MODELLING REPEATED MEASURES DATA IN META ANALYSIS : AN ALTERNATIVE APPROACH

Nik Ruzni Nik Idris

Department Of Computational And Theoretical Sciences
Kulliyyah Of Science, International Islamic University Malaysia
Bandar Indera Mahkota, Kuantan 25200
Pahang, Malaysia
Email : ruzni@iiu.edu.my

Abstract :

A repeated measures design is common in many research areas such as medical and clinical trials, education and psychology. In Meta analysis , where data are typically available at aggregate level, the analysis of repeated measures data is more difficult. One of the limitations of current approaches is in their reliance on the measures at only one or two time points, which involved a considerable loss of data, and do not reflect the trend over time. Another limitation is in estimating the correlation between observations at successive time points. Presently there is no single approach that could address both issues satisfactorily. In this paper, a simulation study is used to develop an alternative approach for meta-analysis based on studies from repeated measures designs which allows utilization of information at all time points. The method uses regression coefficients, estimated from each study, to obtain the study specific estimates of treatment effect. Two approaches of obtaining the overall estimates were considered, namely, separate Meta analyses using the Inverse Variance Weighted method for each coefficient, and using the Multiple Response Model on these regression coefficients. Both approaches generated fixed effects estimates which are in a good agreement when compared to those based on the individual level data. The Multiple Response Model is better as it allows estimates of level 3 random parameters, while the separate Meta analysis only estimates the fixed parameters.

1.0   INTRODUCTION

One of the area in meta analysis where information in the data is not fully utilised is when dealing with repeated measurement data [1]. A repeated measures design occurs when measurements are taken repeatedly over time from the same subject. This is a common design in many research area such as psychology and medical or clinical research. As measurements are repeatedly made from the same individual, the analysis should take into account the serially correlated error terms, in order to distinguish whether the differences between measures in different groups are due to the effect of the treatments or due to the variation between individual studies. Presently there are a number of statistical methods used to analyse repeated measures data at patient level in meta-analysis. The work on data at aggregate level is rather limited although this type of data is very often encountered in meta analysis [2] . There are two main issues in the analysis of aggregate level repeated measures data in meta analysis. The first issue is in estimating the correlation between observations at successive time points, as it is almost never reported in the individual studies. If individual patient data are available, the correlation may be estimated from the data. Otherwise external information such as information from a different study, which may not be as accurate, has to be used. The second issue involves the difficulty in the analysis of the overall trends in the measurements which takes into account all the time points.

Most current meta analysis methodologies do not take into account the information reflected over time, but considers only the differences between two time points of interest. This method involves a considerable loss of data and does not address the issue of any trend in the responses [3]. The method is a particularly popular method in clinical research where the main interest is on the effect of a treatment after a certain time period, and not particularly on the trend in the response over time. For this approach a loss of information will also occur when a study which does not report measurements at the same time points as the other studies has to be excluded from the analysis. Another common approach in dealing with this type of data is to look at the average difference at two time points. In this case, the effect size is defined as the

average difference between pre and post treatment scores divided by the pre-treatment standard deviation. This method is used in many research area where different response variables are used. Meta analysis based on repeated measures data which involve an average effect size have been discussed in a number of articles [3,4,5,6].

The limitation of the approaches discussed above is in their reliance on the measures at only one or two time points. While the first approach involved a considerable loss of data, the standardized mean change approach may not reflect the amount of time between the pre and post testing. Two studies may show the same amount of change, but if the time between testing sessions differs for the studies, the interpretation of these two changes may differ. This however can sometimes be adjusted for by meta regression with the time between pre and post test as an explanatory variable.

In this paper we proposed a method of estimating meta analysis parameters from aggregate level data by utilising the regression coefficients. The regression coefficients are estimated using the information from all time points across the data. The procedure involves estimation of the regression parameters from aggregate-level data from all the studies included in the meta-analysis. These estimates of the regression parameters are then integrated to obtain the overall estimates based on all studies.

2.0   MULTI–LEVEL MODEL FOR REPEATED MEASURES DATA IN META-ANALYSIS

Repeated measures data may be viewed as hierarchical data where the analysis can be carried out within a multilevel framework [7]. In meta analysis, a 3-level hierarchy can be established for repeated measures data with measurement repetitions as level 1 units, nested within patients at level 2 units which is nested within study at level 3.

