# 統合分析文獻綜述暨其在生產與品質管理的應用

# Meta-Analysis: A Review and its Applications in Production and Quality Management

鄭湘芸[1&*] Hsiang-Yun Lisa Cheng,
王事民 Shi-Min Alex Wang[1],
陳惠貴 Huei-Kuei Chen[1],
黃思齊 Szu-Chi Aaron Huang[1],
陳星洋 Hsing-Yang Austin Chen[1],
王銘宗 Ming-Tzong Wang[1&**]
1. Graduate Institute of Industrial Engineering, National Taiwan University

## 摘要

本文旨在提供統合分析的基礎概念，包括其理論基礎、相關軟體比較，並系統性的文獻回顧統合分析在生產管理領域的發展與應用。此外基於統合分析工具的必要性及可用性，本研究還側重於統合分析軟體的比較和歸類。結果顯示在生產與品質管理領域，已有統合分析應用於品質管理(QM)與即時生產系統(JIT)的相關論文，主要應用此研究方法來探討管理實務與企業績效之間的關係，而本文主要利用此兩篇個案論文比較其與一般認定的統合分析標準流程之差異與領域成熟度。

**關鍵字：**文獻綜述、生產與品質管理、統合分析

## Abstract

This paper aims to provide a tutorial on meta-analysis, including basis/theories, software, especially its applications in production and quality management field. A structured literature review is applied on the development of meta-analysis in production management field. Besides, this study also focuses on the software comparison of meta-analysis and categorizes the meta-analysis tools based on the usability and necessity. The results indicate that there are some papers apply the meta-analysis method on the production management such as the quality management (QM) practices and just-in-time (JIT) manufacturing practices. The main applications are to investigate the relationship between management practices and firm performance.

**Keywords:** Literature Review, Product and Quality Management, Meta-Analysis

[*]. First author:國立臺灣大學工業工程所碩士生(聯絡電話：0912918908，E-mail: r00546009@ntu.edu.tw)
[**]. Corresponding author: 國立臺灣大學工業工程所副教授(聯絡電話：(02)3366-3366 ext. 69502，E-mail: mtwang@ntu.edu.tw)

## I. Introduction

In this paper, the methods of Meta-Analysis (MA) are discussed. As the twenty century unfolds, more and more researchers put emphasis on scientific research (Rosenthal & DiMatteo, 2001). This research methodology, meta-analysis, is also high-profile since Gene V. Glass created this term in 1976 (Glass, 1976).

This article gives tutorial instructions on MA intended for beginners. On the whole, findings are often conflicting about central issues in many kinds of domain (Rosenthal & DiMatteo, 2001). For the goal of combining and understanding literatures, MA possesses significant advantages to convince readers (Kavale, 2001). MA is a kind of method could increase power to investigate an overall treatment effect or relationship (Normand, 1999).

We highlight the three parts which cover basis, software and applications for novices. By means of brief introduction, let readers can quickly realize how to conduct the MA procedures and then choose appropriate software. In addition, this paper also provides the applications in production and quality management field, especially for the quality management (QM) practices and just-in-time (JIT) manufacturing practices.

## II. Meta-Analysis: A Primer

Meta-analysis is statistical synthesis method for analyzing and summarizing a series of various studies on a specific topic or a particular research question. Compared with the traditional literature review, it provides more mathematical results in a quantitative way. In addition, there are three rationales to conduct meta-analysis. If this issue has not been summarized quantitatively (Larsson, 2007) or there are some conflicting information against the international or domestic guidelines (Anema, 2008), we consider adapting this statistical method. Besides, when findings are inconsistent among literatures and don’t have enough power (van Mierlo, 2006), it’s a better way to analyze by MA because this method will increase more power and precision by collecting and summarizing researches on a similar research topic.

### 1. History and Evolution

In the beginning of astronomy field, someone found that small-scale data integration is much accurate than the single selected data (Juang, 2011). The first formal use of statistical methods to integrate research data originated since 1904, the famous statistician Karl Pearson performed. He analyzed the results by a group of studies to overcome the problem of no enough statistical power due to small sample sizes ("Meta-analysis," 2012). However, Pearson's algorithm was still not a formal and rigorous way until Gene V. Glass created the term of “Meta-Analysis” in 1976. With the support of his colleague Smith, they published a

psychotherapy study using the method by statistically combing and integrating the results of 375 controlled evaluations. Thus, we may address the Glass as the father of meta-analysis. Since the meta-analysis method was emerged, it was applied on many fields gradually.

