World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:12, No:11, 2018

Bayesian Meta-Analysis to Account for Heterogeneity in Studies Relating Life Events to Disease

Elizabeth Stojanovski

***Abstract*—**Associations between life events and various forms of cancers have been identified. The purpose of a recent random-effects meta-analysis was to identify studies that examined the association between adverse events associated with changes to financial status including decreased income and breast cancer risk. The same association was studied in four separate studies which displayed traits that were not consistent between studies such as the study design, location, and time frame. It was of interest to pool information from various studies to help identify characteristics that differentiated study results. Two random-effects Bayesian meta-analysis models are proposed to combine the reported estimates of the described studies. The proposed models allow major sources of variation to be taken into account, including study level characteristics, between study variance and within study variance, and illustrate the ease with which uncertainty can be incorporated using a hierarchical Bayesian modelling approach.

***Keywords*—**Random-effects, meta-analysis, Bayesian, variation.

I. Introduction

A relationship between life events and physical illness has previously been recognised [1]. It has been speculated that the mechanism by which life stresses affect physical health is via their impact on emotional or mental health. Much of the research in this area, however, has been within the cardiovascular field, where the association between life stresses and both hypertension and myocardial infarction has been documented extensively [2]. Associations between life stresses and various forms of cancers have also been identified, including gastric, lung, and breast cancers [1]. Information about the associations between particular types of cancer with life events, which are considered particular types of life stresses, is sparse and in many cases conflicting [3]. Some of these conflicts may be attributed to differences in study design characteristics and the differing confounding factors that are considered between studies.

The purpose of a recent random-effects meta-analysis by Duijts et al. [4] was to identify studies that examined the association between adverse events and the risk of breast cancer to establish the relationship for various types of life events. These particular events were examined: death of partner or relative or friend, stressful events, personal or non-personal health problems, change in marital or financial status and changes to environment. Only these life events showed a statistically significant effect on breast cancer: death of relative or friend (OR=1.35, 95% CI: 1.09-1.68), death of partner (OR=1.37, 95% CI: 1.10-1.71), and stressful events (OR=1.77, 95% CI: 1.31-2.40).

One of the groups of life events assessed was those associated with changes to financial status, which produced an overall non-statistically significant association with breast cancer risk (SOR=0.90, 95% CI: 0.54-1.50) [5]. The studies, however, differed substantially in terms of several factors related to the study design and study characteristics such as the location in which the study was conducted, time frame over which the study was carried out, and the sample size. The aim of the study was to collate information from several studies that addressed a similar aim, with the intention to identify the characteristics that were able to be attributed to the differential study findings.

Following the method of DuMouchel [6], two random-effects Bayesian meta-analysis models are constructed and implemented for the present study to combine the reported estimates of the four studies described that assess the relationship between breast cancer and life events related to financial status. The proposed model allows three major sources of variation to be taken into account. These include study level characteristics, between study and within study variance. The sensitivity of the overall results to various study characteristics is also investigated.

II. Methods

A summary of study-specific characteristics of the studies considered for the present meta-analyses is provided in Tables I and II. The study-specific estimates of the association between breast cancer and life events related to financial status are presented in Table III. The study estimates are presented graphically in Fig. 1.

In the first model proposed (Model 1), presented in Fig. 2, the observed log odds ratios Y_i, are assumed random samples from study-specific true log odds ratios θ_i, which are themselves assumed distributed around an overall log odds ratio μ and the log odds ratios, Y_i, θ_i and μ are assumed to be normally distributed. For this model, n represents the number of studies (here four), W_Y is the observed precision matrix (the inverse of the observed variance-covariance matrix) describing within-study variation, and W_θ is the prior precision matrix describing between-study variation. The two parameters σ_Y^2 and σ_θ^2 indicate the degree of uncertainty around these precision matrices, as expressed through their respective degrees of freedom ν_Y and ν_θ, respectively. Here, ν_Y is set to the average number of cases in the four studies, and ν_θ is set to the number of studies. Following statistical theory,

E. Stojanovski is with the University of Newcastle, NSW 2308 Australia, (phone: +61-2-4921-5346; fax: +61-2-4921-6898; e-mail: Elizabeth.stojanovski @newcastle.edu.au).

since $\nu S^2/\sigma^2 \sim \chi_\nu^2$, chi-squared distributions are imposed on σ_Y^2 and σ_θ^2 which, when divided by their degrees of freedom, have expected value equal to one. In this way, these variables affect the spread rather than the location of the distributions around W_Y and W_θ, and the studies are assumed to provide independent information.

The second model (Model 2) assumes one or more additional hierarchical levels between the study-specific parameters and the overall distribution and is presented in Fig. 3. Thus, Model 2 can accommodate partial exchangeability between the studies, acknowledging that some studies are more similar due to common designs, locations, and so on.