Suppose there are k = 1, 2, .., N studies with j = 1, 2,..,n_k patients in each study. t_k measurements are taken from each patient, with i = 1, 2,..,t_k. Let y_{ijk} be the ith measurement of the jth patient from study k. A random effect model for this repeated measurements can be written

$$y_{ijk} = \beta_{0k} + \beta_{1k} x_{jk} + \beta_2 T_{ijk} + \beta_3 x_{jk} T_{ijk} + \epsilon_{ijk} \tag{1}$$

β_{0k} represents the overall mean effect which varies across studies, x_{jk} is the dummy covariate for treatment which is assumed to take two values, namely, 0 for control and 1 for the treatment, with β_{1k} as the corresponding random coefficient treatment effect. T_{ijk} is the time variable with the corresponding fixed coefficient β_2, and $x_{ij} T_{ijk}$ represents the interaction between time and treatments. Although in this case it is assumed to have a fixed effect, the time variable need not necessarily be the same in all studies as in practice, the number and location of the time measurements may vary across studies or between patients. Similarly the $x_{jk} T_{ijk}$, representing the interaction between treatment over time, is assumed to have a fixed effect (β_3) although it may not be necessarily so in practice; ϵ_{ijk} is the random error term.

We can write the random coefficients

$$\beta_{0k} = \beta_0 + \nu_{0k}$$
$$\beta_{1k} = \beta_1 + \nu_{1k}$$

with $(\nu_{0k}, \nu_{1k})^T \sim MVN[0, \Sigma_\nu]$, where

$$\Sigma_\nu = \left[\begin{array}{cc} \sigma_{\nu_0}^2 & \sigma_{\nu_0 \nu_1} \\ \sigma_{\nu_1 \nu_0} & \sigma_{\nu_1}^2 \end{array} \right]$$

and $[\epsilon_{1jk}, \ldots, \epsilon_{tjk}]^T \sim MVN(0, \Sigma_\epsilon)$ where these error terms are assumed to be first order autoregressive.

When data are summarized over the patients, the structure of the hierarchy level and random error terms will change. This will have direct effect on the variance and covariance of the mean responses.

It can be shown that the hierarchy has now changed from time, i nested within patients j, nested within study k into time i, nested within treatment h, nested within study k .[8].

At aggregate level, the three level model (1) now becomes

$$\bar{y}_{ihk} = \beta_0 + \beta_1 x_{ihk} + \beta_2 T_{ihk} + \beta_3 x_{jk} T_{ijk} + \nu_{0k} + \nu_{1k} x_{ihk} + \frac{\epsilon'_{ihk}}{n_{ihk}} \tag{2}$$

where $x_{ihk} = 0$ if h = 0 , and $x_{ihk} = 1$ if h=1, and the time points i at level 1 is nested within the treatment group h at level 2 nested within study at level 3. The hierarchy structure for the summary of repeated measures data has changed as the time points are now contained within the treatment group, and the level 1 random error terms are now given by ϵ'_{ihk}/n_{ihk}, with $\epsilon'_{ihk} = \sum_j \epsilon_{ijk}$.

The variance of the mean response at time i, $V(\bar{y}_{ihk})$, is given by

$$V(\bar{y}_{ihk}) = V(\beta_0 + \beta_1 x_{ihk} + \beta_2 T_{ihk} + \nu_{0k} + \nu_{1k} x_{ihk} + \epsilon'_{ihk}/n_{ihk}) \tag{3}$$

$(\nu_{0k}, \nu_{1k})^T \sim MVN[0, \Sigma_\nu]$ where $\qquad \Sigma_\nu = \begin{bmatrix} \sigma^2_{\nu_0} & \sigma_{\nu_0 \nu_1} \\ \sigma_{\nu_1 \nu_0} & \sigma^2_{\nu_1} \end{bmatrix}$

We can show that the variance, given by

$$\begin{aligned} V(\bar{y}_{ihk}) &= (\sigma^2_{\nu 0} + \sigma^2_{\nu 1} x_{ihk} + 2\sigma_{\nu 01} x_{ihk}) + V(\sum_{j=1}^{n_{ihk}} \epsilon_{ijkh}/n_{ihk}) \\ &= \sigma^2_{\nu 0} + \sigma^2_{\nu 1} x_{ihk} + 2\sigma_{\nu 01} x_{ihk} + \frac{\sigma^2_\epsilon}{n_{ihk}} \end{aligned} \tag{4}$$

depends on both the between study variance, within study variance and number of patients at the timepoints i treatment h and study k.