Meta-analysis is more and more popular research method for integrating research findings nowadays. With increasing amounts of meta-analysis articles, we could find its importance in the research field. Especially, it is pushed by two organizations, the Cochrane Collaboration in England and Campbell Collaboration in America. Its situation suggests us to know how to conduct the meta-analysis.

## 2. Advantages and Disadvantages of Meta-Analysis

Compared to traditional literature reviews, meta-analysis has definite methodology and uses a quantified standard index to summarize and compare among the included studies (Marsh, H. W. et al., 2008). The main contribution is that it can synthesize many small-scale studies especially for medical field, due to lack of statistical power for small-scale study. We can conduct meta-analysis to summarize them in an objective and impartial way to increase power (Juang, 2011). Furthermore, we have ability to control between-study variation, find relationships across studies (Lipsey & Wilson, 2001), and provide evidential subgroup analysis by moderators. Then, it's easy to interpret results by summary statistics.

However, meta-analysis still has its weakness. First, the results of meta-analysis are not completely reliable because the sources of bias cannot be controlled in this method ("Meta-analysis," 2012). Secondly, Meta-analysis results may contravene the results of randomized controlled trials (RCT). If this event happened, we should spend much time to find the discrepancy between MA and RCT (Juang, 2011). Thus, we should cautiously evaluate the effect of heterogeneity.

The possible limitations of meta-analysis are that we could not find out the publications which have negative or null findings, and we cannot avoid that the difference between-studies could be correlational. It was criticized by other people because the theory of "Apple and Oranges" or "Garbage in, Garbage out." Actually, it depends on how researchers define their research ranges and criteria.

## 3. Meta-Analysis Procedures

There are many different kinds of procedures of MA from the published books or websites. Especially, Cochrane collaboration, which is the reputable MA collaboration, has its regularization form. It requires that every Cochrane review should follow the PRISMA checklist which covers 27 steps. As of December, 2011 in Cochrane Library, cited from Cochrane website,

there are 4892 Cochrane reviews and total 9069 Cochrane reviews plus protocols, covered a wild range of medical discipline. This database can't be underestimated.

Therefore, this article just provides readers a general analytic process to apply different cases. Here we separate research process into eight steps in the following:

i. Defining a clear research question

ii. Setting up clear inclusion and exclusion criteria for selection of studies

iii. Conducting a systematic literature search

iv. Defining/Coding the independent and dependent variables

v. Entering data

vi. Analyzing data

vii. Assessing publication bias/ the effect of unpublished data

viii. Interpreting and concluding the results

Basically, meta-analysis is generally followed by the above flows. In the beginning of the analysis, defining a clear research question is very important. It affects the criteria (step 2) you have to include studies into the research. Then, after defining problem, we can set up clear eligibility to search amount of studies in the literature searching database and clearly define the keywords we use in the databases. A systematic literature research will help you search in an efficient way. By the MA software, you can define the variables then enter the data into programs and make use of the functions of MA software to analyze data. In addition, assessing the publication bias such as calculating fail-safe N number is to see whether you search the enough data already. Finally, we can interpret our results for readers to integrate the conclusion on the research question.

## 4. Statistical Methods

This portion includes effect size and variance, fixed-effect versus random-effects models, heterogeneity, and bias issue.

### (1) Effect Size and Variance

Effect size is a common index in the meta-analysis. Borenstein et al. (2009) provide a viewpoint that effect size includes the meaning of treatment effects, single group summaries, or just a generic statistic. The synthesis of studies should be in the same level; therefore, any

standardized index could be an effect size.

We could choose the appropriate effect size based on three major considerations (Borenstein et al., 2009). The first one is the effect size should be comparable across studies you include. The second one is the effect size should be computable. The final factor we consider is the index should have good mathematical properties. For example, the sample distribution of effect size should be known. In the basis of the above principles, you could find an effect size for conducting meta-analysis.

However, different kinds of data types have their appropriate effect size. Here, the data types could be divided into three cases to discuss.

Table 1 There are different cases to use the different effect size based on data types

| Data Types | Effect sizes |
|---|---|
| Means and Standard Deviations | Raw (unstandardized) mean difference (D)<br>Standardized mean difference (d or g)<br>Response ratio (R) |
| Binary data | Risk ratio (RR)<br>Odds ratio (OR)<br>Risk Difference (RD) |
| Correlation | Correlation (r) |

Effect size is our main standardized index when we conduct the meta-analysis. However, the selecting of effect size is very straightforward. We have to choose the interpretable one.

### (2) Fixed-Effect versus Random-Effects Models

In the process of conducting meta-analysis, we have to decide a statistical model to calculate the summary effects synthesizing the effect sizes gotten before. The mainly used models are fixed-effect model and random-effects model. The former model is assumed that the true effect of population is the same within all studies we include. The observed bias of this model only results from the sampling error. Therefore, we could also call it the common-effect model.

