.doi.org/10.11919/j.issn.1002-0829.215063> on August 6, 2015.

where  $\omega_k=1/\sigma_k^2(k=1,\dots,K)$ . The variance of this FE-MA global estimate of ES can be expressed as:

$$\text{Var}(\hat{\delta})=1/\sum_{k=1}^K W_k \quad (2.3)$$

Using this weighted-mean estimate in (2.2) and its variance in (2.3), the 95% confidence interval for the global ES of  $\delta$  can be constructed as:

$$\hat{\delta}\pm 1.96\sqrt{\text{Var}(\hat{\delta})} \quad (2.4)$$

and a test-statistic can be formulated as

$$z = \frac{\hat{\delta}-\delta}{\sqrt{\text{Var}(\hat{\delta})}} \quad (2.5)$$

to test the hypothesis that the global ES is zero as  $H_0: \delta=0$  versus  $H_a: \delta \neq 0$ .

### 2.2 Random-effects MA model

The FE-MA is simple, but the fundamental homogeneity assumption of no between-study variation is often too restrictive. Intuitively when we synthesize a group of studies with meta-analysis, we expect these studies to have enough in common to merit combining the information for statistical inference, but it is impractical to require that all included studies have identical true effect size. The homogeneity assumption in FE-MA needs to be relaxed for practical applications of MA to include situations where there is between-study variation in the true effect size of the studies included in the MA, that is, when the study results are ‘heterogeneous’. The random-effects meta-analysis (RE-MA) model is used when the included studies are heterogeneous.

To incorporate heterogeneity as random-effects, it is assumed that the underlying population effect sizes  $\delta_k$  are normally distributed with a global mean of  $\delta$  and a between-study variance of  $\tau^2$ , that is,  $\delta_k \sim N(\delta, \tau^2)$ . Therefore the FE-MA model in (2.1) can be extended to the RE-MA model as:

$$\hat{\delta}_k = \delta_k + \varepsilon_k = \delta + v_k + \varepsilon_k \quad (2.6)$$

where  $v_k \sim N(0, \tau^2)$ . The RE-MA model in (2.6) can also be expressed in a two-level model as follows:

$$\text{Level 1: } \hat{\delta}_k \sim N(\delta_k, \sigma_k^2) \quad (2.7)$$

$$\text{Level 2: } \delta_k \sim N(\delta, \tau^2) \quad (2.8)$$

The global population ES for RE-MA can be estimated using a weighted-mean methods similar to that in the FE-MA model (2.2), but the weights in RE-MA must incorporate both within-study variance ( $\sigma_k^2$ ) and between-study variance ( $\tau^2$ ) as  $\omega_{kR}=1/(\sigma_k^2 + \hat{\tau}^2)$  where the subscript ‘R’ represents notations for the random-effects meta-analysis model to distinguish them from notations for the fixed-effects meta-analysis model. With these new weights, the estimate of global ES in RE-MA is:

$$\hat{\delta}_R = \frac{\sum_{k=1}^K W_{kR} \hat{\delta}_k}{\sum_{k=1}^K W_{kR}} \quad (2.9)$$

The variance of this RE-MA estimate can be expressed as:

$$\text{Var}(\hat{\delta}_R) = 1/\sum_{k=1}^K W_{kR} \quad (2.10)$$

The 95% confidence interval for the global ES of  $\delta$  in RE-MA can be constructed using (2.9) and (2.10) as:

$$\hat{\delta}_R \pm 1.96\sqrt{\text{Var}(\hat{\delta}_R)} \quad (2.11)$$

and a test-statistic can be formulated as

$$z = \frac{\hat{\delta}_R - \delta}{\sqrt{\text{Var}(\hat{\delta}_R)}} \quad (2.12)$$

to test the hypothesis that the global ES is zero as  $H_0: \delta=0$  versus  $H_a: \delta \neq 0$  in RE-MA.