The covariance between two mean responses, when the number of patients assumed to remain constant over the time points, is given by Idris (2006) [8].

$$\begin{aligned} cov(\bar{y}_{1hk}, \bar{y}_{2hk}) &= \frac{1}{n^2_{hk}}(n_{hk}\sigma_{\epsilon 12}) \\ &= \frac{\sigma_{\epsilon 12}}{n_{hk}} \end{aligned}$$

This implies that if the number of patients are constant over the timepoints, the covariance between two successive mean responses given by $\frac{\sigma_{\epsilon 12}}{n_{ihk}}$ is a function of $\sigma_{\epsilon 12}$ the covariance between two successive responses, and n_{ihk}, the number of patients at time point i, under treatment h and study k.

In practice some patients may drop out over the time period, resulting in a reducing number of n, thus increasing the variance of the level 1 error terms over time. In this case, the covariance is given by

$$cov(\bar{y}_{1hk}, \bar{y}_{2hk}) = \frac{\sigma_{\epsilon 12}(n_{2hk})}{n_{1hk}n_{2hk}}$$

where n_{1hk} is the larger number of patients of the two time points. This implies that in this particular case, even if the number of patients are reducing over the time period, for instance, $n_{1jk} > n_{2jk}$, the covariance between successive mean responses, $cov(\bar{y}_{ihk}, \bar{y}_{(i+1)hk})$ is not much affected.

3.0   MATERIAL AND METHOD

The data used in this study is based upon simulation of hypothetical meta analyses. The procedure for data generation and parameters estimation are detailed below ;

3.1   Individual level Data

i)   Generation of data

An individual level data was simulated based on model given by Equation (1). The data is generated with the following specifications
- Number of studies included, N = 30
- Number of patients in each study, n = 10, 20, 60 (at 37%, 47%, 16%, respectively)
- Number of equally spaced time points = 4
- Number of treatment arms

As been done in other studies [9], the coefficient parameters are assigned with the following arbitrarily chosen values

- $\beta_0 \sim N(0, 2^2)$ (Intercept)

- $\beta_1 \sim N(2, 1^2)$ (treatment)

- $\beta_2 = 2$ (time)

- $\beta_3 = 1$ (time by treatment)

The random effect at level 2 (patient level) is specified as normally distributed with mean 0 and variance 1, while the error terms in all three data are autoregressively correlated

$$\left[\epsilon_{1jk}\epsilon_{2jk}\epsilon_{3jk}\epsilon_{4jk}\right]^T \sim MVN(0, \Sigma_\epsilon)$$

An assigned value of the correlation coefficient of $\rho = 0.7$ and level 1 variance $\sigma_\epsilon^2 = 1$ were given.

With this specification, 5,520 responses were generated from 1,380 patients. Half of the patients in each study are on control, and the other half are treated.

For instance, the data generated for the Study 1, with n = 20 is

no	study	pat	time	treatment	y
[1,]	1	1	0	0	−1.99728
[2,]	1	1	1	0	−1.03282
[3,]	1	1	2	0	−0.40679
[4,]	1	1	3	0	1.28884
[5,]	1	2	0	1	−1.17423
[6,]	1	2	1	1	2.02377
[7,]	1	2	2	1	4.83525
[8,]	1	2	3	1	6.06153
[9,]	1	3	0	0	−0.77352
.	.		.		.
.	.		.		.
.	.		.		.
.	.		.		.
[54,]	1	14	1	1	0.95060
[55,]	1	14	2	1	4.05916
[56,]	1	14	3	1	5.98171
[57,]	1	15	0	0	−1.86537
[58,]	1	15	1	0	0.25889
[59,]	1	15	2	0	1.53976
[60,]	1	15	3	0	2.67248
[61,]	1	16	0	1	0.80140
[62,]	1	16	1	1	2.42819
[63,]	1	16	2	1	4.43984
[64,]	1	16	3	1	6.94998
[65,]	1	17	0	0	−1.56506
[66,]	1	17	1	0	−0.31588
[67,]	1	17	2	0	−0.18731
[68,]	1	17	3	0	0.44211
[69,]	1	18	0	1	−0.26411
[70,]	1	18	1	1	1.50568
[71,]	1	18	2	1	2.47900
[72,]	1	18	3	1	5.08469
[73,]	1	19	0	0	−1.31831
[74,]	1	19	1	0	−0.24813
[75,]	1	19	2	0	1.93332
[76,]	1	19	3	0	3.94674
[77,]	1	20	0	1	−2.35337
[78,]	1	20	1	1	0.34675
[79,]	1	20	2	1	2.07463
[80,]	1	20	3	1	4.53804

ii)   Estimation of parameters

Next, a three level model given by Equation (1) were fitted into the data the software using MLwiN [7] with first order autoregressive error term to obtain estimates of the parameters. The procedure for fitting is discussed in detail by Goldstein And Yang (2000) [10]