With regard to the latter model, it allows the true effect could change from study to study. It means that it will have different effect sizes in different studies. Unless we could get infinite studies (sample sizes) in our research, the true effect would be composed of some distribution about mean. On the whole, the fixed-effect model is easier than the other. The former model relatively needs more strong assumptions.

Depending on different models, we can get our summary effects. In general, we would give different weights for studies, and then we could calculate the summary effects we want. According to different model's definition and process, we could consider a lot of factors

affecting the model then defining a correct model. Therefore, the formulas of these models are of course different. Through the test we could decide the appropriate model.

Fixed-Effect Model

After a series of considerations, we determine to choose the fixed-effect model as our research model. This model assumption is that all of studies share a same true effect (setting theta, $\theta$.) Generally speaking, the observed effect $Y_i$ is composed of population mean plus the sampling error. In other words,

$$Y_i = \theta + \epsilon_i \tag{1}$$

Here we use $Y_i$ to estimate the population effect of all of studies. Nevertheless, if we give different studies the same weight, it could result in less precise of estimation. Thus, combining with the way of Inverse Variance Weight, we compute a weighted mean in the analysis. First of all, using the Inverse Variance Weight method, the weight assigned to each study is

$$W_i = {}^{1}/_{V_{Y_i}} \tag{2}$$

where $V_{Y_i}$ is the variance of within-study for study (i). Then, we could calculate the weighted mean is

$$\mathrm{M} = \sum_{i=1}^{k} W_i Y_i / \sum_{i=1}^{k} W_i \tag{3}$$

The above formula (3) is the summary effect we pursue. According to formula (3), we apply for its variance is

$$V_M = 1/\sum_{i=1}^{k} W_i \tag{4}$$

Naturally, 95% confidence interval for the summary effect is in the following,

$$LL_M = M - 1.96 * SE_M \tag{5}$$

and

$$UL_M = M + 1.96 * SE_M \tag{6}$$

Finally, testing the null hypothesis (the true effect $\theta$ is zero) by Z statistic,

$$\mathrm{Z} = M/SE_M \tag{7}$$

In the case of one-tail test, the p-value will be

$$p = 1 - \Phi(\pm|Z|) \tag{8}$$

where we hope that the difference is '+' and otherwise, '-' direction will be the case of two-tail test, the p-value will be

$$p = 2 * [1 - \Phi(|Z|)] \tag{9}$$

where $\Phi(Z)$ is the standard normal cumulative distribution.

The above process is the analysis based upon the fixed-effect model. This model is really easy to know and use. However, the real cases we met are usually not good at all. Therefore, we could consider using another statistical model, random-effect models, to continue our analysis.

Random-Effects Model

We further describe the random-effects model and how we use in our research process. Due to no limitations on the same true common effect in this model, it allows variation among studies. However, this model assumes the true effect follows the normal distribution. The notation is the true variation in effect sizes ( $\zeta_i$) and sampling error ( $\varepsilon_i$) (Juang, 2011).

More generally, the observed effect $Y_i$ is composed of the grand mean, the deviation of the true effect from the grand mean, and the sampling error. In other words,

$$Y_i = \mu + \zeta_i + \epsilon_i \tag{10}$$

The above formula is obviously composed of within-study variance and between-studies variance ( $\tau^2$).

Thus, we have to estimate the between-studies variance at first. According to the method of moments, we compute the T2 statistic in order to estimate $\tau$ 2

$$T^2 = (Q - df)/C \tag{11}$$

where

$$Q = \sum_{i=1}^{k} W_i Y_i^2 - \left(\sum_{i=1}^{k} W_i Y_i\right)^2 / \sum_{i=1}^{k} W_i \tag{12}$$

$$df = k-1 \tag{13}$$

where k is the number of studies, and

$$C = \sum_{i=1}^{k} W_i - \sum_{i=1}^{k} W_i^2 / \sum_{i=1}^{k} W_i \tag{14}$$

In the random-effects model, we also use the method of "Inverse Variance Weight" to calculate the weighted mean. The only difference is the random-effects model extra includes the between-studies variance. Therefore, we could just change the part of variance and follow the process of fixed-effect model in the previous section. The variance is

$$V_{Y_i}^{*} = V_{Y_i} + T^2 \tag{15}$$

We could find out this model is relatively not easy but may common use. However, we could meet the difficult and complex cases in the future. Therefore, we could consider using the mix-effect models, combined with the fix-effect and random-effects. Latter, we will detail describe the difference between fix-effect and random-effects model.