To make use of the estimation in RE-MA from equations (2.9) to (2.12), the estimate of the between-study variance  $\tau^2$  is required. There are several methods to estimate  $\tau^2$ , including the DerSimonian and Laird’s method of moments (MM),<sup>[8]</sup> the maximum likelihood estimation (MLE) method,<sup>[9]</sup> the restricted maximum likelihood (REML) method,<sup>[10]</sup> and the Sidik-Jonkman (SJ) estimator.<sup>[11]</sup> Among these estimators, MM and SJ are distribution-free and non-iterative, whereas both MLE and REML are parametric methods that need multiple iterations to estimate  $\tau^2$ . A discussion that compares these estimators can be found in the paper by Sidik and Jonkman.<sup>[12]</sup>

The most commonly used estimate is the DerSimonian-Laird method of moments which is also commonly called the Cochran-DerSimonian-Laird procedure. This estimate is given as:

$$\hat{\tau}^2 = \max(0, \frac{Q-(K-1)}{U}) \quad (2.13)$$

where  $U$  is a constant defined as  $U = \sum_{k=1}^K w_k - \frac{\sum_{k=1}^K w_k^2}{\sum_{k=1}^K w_k}$  and  $Q$  is the weighted sum of squared errors defined as  $Q = \sum_{k=1}^K \omega_k (\hat{\delta}_k - \hat{\delta})^2$ . This Q-statistic is commonly used to test the statistical significance of heterogeneity across studies; it has a  $\chi^2$  distribution with  $K-1$  degrees of freedom.

It is impossible for a group of independent studies to be identical in every respect, so even when the chi-square value for the Q-statistic suggests that the studies included in the review are homogeneous it is recommended that the RE-MA model be used to combine the results of the studies because this method considers both within-study and between-study variation.

### 3. Meta-analysis about lamotrigine for the treatment of bipolar depression

#### 3.1 Data from five clinical trials on lamotrigine

Although there is definitive evidence of the long-term efficacy of lamotrigine in the maintenance treatment for bipolar I disorder, five placebo-controlled clinical trials of lamotrigine in the acute phase of the illness did not find statistically significant benefit of lamotrigine over

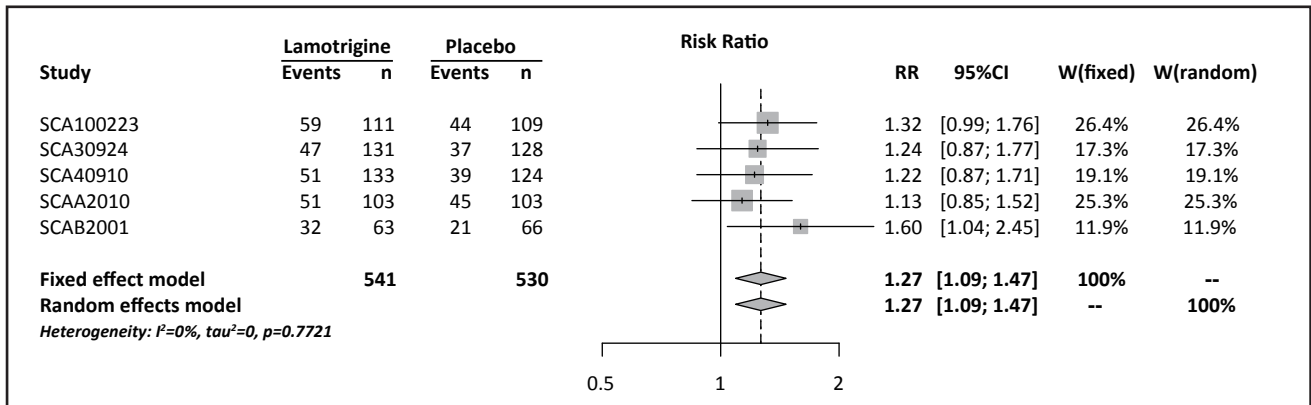
placebo.<sup>[7]</sup> However, Geddes and colleagues<sup>[6]</sup> pooled the patient-level results of these five studies using MA to demonstrate that lamotrigine is, in fact, superior to placebo in the acute phase of bipolar I disorder. A comprehensive description of this analysis is provided in Chen and Peace,<sup>[5]</sup> but the following discussion will be limited to the part of the paper that pooled results for the Hamilton Rating Scale for Depression (HRSD)<sup>[13]</sup> from the five studies. In this analysis, a patient was considered a ‘responder’ if he or she experienced at least a 50% reduction from baseline in the HRSD. The basic results for the lamotrigine group and the control group from the five studies are shown in the first five columns of Figure 1; in this figure the ‘Total’ columns represent the sample size in each group and the ‘Events’ columns represent the number of individuals in each group who met the ‘responder’ criteria at the end of the trial.