3.2 Aggregate level Data

i) Summarizing the data into aggregate level

To summarize the data over patients, we let the mean, $\bar{y}_{ihk}$ and standard deviation, s_{ihk}, for a study k, at the time i, under treatment h, of the responses. If there are t time points and h treatment arms, then for each study k, there are (t x h) mean responses. For instance, in the meta analysis that was generated, there are N=30 studies, two treatment arms, h=2, and four time points, t=4, thus for each study there 8 responses and the corresponding standard deviations. The summary of patient level data for Study 1 is given below :

no	study	time	treat	mean y
1	1	0	0	-0.76932
2	1	1	0	0.29151
3	1	2	0	1.20096
4	1	3	0	2.60187
5	1	0	1	-1.39224
6	1	1	1	0.50351
7	1	2	1	2.67885
8	1	3	1	4.96498

ii) Estimation of Regression Coefficients

To get the estimates of regression coefficient for this summary level data, each study is fitted with a weighted linear regression on the mean response $\bar{y}_{ihk}$, taking the inverse of the standard error of the responses as the weights.

This model was fitted,

$$\bar{y}_{ihk} = \beta_{0k} + \beta_{1k}x_{hk} + \beta_{2k}T_{ihk} + \beta_3 x_{jk}T_{ijk}$$

and extract the estimates of β_{0k}, β_{1k}, β_{2k} and β_{3k} and the corresponding standard error. β_{0k} is the overall mean response in study k, β_{1k} is the estimate of treatment effect, β_{2k} are the estimate of time effect in study k and β_{3k} is the coefficient for the interaction term.

The above procedure generated 30 sets of estimates of regression parameters comprising the estimates of β_{0k}, β_{1k}, β_{2k} and β_{3k} and their corresponding standard error. The estimates for study 1 is presented below for illustrations. We can also extract the correlation matrix between β_{0k}, β_{1k} and β_{2k}.

Study	Regression Parameters	Estimates
1	Const - $\beta_0(SE)$	-0.848 (0.0490)
1	Treat - $\beta_1(SE)$	2.109 (0.0702)
1	Time - $\beta_2(SE)$	2.513 (0.02839)
1	Time x Treat - $\beta_3(SE)$	0.882 (0.0388)

Table 1 :   Estimates of regression coefficients from Study 1

iii) Estimation of Meta analysis parameters

The overall estimates based on this aggregate level data can now be estimated from these coefficients using the weighted inverse variance method, or by fitting a Multiple Response model.

Weighted Inverse variance method
This is a standard method used in meta analysis and is applied separately to each set of estimates and their corresponding variances. The meta analysis estimates computed here are based on a random effect model.

Multiple Response model
Fitting of a multiple response model to this data as describe in [8]. using the MLwIN software. In this data, the coefficient estimates of β_{0k}, β_{1k}, β_{2k} and β_{3k} are treated as multiple response variable, and the data are restructured such that the lowest level (level 1) is the response variable, nested within treatment at level 2 and nested within study at level 3.

4.0   RESULTS

In Table 1 three sets of estimates are presented for comparison, namely estimates from the individual level data, the estimates from aggregate level data obtained by performing a separate meta analysis on the regression coefficients and those from fitting a Multiple Response model to the regression coefficients.

If repeated measure data are available at individual level for all studies, all fixed and random parameter estimates, including an estimate of the within study correlation ρ may be obtained using the MLwiN software.

For data at aggregate level, generally both Multiple Response meta analysis and separate meta analysis using the standard Inverse Variance Method (IVM) are capable of providing good estimates of the fixed effect parameters. The treatment effect estimate, β_1 from separate meta analysis is 2.011 ± 0.223, and those from the Multiple response model is 1.996 ± 0.237, compared to 1.990 ± 0.214 from the individual level data. Additionally, the estimates of time effect, β_2, are also in good agreement at 2.021(0.044) (against 2.014 ± 0.015 at patient level) for both approaches, and the time by treatment interaction, β_3, at 0.990 ± 0.064 and 0.987 ± 0.064 for IVM and MRM, respectively, compared to 0.995 ± 0.021 at patient level.