Differences in Fixed-Effect and Random-Effects Models

According to the above two sections, we can realize what the fixed-effect and random-effects models are. Here, we provide the differences among these models to help researchers to choose appropriate model. The comparable table Table 2 is shown in the following table.

Table 2 The differences between fixed-effect and random-effects model

| | **Fixed-effect model** | **Random-effects model** |
|---|---|---|
| **Summary effect** | the estimate of the true effect size | the estimate of the mean of the effect size |
| **Weight assigned** | The larger the effect size, the higher weight it will be assigned. The extreme case exists. | More balanced than fixed-effect model. If the study is large, it would lose influence, vice versa. |
| **C.I.'s width** | The width of confidence interval is zero. | The width of confidence interval would not approach zero. |
| **Standard error of summary effect** | $SE_M = \sqrt{\sigma^2/(k*n)}$ | $SE_M = \sqrt{\sigma^2/(k*n) + \tau^2/k}$ (The second term means between-studies variance.) |
| **Null hypothesis** | There is zero effect in every study. | The mean effect is zero. |

The above table tells that these two models are different in summary effect's calculation, confidence interval's calculation and null hypothesis' definition. The calculating process of standard error of summary effect is in the Borenstein's book (2009).

How to Choose the Appropriate Model

Although we can use the homogeneity test to assist us to choose the model, we should base upon the study design whether all of studies have the identical effect size or not. In general, we will test these two models simultaneously by software packages in the research. Therefore, determining the appropriate model mainly depends on your understanding of the study design and supported with statistical test.

### (3) Heterogeneity

In general, we use the test of homogeneity to assess whether the heterogeneity exists or not. The test should be conducted before combing the included studies. The variation among any kinds of studies is heterogeneity (Higgins & Green, 2008). It also means heterogeneity in the true effects of all studies (Borenstein et al., 2009). Without doing this assessment, the process of combining the studies would be wrong. The goal of heterogeneity is aim to assess whether among studies is similar each other. If it does exist, it could have differences in research methodology, sample size, interventions, and/or the level of quality.

There are two common methods to test for heterogeneity. The first method is using Q statistic (chi-square test) and the second method is conducting $I^2$ test ($I^2$ is the ratio of true heterogeneity to total observed variation). Although we want to assess the heterogeneity in the process of meta-analysis, the source of heterogeneity is more important than just assess whether our studies have heterogeneity. The possible sources result from clinical diversity, methodology diversity or statistical heterogeneity (Higgins & Green, 2008). By means of realizing the source we could have, we can consider using random-effects model or other methods to continue the analysis. Besides, the L'abbé plot also can assess the heterogeneity by graphic method (Tsao, 2006). However, these measures only tell us the variation or dispersion of between-studies.

In conclusion, the assessment of heterogeneity is indispensable when conducting the meta-analysis. Taking care of the heterogeneity will help us to increase the reliability of our meta-analysis.

### (4) Other Issues - Bias

In the process of doing meta-analysis, the bias issue is quite important for researchers. While setting the inclusion and exclusion criteria, you have to avoid the latent bias. According to Cochrane handbook, it provides some types of reporting bias which includes publication bias, time lag bias, multiple (duplicated) publication bias, location bias, citation bias, language bias, and outcome reporting bias (Higgins & Green, 2008).

In general, we would prefer to search for the most-cited publication, publication in English, publication published more than once, and so on. These actions could bring about bias emerging. The following is the definitions of bias, respectively.

- Publication bias: because significant findings are more likely accepted to publish, we are more likely to find out significant research findings.
- Time lag bias: some articles are published rapidly, so we are more likely to find these.

- Multiple (duplicate) publication bias: If the same paper is published more than once, we are more likely to find it.
- Location bias: Some databases usually include some journals. If we only search one or two databases, we could not find the papers not included in databases.
- Citation bias: Generally, we are more likely to find the publication cited by others.
- Language bias: due to language limitation, we usually find the papers we can read.
- Outcome reporting bias: Some researchers could only report (selective) some results from their research.

Cochrane Collaboration, which is top1 collaboration of MA in medicine field, much care about this part; other organizations or journals seldom mention this issue. Moreover, CMA also focuses on the "publication bias" in this issue. Therefore, when we use RevMan and CMA, they will have more functions about bias.

