

## DOCTORAL THESIS

### Statistical learning of median in meta-analysis

Luo, Dehui

*Date of Award:*  
2017

[Link to publication](#)

#### **General rights**

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

# Abstract

As one of the most recommended strategies in decision making of contemporary medicine, evidence-based medicine (EBM) is attracting more and more attention. For EBM, the scientific evidences are obtained mainly from the randomized controlled trials (RCTs), systematic review and meta-analysis. In particular, meta-analysis can help researchers statistically combine several independent studies for a same clinical problem. In meta-analysis, Cohen's  $d$  and Hedges'  $g$  are among the most commonly used effect size measurements for continuous data. To compute these mean difference criteria, the sample mean and standard deviation are two conventional statistics reported in the literature. However, some other clinical studies may instead report the median, minimum and maximum values, and/or sample quartiles. Such a situation requires researchers to estimate the sample mean and standard deviation from these reported summary statistics. We note, however, that most existing estimators in the literature have some serious limitations. For this, we propose to improve the existing methods and extend them to three frequently encountered scenarios.

In this thesis, we developed the optimal sample mean estimators, the normality test statistics and the updated Cohen's  $d$  mean difference criterion for three commonly encountered scenarios in meta-analysis. In Chapter 1, we gave a brief introduction on evidence-based medicine, meta-analysis and the summary data. In Chapter 2, we introduced our optimal estimators of the sample mean under the three scenarios, respectively. In Chapter 3, we proposed several methods for testing the normality of the underlying data. And in Chapter 4, we proposed to improve the famous Cohen's  $d$  and its relevant parameters. To assess the practical performance of our newly proposed methods, we also chose a few real data at the end of each chapter as illustrating examples. Numerical results of those studies indicated that our proposed methods have satisfactory performance both in theory and in practice. Following our new methodology, we also recommended an improved procedure for medical researchers to conduct meta-analysis. For illustration, we chose a meta-analysis in Chapter 5 on the effect of phytosterols to plasma CRP level (Rocha et al., 2016) to compare the results obtained from our recommended procedure and from the original methods.

The results showed that our recommended procedure may lead to distinctly different results for a same clinical problem. To conclude the thesis, we expect that our newly proposed methods can be regarded as “rules of thumb” and will soon be widely applied in meta-analysis and evidence-based medicine.

**Keywords:** Cohen’s  $d$ , Fixed-effect model, Median, Meta-analysis, Mid-range, Mid-quartile range, Normality test, Random-effects model, Sample mean, Sample size.

# Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
Chapter 1 Introduction	1
1.1 Evidence-based medicine . . . . .	1
1.2 Meta-analysis . . . . .	2
1.3 Median and summary statistics . . . . .	2
1.4 Recommended procedure of meta-analysis . . . . .	3
1.5 Outline of the thesis . . . . .	4
Chapter 2 Optimally Estimating the Sample Mean from the Median, Extrema and/or Quartiles	7
2.1 Introduction . . . . .	7
2.2 Sample mean estimation . . . . .	8
2.2.1 Existing methods . . . . .	9
2.2.2 Improved methods . . . . .	11
2.3 Simulation studies . . . . .	20
2.3.1 Simulation study for $\mathcal{S}_1 = \{a, m, b; n\}$ . . . . .	20
2.3.2 Simulation study for $\mathcal{S}_2 = \{q_1, m, q_3; n\}$ . . . . .	23
2.3.3 Simulation study for $\mathcal{S}_3 = \{a, q_1, m, q_3, b; n\}$ . . . . .	23
2.4 Real data analysis . . . . .	23

2.4.1	Data description . . . . .	28
2.4.2	Results and comparison . . . . .	29
2.5	Conclusion . . . . .	30
2.6	Appendix . . . . .	33
Chapter 3	Testing Symmetry of underlying data using the Median and Quartiles with Application to Meta-analysis	43
3.1	Introduction . . . . .	43
3.2	Test methods . . . . .	47
3.2.1	Hypothesis test for $\mathcal{S}_1$ . . . . .	47
3.2.2	Hypothesis test for $\mathcal{S}_2$ . . . . .	48
3.2.3	Test statistic for $\mathcal{S}_3$ . . . . .	50
3.3	Simulation studies . . . . .	51
3.3.1	Simulation study for $\mathcal{S}_1$ . . . . .	51
3.3.2	Simulation study for $\mathcal{S}_2$ . . . . .	52
3.3.3	Simulation study for $\mathcal{S}_3$ . . . . .	54
3.4	Real data analysis . . . . .	56
3.4.1	Case study of investigating the association of asthma diagnosis with leptin and adiponectin . . . . .	58
3.4.2	Case study of investigating the effects of statin therapy on plas- ma lipid concentrations in HIV-infected patients . . . . .	62
3.4.3	Case study of exploring the impact of statin therapy on plasma MMP-3, MMP-9 and TIMP-I concentrations . . . . .	68
3.5	Conclusion . . . . .	71
3.6	Appendix . . . . .	75
Chapter 4	Improved Cohen's $d$ and Relevant Parameters in Meta-analysis by using the Median and Quartiles	80
4.1	Introduction . . . . .	80
4.2	Updated Cohen's $d$ value and its variance in meta-analysis . . . . .	82
4.2.1	Updated Cohen's $d$ value for $\mathcal{S}_1$ . . . . .	83
4.2.2	Updated Cohen's $d$ value for $\mathcal{S}_2$ . . . . .	86

4.2.3	Updated Cohen's $d$ value for $\mathcal{S}_3$ . . . . .	87
4.3	Fixed-effect model . . . . .	91
4.3.1	Updated weight of each study for $\mathcal{S}_1$ . . . . .	92
4.3.2	Updated weight of each study for $\mathcal{S}_2$ . . . . .	94
4.3.3	Updated weight of each study for $\mathcal{S}_3$ . . . . .	95
4.4	Random-effects model . . . . .	96
4.4.1	Updated $\tau^2$ estimators and weight of each study for $\mathcal{S}_1$ . . . . .	101
4.4.2	Updated $\tau^2$ estimators and weight of each study for $\mathcal{S}_2$ . . . . .	104
4.4.3	Updated $\tau^2$ estimators and weight of each study for $\mathcal{S}_3$ . . . . .	107
4.5	Real data analysis . . . . .	110
4.5.1	Data description . . . . .	110
4.5.2	Results and comparison . . . . .	111
4.6	Conclusion . . . . .	116
4.7	Appendix . . . . .	118
Chapter 5 Real Data Analysis Following the Recommended Procedure of Meta-analysis . . . . .		125
5.1	Introduction . . . . .	125
5.2	Data description and preparation . . . . .	126
5.3	Results and comparison of meta-analysis . . . . .	128
5.4	Conclusion . . . . .	130
Chapter 6 Conclusion and Discussion . . . . .		133
6.1	Summary of the thesis . . . . .	133
6.2	Future work . . . . .	134
Curriculum Vitae . . . . .		151