### 3.2 Meta-analysis with risk-ratio

The purpose of meta-analysis is to combine individual estimates of treatment effect or effect sizes (ESs) across studies. If estimates of the treatment effect or effect size are not provided for the individual studies but the number of patients who respond to treatment are provided (as in this example), it is possible to calculate the effect size for each study and to subsequently pool the estimated effect sizes across all of the studies in a meta-analysis. For binomial outcome measures, such as response to treatment versus non-response to treatment, the most commonly use estimator of effect size is the risk-ratio. The risk-ratios for the studies included in the MA of the effectiveness of lamotrigine are defined as:

$$RR = \frac{P_L}{P_C} = \frac{E_L/T_L}{E_C/T_C} \quad (3.1)$$

Figure 1. Forest plot for meta-analysis of five lamotrigine clinical trials



where the risk (i.e., the probability,  $p$ ) in each group is the proportion of the ‘total’ sample (i.e., ‘ $T$ ’) in each group that experience the ‘event’ (i.e., ‘ $E$ ’) of interest (in this case, respond to treatment); for the lamotrigine group  $p_L=E_L/T_L$  and for the control group  $p_C=E_C/T_C$ . The method of estimating the variance of this risk ratio is based on the normal distribution approximation; the RR is transformed using the natural logarithm and the variance of the natural log of RR is estimated using the delta method:

$$Var(\ln(RR)) = \frac{1}{E_L} - \frac{1}{T_L} + \frac{1}{E_C} - \frac{1}{T_C} \quad (3.2)$$

Subsequently, the point estimate for  $\ln(RR)$  and the corresponding confidence intervals are transformed back to RR and the confidence interval for RR.

When conducting the MA using R, data from column 1 to 5 in Figure 1 would first be loaded into R as:

```
> evlamo = c(59,47,51,51,32);
nlamo = c(111,131,133,103,63)
> evcont = c(44,37,39,45,21);
ncont = c(109,128,124,103,66)
> trial = c("SCA100223", "SCA30924",
"SCA40910", "SCAA2010", "SCAB2001")
```

The MA can be done using the function ‘metabin’ (denoting meta-analysis for binary variables) in the R library labelled ‘meta’. [For explanation about the use of this R library, activate the help function by entering ‘library(help=‘meta’)', which will display all the functionalities for this library.] The R coding to meta-analyze the five lamotrigine trials using this function would be as follows:

```
# load the library
> library(meta)
# call ‘metabin’ for MA
> RR1.Lamo = metabin(evlamo, nlamo, evcont,
ncont, studlab=trial,
label.e='Lamotrigine', label.c='Placebo',
method='inverse', sm='RR')
```

In the above coding, ‘studlab’ is the data field when the labels for the different studies are located (i.e., ‘trial’); ‘label.e’ is the label assigned to the experimental group (‘Lamotrigine’); ‘label.c’ is the label assigned to the control group (‘Placebo’); ‘method’ is the method used to pool the studies (‘inverse’, inverse-weighting as described above for FE-MA [2.2] and RE-MA [2.9]); and ‘sm’ is the summary measure being used (‘RR’, risk-ratio). Alternative methods that can be selected in R

to combine the target measure included the Mantel-Haenszel method (which was used in Geddes and colleague’s report<sup>[6]</sup>) or the Peto method. The printout for this coding would be as follows:

```
# Print the summary
> RR1.Lamo
# Summary information:
                RR      95%CI      % W(fixed)
SCA100223  1.3167 [0.9879; 1.7551]  26.41
                26.41
SCA30924   1.2412 [0.8700; 1.7706]  17.27
                17.27
SCA40910   1.2192 [0.8699; 1.7088]  19.13
                19.13
SCAA2010   1.1333 [0.8451; 1.5198]  25.32
                25.32
SCAB2001   1.5964 [1.0398; 2.4509]  11.86
                11.86
```

Number of studies combined: k=5

```
                RR      95%CI      z
Fixed effect model  1.265 [1.0914; 1.4663]  3.1205
                0.0018
Random effects model 1.265 [1.0914; 1.4663]  3.1205
                0.0018
```

Quantifying heterogeneity:

```
tau^2 = 0; H = 1 [1; 1.47]; I^2 = 0% [0%; 53.8%]
```

Test of heterogeneity:

```
Q      d.f.  p.value
1.8    4    0.7721
```

Details on meta-analytical method:

- Inverse variance method
- DerSimonian-Laird estimator for tau^2

The first part of the summary provides the risk-ratios, associated 95% confidence intervals, and the weightings for fixed-effects and random-effects MA models for each individual study based on the formulae in (3.1) and (3.2). Notice that the weights for both fixed-effects and random-effects are the same since the heterogeneity is not statistically significant and therefore the estimated  $\tau^2=0$ . The first four trials are not statistically significant (i.e., the 95% confidence intervals of the RRs include 1, which is the RR at which lamotrigine and placebo are equally effective), but the fifth trial is statistically significant (RR=1.60, 95% CI=1.04,2.45), indicating a significant advantage for lamotrigine over placebo. The second part of the summary provides the pooled RR and the associated 95% confidence intervals both for the fixed-effects and random-effects models; in this example both pooled RRs were statistically significant and the values using the FE-MA and RE-MA models were identical.

The last part of the summary first quantifies the level of heterogeneity of the included studies and then tests whether or not there is statistically significant heterogeneity. In the ‘Quantifying heterogeneity’ section, between study variance ( $\tau^2$  or ‘tau^2’) is estimated to be 0; the standardized heterogeneity index (H) is estimated to be 1 with a 95% CI of 1 to 1.47, and the proportion of the total variance attributed to between-study heterogeneity ( $I^2$ ) if 0% with a 95% CI of

0% to 53.8%. In the ‘Test of heterogeneity’ section, the p-value for the Q-statistic, 0.772, is not significant which indicates that there was no significant heterogeneity between the five studies (which is the reason the pooled RRs for the fixed-effect and random-effect models were so similar).

The Q statistic only assesses the presence or absence of heterogeneity. Test can only tell us about the presence versus the absence of heterogeneity and ‘Quantifying heterogeneity’ is then to report the extend of such heterogeneity which shows that the between-study heterogeneity  $\tau^2$  (‘tau^2’) is estimated to be 0; the standardized heterogeneity index H is estimated to be 1 with 95% CI [1; 1.47]; the measure of proportion of observed heterogeneity from the total heterogeneity  $I^2 = 0\%$  with 95% CI [0%; 53.8%] indicating again that there is no statistically significant heterogeneity for this MA.

For any MA, a forest plot is typically produced for summary and publication purposes. The forest plot for this dataset shown in Figure 1 can be simply produced in R by using the ‘forest’ function as follows:

```
> forest(RR1.Lamo)
```

### 3.3. Meta-analysis with risk-difference and odds-ratio

The risk-ratio is probably the most commonly used measure of ES in MA for binomial data, but other measures of ES include the risk-difference and odds-ratio. The definition of risk difference (RD) is simply the difference of the risks between a treatment (or intervention) group and control group, defined as:

$$RD = P_T - P_C \quad (3.3)$$

with the risk (or probability) of a target outcome defined as in (3.1). The statistical inference for RD is to test whether this RD is statistically significant different from zero.

The odds-ratio (OR), which is familiar because of its use in logistic regression, is intuitively less appealing than the RR or RD. The odds-ratio (OR) associated with an event is defined as the ratio of the odds of the event in one study group to the odds of the event in another study group. The odds of the event is defined as:

$$odds = \frac{p}{1-p} \quad (3.4)$$

Thus the odds-ratio (OR) of the treatment group (such as, lamotrigine) to the control group for kth study can be formulated as follows:

$$OR = \frac{odds_T = \frac{p_T}{1-p_T}}{odds_C = \frac{p_C}{1-p_C}} \quad (3.5)$$

The statistical inference for the OR in meta-analysis is usually conducted by converting the odds-ratio to the log scale and estimating the log odds-ratio and its standard error based on an approximate normal distribution.

The implementation of these alternative methods for estimated ES in R is very straightforward. It is done by specifying  $sm='RD'$  for risk-difference and  $sm='OR'$  for odds-ratio in the R function ‘metabin’ in the coding block shown in section 3.2.

#### 4. Discussion

This paper provides an overview for meta-analysis using the classical fixed-effects and random-effects models and illustrates the models using the 'meta' package in R. Other commonly used R packages for conducting meta-analysis, such as 'rmeta' and 'metafor', and extensive illustrations of their application can be freely accessed from <http://www.r-project.org>. Further description of this meta-analysis methodology and its implementation

using R is available in *Applied Meta-Analysis with R* by Chen and Peace.<sup>[5]</sup>

#### Conflict of interest

The author declares no conflict of interest.