## III. MA Software and a Comparison

Extending from the studies of Bax (2007) and Wallace et al. (2009), we have investigated ten MA-related software, including Comprehensive Meta-Analysis (CMA), RevMan, MetaWin, MIX, WEasyMA, Meta-Analyst, Stata, SPSS, SAS, and R. The former six packages are dedicated MA software, the last four general statistical ones. Because of page limitation, please refer to references for detail information. The comparison of MA-related software and packages has four indexing categories: (1) General, (2) Necessary/Must, (3) Second/Advanced, (4) Complementary/Auxiliary. Each category has several factors which are related to MA theories.

An MA Software Comparison/A Meta-Analysis Software Comparison

The factors of category classifications refer from Leon Bax (2007; 2006), Cheng and Wang (2012), Geyskens et al. (2008) and Wallace et al. (2009).

Table 3 Software Comparison

| | Professional Meta-Analysis Software | | | | | | Statistical Software | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CMA v.2 | RevMan 5.0.18 | MetaWin 2.1 | MIX 1.7 | WEasyMA 2.5 | Meta-Analyst Beta 1.0 | Stata 10 | R 2.6 | SAS | SPSS |
| **General**- Free (Noncommercial) | No | Yes | No | Yes | No | Yes | No | Yes | No | No |
| **Necessary/Must** | | | | | | | | | | |
| Fixed effects | x | x | x | x | x | x | x | x | x | x |
| Random effects | x | x | x | x | x | x | x | x | x | x |
| Diagnostic test data | | x | | | | x | x | x | | x |

| | Professional Meta-Analysis Software | | | | | | Statistical Software | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CMA v.2 | RevMan 5.0.18 | MetaWin 2.1 | MIX 1.7 | WEasyMA 2.5 | Meta-Analyst Beta 1.0 | Stata 10 | R 2.6 | SAS | SPSS |
| Forest plot | x | x | x | x | x | x | x | x | | x |
| Funnel plot | x | x | x | x | x | x | x | x | x | x |
| Small study effect/publication bias | x | | x | x | x | | x | x | x | x |
| Heterogeneity (Q, $I^2$, etc.) | x | x | x | x | x | | | | x | |
| Constant continuity correction | x | x | x | x | x | | | | | x |
| **Secondary/advanced** | | | | | | | | | | |
| SROC (Summary ROC approach) | | x | | | | x | x | x | x | x |
| HSROC-bivariate meta-analysis | | | | | | x | x | x | x | |
| Subgroup analysis | x | x | | | x | x | x | x | x | x |
| Meta-regression | x | | x | | | x | x | x | x | x |
| Leave one out sensitivity/ sensitivity analysis | x | | x | | | x | x | x | x | x |
| Point and click plot editing | | | | | | x | x | | | |
| L'Abbe plot | | | | x | x | | | | | |
| Trim and fill plot | x | | | x | | | | | | |
| Cumulative forest plot | x | | x | x | x | x | x | x | | |
| **Complementary/supplementary/seldom used/auxiliary** | | | | | | | | | | |
| Individual study data | x | x | x | x | x | | | | | |
| Bootstrap confidence intervals | | | x | | | | | | | |

We classify two kinds of software, professional one and statistical one. Table 3 shows that for the necessary and complementary function, Professional Meta-Analysis Software perform well than the Statistical Software. In general, these software can conduct the basic meta-analysis well. However, for the advanced level, some MA-related software are not flexible than the Stata and R (be close to programming language).

## 1. MA Applications in Production and Quality Management

The MA applications in Production and Quality Management are still in their early stage of infantry. There are not that much literature and existing works on these related fields. In this article, we will introduce two examples in the following sections. Differentiate with the major applications in medicine or clinical fields, MA studies in Production and Quality Management tend to concentrate on more correlation flows than to investigate the true effects of interventions.

The goal of MA of correlations is to elaborate the distribution of true correlations between independent and dependent variables (Hunter and Schmidt, 2004). Hunter and Schmidt (1990,

p.45) firstly identify 11 artifacts that vary the size of the study correlation to differ from the true correlation. By replicating studies could potentially reduce the effects of artifacts. The details of MA of correlations were shown in Hunter and Schmidt (1990).

The following sections we are going to introduce two classic MA studies in Production and Quality Management: Nair (2006) *Meta-analysis of the relationship between quality management practices and firm performance—implications for quality management theory development* published by Journal of Operations Management; and Mackelprang and Nair (2010) *Relationship between just-in-time manufacturing practices and performance: A meta-analytic investigation*, also published by Journal of Operations Management.

## 1. Quality Management practices and firm performance

Quality management has received highly attention nowadays. Nair (2006) addressed the following questions to see the magnitude of QM practices affect firm performance: 1) Which QM practices are positively correlated with aggregate firm performance? 2) Which QM practices are positively correlated with individual dimensions of performance? 3) Are there other moderating factors that influence the relationship between QM practices and performance? According to research questions, the following hypotheses were developed in the original study (Nair, 2006):

H1. QM practices in a manufacturing plant are positively correlated with aggregate performance.