	INDIVIDUAL LEVEL DATA	SUMMARY LEVEL DATA	
		Multiple Response method	Inverse variance method
Fixed Par			
Cons (SE)	0.348 (0.414)	0.337 (0.420)	0.348 (0.432)
Treat (SE)	1.990(0.214)	1.996(0.237)	2.011(0.223)
Time (SE)	2.014(0.015)	2.021(0.044)	2.021(0.044)
Time x Treat(SE)	0.995(0.021)	0.987(0.064)	0.990(0.064)
Random Par			
Level 3 (study)			
Cons (SE)	5.031(1.325)	5.083(1.333)	
Treat (SE)	1.148(0.347)	1.240(0.363)	
Cons/treat (SE)	0.278(0.481)	0.229(0.493)	
Level 2 (pat)			
Cons	0.970(0.113)		
Level 1 (time)			
Cons	1.027(0.101)		
P	0.336(0.715)		

Table 3 : Estimates of meta analysis parameters at individual level and aggregate level data

For random parameters, the Multiple Response meta analysis have been able to provide estimates of random parameters at level 3, but fail to give the level 2 (between patients) and level 1 (between time points) estimates. These estimates are again relatively close those obtained from the individual level data.

A separate meta analysis, as expected, would not be able to provide the random parameter estimates as the integration of the coefficients are done separately for each coefficients. Both approaches fail to provide the level 2 and level 1 random parameters.

5.0   DISCUSSION

We have shown that it is possible to obtain meta analysis estimates that are close to those from individual level data from the regression coefficients. The overall estimates may be computed from these regression coefficients through separate meta-analysis or by fitting a Multiple response model.

Both meta analysis and multiple response model performed on the regression coefficients provide a very good fixed effect estimates when compared with the estimates produced based on data at individual level. Performing separate meta analysis using the IVM of each coefficient is an easier option as it is quite straightforward. However it only provides the estimates of the fixed parameters. The Multiple Responses model may be preferable as it allows additional level 3 random parameter estimates. The application of the Multiple response model however requires initial assumptions on the level 2 covariance matrix, and involves a more complex statistical basis. Nevertheless, it is still considered as a good alternative to the separate meta analysis.

The regression coefficient estimates are expected to be more precise if the individual studies have a relatively large sample size as the estimates of the regression coefficients will have smaller standard error. This will in turn, provide better overall estimates.

One of the limitations of this method is that it requires two stages of parameter estimates, namely, the estimates of the regression coefficient, and then the estimates of the overall parameters. Additionally, the assumptions of independence between the coefficients, and the linearity assumptions in the mean response over the time period are required in obtaining the estimates of the regression coefficients. Additional assumption of $\rho = 0$ was also made in estimating the regression coefficients.

6.0   CONCLUSION

In general this approach can be considered as an alternative to the existing approaches in estimating meta analysis estimates based on summary level data. The proposed method has the advantages of utilizing the data at all time points, and thus provides an overall effect of the treatment against the current approaches which uses only two time point which provides effects over these two period only. Additionally the method allows estimates of the effect of time as well as the interaction of of treatment over time.

Another advantage of this method is that it may be applied even when a small number of studies are available for meta analysis which is a common problem for meta analysis data. The method produces better estimates if the number of patients in each study is relatively large (>30), as these studies would be able to provide more precise regression coefficients. The approach however fail to estimate the within study correlation, one of main issues in modeling aggregate level repeated measure data in meta-analysis, the even multiple response model was used

.

References

[1]   Whitehead, A. (2002) Meta-analysis of controlled clinical trials. London : John Wiley & Sons

[2]   Sutton, A.J., Keith, R.A., David, R.J., Trevor, A.S., Fujian, S. (2002). Methods for meta-analysis in medical research. New York : John Wiley.

[3]   Becker B.J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematics and Statistical Psychology*, **41**, 257-278.

[4]   Dunlap, W.P., Cortina, J.M., Vaslow, J.B., Burke, M.J. (1996). Meta analysis of experiments with matched group or repeated measures designs.*Psychological Method*},**1**, 170-177.

[5]   Scott, B.M. (2000). Distribution of the standardized mean change effect size for meta analysis. *British Journl of Methametical and Statistical Psycholog*}, **53**, 17-29.

[6]   Morris, B.S., DeShon, R.P. (2002) Combining effect size in meta analysis with repeated measures and independent-groups designs. *Psychological Methods*,**7**(1)}, 105-125.

[7]   Goldstein, H. (2003). *Multilevel statistical models, Third Edition*. London : Arnold.

[8]   Idris, N.R.N (2006). Estimating meta analysis parameters in non standard data. PhD Thesis. Glasgow: University of Strathclyde.

[9]   Higgins, J.P.T., Whitehead, A.,Turner, R.M., Omar, R.Z., Thompson, S.G. (2001). Meta-analysis of continuous outcome data from individual patients. *Statistics In Medicine*, **20**, 2219-2241.

[10]   Goldstein, H.,Yang, M. (2000) Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistics*, **49**(3), 399-412.