#### Funding

No funding was provided to prepare this paper.

## 使用免费软件 R 进行精神病学研究的 Meta 分析

Chen DG

概述：本文对 meta 分析（meta-analysis, MA）作了简要概述，并着重于经典的固定效应和随机效应 MA 模型。文章以综合分析 5 项使用拉莫三嗪治疗双相障碍抑郁发作的研究的公开数据为范例，介绍了 MA 模型的应用及免费软件 R 在 meta 分析中的使用方法，该 meta 分析证实拉莫三嗪的效果优于安慰剂，且有统计

学显著性，而在单个的研究中此优势并不明显。

**关键词：**meta 分析；固定效应模型；随机效应模型；双相情感障碍；拉莫三嗪

本文全文中文版从 2015 年 08 月 06 日起在

<http://dx.doi.org/10.11919/j.issn.1002-0829.215063> 可供免费阅读下载

#### References

- Whitehead A. *Meta-Analysis of Controlled Clinical Trials*. New Jersey: John Wiley & Sons, Inc.; 2003
- Hartung J, Knapp G, Sinha BK. *Statistical Meta-Analysis with Applications*. New Jersey: John Wiley & Sons, Inc.; 2008
- Borenstein M, Hedges, LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. New Jersey: Wiley; 2009
- Pigott TD. *Advances in Meta-Analysis*. New York: Springer; 2012
- Chen DG, Peace KE. *Applied Meta-Analysis with R*. Chapman and Hall/CRC Biostatistics Series. London: Taylor & Francis Group, Inc.; 2013
- Geddes JR, Calabrese JR, Godwin GM. Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomized trials. *BJP*. 2009; **194**: 4-9
- Calabrese JP, Huffman RF, White RL, Edwards S, Thompson TR, Ascher JA. Lamotrigine in the acute treatment of bipolar depression: results of five double-blind, placebo-controlled clinical trials. *Bipolar Disorder*. 2008; **10**(2): 323-333. doi: <http://dx.doi.org/10.1111/j.1399-5618.2007.00500.x>
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986; **7**: 177-188. doi: [http://dx.doi.org/10.1016/0197-2456\(86\)90046-2](http://dx.doi.org/10.1016/0197-2456(86)90046-2)
- Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med*. 1998; **17**(8): 841-856. doi: [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19980430\)17:8<841::AID-SIM781>3.0.CO;2-D](http://dx.doi.org/10.1002/(SICI)1097-0258(19980430)17:8<841::AID-SIM781>3.0.CO;2-D)
- Raudenbush SW, Bryk AS. Empirical Bayes meta-analysis. *J Educ Stat*. 1985; **10**(2):75-98. doi: <http://dx.doi.org/10.2307/1164836>
- Sidik K, Jonkman JN. Simple heterogeneity variance estimation for meta-analysis. *J R Stat Soc Series C*. 2005; **54**(2): 367-384. doi: <http://dx.doi.org/10.1111/j.1467-9876.2005.00489.x>
- Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. *Stat Med*. 2007; **26**(9): 1964-1981. doi: <http://dx.doi.org/10.1002/sim.2688>
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; **23**(1): 56-62

(received 2015-05-29; accepted, 2015-06-18)



Ding-Geng (Din) Chen received his Ph.D. in Statistics from University of Guelph (Canada) in 1995 and he is now the Wallace H. Kurlalt Distinguished Professor and Director of Statistical Development and Consultation at the School of Social Work, University of North Carolina at Chapel Hill. Professor Chen was a professor in biostatistics at the University of Rochester from 2010 to 2015 and the Karl E. Peace endowed eminent scholar chair in biostatistics from the Jiann-Ping Hsu College of Public Health at the Georgia Southern University from 2009 to 2010. Professor Chen is also a senior biostatistics consultant for biopharmaceuticals and government agencies with extensive expertise in clinical trials and bio-informatics. He has more than 100 referred professional publications and co-authored/co-edited six books on 'Clinical Trial Methodology (2010)', 'Clinical Trial Data Analysis Using R(2011)', 'Interval-Censored Time-to-Event Data: Methods and Applications (2012)', 'Applied Meta-Analysis Using R(2013)' and 'Clinical Trials Biostatistics and Biopharmaceutical Application (2014)' published by Chapman and Hall/CRC and 'Innovative Statistical Methods for Public Health Data (2015)' by Springer.