H2. QM practices in a firm are positively correlated with aggregate performance.

H3. The correlation between QM practices and aggregate performance is influenced by moderating factors.

H4. Individual QM practices are positively correlated with aggregate performance.

H5. The correlation between individual QM practices and aggregate performance is influenced by moderating factors.

H6. Individual QM practices are positively correlated with different performance measures.

H7. The correlation between individual QM practices and different performance measures is influenced by moderating factors.

Literature was searched and was thoroughly examined from the ABI/INFORMS database. By searching through “total quality management” or “quality management” and “performance”, only the studies that using perceptual data and published from 1995 to 2004 were selected. Finally, a total of 23 empirical studies were included in meta-analytic procedure. Each empirical study was firstly examined by revealing its “sample and unit of analysis”, “method”,

"operationalizing TQM practices", "operationalization of performance", and "findings". The aggregate result is presented in Nair's original study.

In the meta-analytical procedure, researcher employed two heuristics which developed by Hunter and Schmidt (1990) to test the hypotheses: *RATIO1* to exam the population correlation and *RATIO2* to exam the existence of moderator. Owing to the page limitation, we won't look into the detail of heuristics and analytical procedure.

As a whole, the results showed a positive correlation between several QM practices and firm performance dimensions. It encouraged practitioners to continue adopting QM practices in their organizations. Furthermore, some moderating effects between variables were also pointed out and it provided directions for future studies to work on. However, limitations like analyzer didn't perform "reporting on transcriptional error", which is one of the artifacts suggested by Hunter and Schmidt, in the study was confessed by the author.

## 2. JIT manufacturing practices and performance

Just-In-Time (JIT) manufacturing has been deeply studied in the area of operations management. JIT practice is considered a boost for an organization to keep competitiveness in global market. However, skeptics also question the adoption of JIT manufacturing. Mackelprang and Nair (2010) tried to investigate the true correlations between JIT practices and performance by conducting a meta-analysis. They addressed some important issues that remain unanswered: 1) Do all JIT practices positively relate to all performance outcomes? 2) Are there JIT practices that have greater impact on various performance measures than others? 3) Which JIT practice to performance links are influenced by moderating factors?

Authors conducted a citation analysis first to obtain samples for MA study. The articles published from 1992 to 2008 which Mehra and Inman (1992) or Sakakibara et al. (1993) were included. Keywords such as 'just-in-time", "JIT'' and ''Lean" were used to advanced search specific literature. Finally, 23 operations management related journals formed a database and 25 empirical studies were selected.

In Mackelprang and Nair's (2010) study, ten JIT practices are considered as variables in MA procedure: setup time reduction, small lot sizes, JIT delivery from suppliers, daily schedule adherence, preventive maintenance, equipment layout, Kanban, JIT link with customers, pull system, and repetitive nature of master schedule. In terms of performance dimensions, quality performance, manufacturing cost, inventory, cycle time, manufacturing flexibility, and delivery performance were considered variables. Besides, for each individual study, JIT-performance correlation, study sample size, and attenuation factor were directly obtained as a code to be entered to MA procedure. Analyzers further, according to the MA of correlation procedures

suggested by Hunter and Schmidt (2004), calculated the error variance, study weight, and corrected correlation in order to do the heuristics. A step-by-step meta-analytic procedure to correct for sampling and measurement error is outlined by Mackelprang and Nair (2010) as shown in Table 4.

Table 4 Meta-analytic procedures to correct for sampling and measurement error (Mackelprang and Nair, 2010)

| Process steps | Input variables | Formula | Purpose |
|---|---|---|---|
| Step 1. Attenuation factor | 1a. Reliability of JIT practices ($\alpha_{xx}$).<br>1b. Reliability of Performance outcome ($\alpha_{yy}$). | $A = (\alpha_{xx})^{1/2} \times (\alpha_{yy})^{1/2}$ | The attenuation factor is used to correct the correlation for measurement error, create the error variance across studies and to weight the studies. |
| Step 2. Correct study correlations | 2a. Attenuation factor ($A$).<br>2b. Study correlations ($r$). | $r' = r/A$ | The corrected correlations are used in calculating RATIO1, which is used to identify significant population correlations. |
| Step 3. Individual study weights | 3a. Study sample size ($N$).<br>3b. Attenuation factor ($A$). | $W_i = N \times A^2$ | The study weight is used to find the average corrected correlations, average error variances and variance of the corrected correlations. |
| Step 4. Corrected study sampling error | 4a. Weighted sample mean correlations ($\bar{r}$).<br>4b. Study sample size ($N$).<br>4c. Attenuation factor ($A$). | $e_i = (1 - \bar{r}^2)^2/(N-1)A^2$ | Each study's corrected sampling error variance is used to calculate the weighted mean sampling error variance across studies. |
| Step 5. Weighted mean sampling error variance | 5a. Study weight ($W_i$).<br>5b. Study error variances ($e_i$). | $\bar{e} = \sum W_i e_i / \sum W_i$ | The weighted mean error variances is used to estimate the population SD. |
| Step 6. Weighted mean corrected correlations | 6a. Study weight ($W_i$).<br>6b. Corrected study correlations ($r'$). | $\bar{r}' = \sum W_i r' / \sum W_i$ | The weighted mean corrected correlations is used to find both the variance of the corrected correlations as well as RATIO1 |
| Step 7. Variance of the corrected correlations | 7a. Study weight ($W_i$).<br>7b. Corrected study correlations ($r'$).<br>7c. Weighted mean corrected correlations ($\bar{r}'$). | $\sigma_r^2 = \sum W_i [r' - \bar{r}']^2 / \sum W_i$ | The variance of the corrected correlations is used to estimate the population SD. |
| Step 8. Estimate the population SD | 8a. Variance of the corrected correlations ($\sigma_r^2$).<br>8b. Mean error variances ($\bar{e}$). | $S_\rho = [(\sigma_r^2) - (\bar{e})]^{1/2}$ | The estimate of the population standard deviation is used to calculate RATIO1. |
| Step 9. Calculate RATIO1 | 9a. Average corrected correlations ($\bar{r}'$).<br>9b. Estimated population standard deviation ($S_\rho$). | $\text{RATIO1} = \bar{r}'/S_\rho$ | RATIO1 values greater than 2 imply that a positive correlation exists between the variables considered. |
| Step 10. Calculate RATIO2 | 12a. Weighted mean sampling error variances ($\bar{e}$).<br>12b. Variance of the corrected correlations ($\sigma_r^2$). | $\text{RATIO2} = \bar{e}/\sigma_r^2$ | RATIO2 values greater than or equal to 0.75 imply that there is only one population correlation and that the relationship is not subject to moderating factors. |
| Step 11. Credibility interval | 11a. Estimated population standard deviation ($S_\rho$).<br>11b. Average corrected correlations ($\bar{r}'$).<br>11c. $Z$-value of desired credibility level ($Z$). | $\text{Credibility Interval} = \bar{r}' \pm Z \times S_\rho$ | The credibility interval returns the endpoints whereby the percentage selected of the values in the correlation distribution are contained. |

As a whole, the results of Mackelprang and Nair's MA investigation support a positive relationship between JIT manufacturing practices and aggregate performance. However, the results also revealed that several of the correlations between JIT practices and performance have yet to be subjected to rigorous empirical investigation because 12% (7 of 60) possible relationships didn't have sufficient data available for analysis. Moreover, the results of this meta-analysis indicate that several JIT practice to performance relationships are subject to moderating factors. It provided directions for future JIT theory building.

## 2. Conclusions and Discussions

This portion includes five parts which cover conclusions, contribution, significance, justification and implication, and limitation and future research.

### 1. Conclusions

For the software part, most of software has not complete functions to provide for the users. Therefore, it mainly depends on customers' needs to choose an appropriate one.

According to the result of applications, they tend to focus on more correlation flows than the standard MA steps. By conducting MA of correlation, Nair (2006) and Mackelprang and Nair (2010) are all conclude that both QM practices and JIT manufacturing practices have positive correlation with aggregated performance. Besides, two studies also pointed out the potential of moderating effects which both provided directions for future studies.

### 2. Contribution

This study helps readers to know the MA applications in production and quality management field and further provides a brief introduction of software selection. We synthesize the basic ideas of meta-analysis to let the beginners systematically realize and learn.

### 3. Significance

The importance of this research is to present the current state of meta-analysis and the state-of-the-art of MA software development. This study also is a primer of introducing MA in varied research fields. We introduce two MA applications and analyze them step by step. Hope to give a brief guide for beginners of how to understand a meta-analysis in production and quality management fields.

### 4. Justification and Implication

This study gives a general analytic process of conducting a MA study. And we further use this process to exam two famous MA studies in production and quality management fields. Two studies are both failed on reporting publication bias. It implies the MA applications in production and quality management are still in their early stage of infantry.

### 5. Limitation and Future research

Limitation

First of all, due to limited contexts available, the theory of meta-analysis and detail software comparison cannot describe too much. Here just provide the core of method. Secondly,

thanks to time limitation, we just compared and analyzed the differences between general steps and the research flows of QM and JIT papers.

Future research

We do not deny the limitations of the present study. In order to obtain more reliable and objective data, future research that extensions can be conducted in at least two directions: a serious search strategy to collect the related field's papers such as production and quality management and furthermore investigate the situation of other fields like biomedical, education and psychology that how mature they are. Moreover, to develop an MA software based on R will be a worthy exploration.

## References

Anema, A., Mills, E., Montaner, J., Brownstein, J.S., Cooper, C. (2008). Efficacy of influenza vaccination in HIV-positive patients: a systematic review and meta-analysis. HIV Medicine, 9, 57-61.

Bax, L., Yu, L.-M., Ikeda, N., & Moons, K. G. M. (2007). A systematic comparison of software dedicated to meta-analysis of causal studies. BMC Medical Research Methodology, 7(1), 40. doi: 10.1186/1471-2288-7-40

Bax, L., Yu, L. M., Ikeda, N., Tsuruta, H., & Moons, K. G. (2006). Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. [Research Support, Non-U.S. Gov't Validation Studies]. BMC Med Res Methodol, 6, 50. doi: 10.1186/1471-2288-6-50

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to Meta-Analysis.

Cheng, H.-Y. L., & Wang, M.-T. (2012). Meta-Analysis: A Literature Review and its Applications in Management Science in Taiwan.

Geyskens, I., Krishnan, R., Steenkamp, J. B. E. M., & Cunha, P. V. (2008). A Review and Evaluation of Meta-Analysis Practices in Management Research. Journal of Management, 35(2), 393-419.

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 5.

Higgins, J. P., & Green, S. (2008). Cochrane Handbook for Systematic Reviews of Interventions.

Hunter, J.E., Schmidt, F.L., 1990. Methods of Meta-Analysis: Correcting Bias in Research Findings, first ed. Sage Publications, Thousand Oaks, California.

Hunter, J.E., Schmidt, F.L., 2004. Methods of Meta-Analysis: Correcting Bias in Research Findings, second ed. Sage Publications, Thousand Oaks, California.

Juang, C.-M. (2011). 臨床醫師如何閱讀統合分析(Meta-analysis)的論文. Taiwan Medical Journal, 54(2), 74-82.

Kavale, K. A. (2001). Meta-Analysis: A Primer. EXCEPTIONALITY, 9(4).

Larsson, S. C., Wolk, A. (2007). Overweight, obesity and risk of liver cancer: a meta-analysis of cohort studies. [Meta-Analysis Research Support, Non-U.S. Gov't]. Br J Cancer, 97(7),

1005-1008.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis, from http://en.wikiversity.org/wiki/Meta-analysis

Mackelprang, A. W., & Nair, A. (2010). Relationship between just-in-time manufacturing practices and performance: A meta-analytic investigation. Journal of Operations Management, 28(4), 283-302.

Marsh, H. W., O'Mara, Alison, Malmberg, Lars-Erik (2008). What is Meta-analysis? [PowerPoint slides], from http://www.esrc.ac.uk/my-esrc/grants/RES-035-25-0045/outputs/read/bbfcc1e3-bb27-490d-b0df-34cad67fb7dd

Meta-analysis. (2012). In Wikipedia. Retrieved May 9, 2012, from http://en.wikipedia.org/wiki/Meta-analysis#History

Nair, A., 2006. Meta-analysis of the relationship between quality management practices and firm performance—implications for quality management theory development. Journal of Operations Management 24 (6), 948–975.

Normand, S.-L. T. (1999). Tutorial in Biostatistics Meta-Analysis: Formulating, Evaluating, Combining, and Reporting. STATISTICS IN MEDICINE, 18.

Rosenthal, R., & DiMatteo, M. R. (2001). META-ANALYSIS: Recent Developments in Quantitative Methods for Literature Reviews. Annual review of psychology, 52(1).

Tsao, L.-Y. (2006). Meta-analysis (統合分析). The official bulletin of the Society of Neonatology, Taiwan, 15(3).

van Mierlo, L. A. J., Arends, L. R., Streppel, M. T., Zeegers, M. P. A., Kok, F. J., Grobbee, D. E., Geleijnse, J. M. (2006). Blood pressure response to calcium supplementation: a meta-analysis of randomized controlled trials. Journal of Human Hypertension, 20(8), 571-580.

Wallace, B. C., Schmid, C. H., Lau, J., & Trikalinos, T. A. (2009). Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. BMC Medical Research Methodology, 9(1), 80.