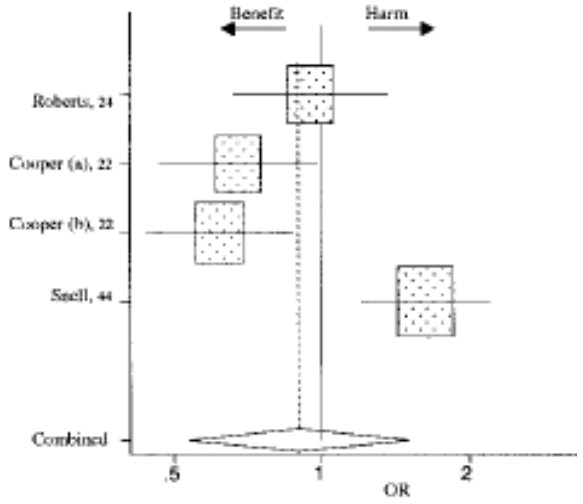

Fig. 1 Study specific summary odds ratios (diamonds) for life events related to changes in financial status. Vertical lines and width of diamonds indicate the 95% CI. Solid and dashed reference lines indicate no effect and summary odds ratios, respectively. Benefit denotes decreased effect and harm denotes increased effect

The study characteristics that were considered using Model 2 are defined as follows: C1: Study design: case control or cohort; C2: Study location: USA or UK and C3: whether there was correction for confounding (yes or no). Thus, for each of these three situations, each true study-specific log odds ratio arises from one of two subgroups with subgroup mean log odds ratio θ_j (j=1,2). Independence is also assumed between studies with this model, so that the precision matrices are all diagonal. Within the model, n represents the number of studies (here four), W_Y is the observed precision matrix describing within-study variation, and W_θ is the prior precision matrix describing between-study variation. The two parameters σ_Y^2 and σ_θ^2 indicate the degree of uncertainty around these precision matrices, as expressed through their respective degrees of freedom ν_Y and ν_θ respectively. Here, ν_Y is set to the average number of cases in the four studies, and ν_θ is set to the number of studies. Following statistical theory, since $\nu S^2/\sigma^2 \sim \chi_\nu^2$, chi-squared distributions are imposed on σ_Y^2 and σ_θ^2, which, when divided by their degrees of freedom, have expected value equal to one. In this way, these variables affect the spread rather than the location of the distributions around W_Y and W_θ. The number of subgroups is represented by m and ξ_j represents the log odds ratio of subgroup j with corresponding precision parameters σ_θ^2 and ν_ξ, and prior between-subgroup precision matrix W_ξ.

$$
\begin{aligned}
&\textit{Model: } Y_i \sim N(\theta_i, \sigma_Y^2 W_Y) \quad i=1,..,n\\
&\theta_i \sim N(\mu, \sigma_\theta^2 W_\theta)\\
&\mu \sim N(0, \Delta \to \infty)\\
&\sigma_\Psi^2 \sim \chi_{\nu_\Psi}^2 / \nu_\Psi\\
&\sigma_\theta^2 \sim \chi_{\nu_\theta}^2 / \nu_\theta
\end{aligned}
$$

Fig. 2 Random-Effects Bayesian Model 1

The prior precision matrices are defined to have diagonal entries equal to one, reflecting little information and therefore strong uncertainty about between-study variation.

Bayesian analysis involves integration over potentially high dimensional probability distributions of model parameters to enable inferences about model parameters [7]. MCMC may be used instead to draw samples from the required distributions and then form sample averages to approximate expectations.

Model:

$$
\begin{aligned}
&Y_i \sim N(\theta_i, \sigma_Y^2 W_Y) && i=1,..,n\\
&\theta_i \sim N(\xi_j, \sigma_\theta^2 W_\theta) && j=1,..,m\\
&\xi_j \sim N(\mu, \sigma_\xi^2 W_\xi) && j=1,..,m\\
&\mu \sim N(0, D \to \infty)\\
&\sigma_Y^2 \sim \chi_{\nu_Y}^2 / \nu_Y\\
&\sigma_\xi^2 \sim \chi_{\nu_\xi}^2 / \nu_\xi\\
&\sigma_\theta^2 \sim \chi_{\nu_\theta}^2 / \nu_\theta
\end{aligned}
$$

Fig. 3 Random-Effects Bayesian Model 2

The analyses was undertaken in WinBUGS [8], with a burn-in of 100,000 iterations, which are excluded, and a collection period of 100,000 iterations to estimate the odds of developing breast cancer as a result of having experienced life events related to financial status; and initial values were set at the maximum likelihood values. These were more than sufficient to confirm convergence, as indicated by the diagnostics within WinBUGS.

Summaries of the posterior distributions were assessed graphically using kernel density plots and are presented numeri-cally by calculating summary statistics such as the mean, variance, and quantiles of the sample. WinBUGS trace and history functions offer serial plots of the actual sequence of simulated values to diagnose convergence. The full empirical distribution function is used for hypothesis testing.

III. Results

Trace plots of the MCMC iterations of the simulated variables θ, τ_θ, μ, α, ν, δ, τ_μ for the first random-effects model

(Model 1) showed all parameter estimates as being quite stable, although the overall precision had some large values consistent with its vague Gamma prior. The plots of the posterior densities showed the simulated posterior distribution for the parameters. Estimates of the posterior mean, standard deviation and 95% credible interval for the study-specific log odds ratios θ_i, i=1,...,4, and the overall mean log odds ratio, μ, for Model 1 are presented in Table IV. The overall posterior mean log odds ratio point estimate is -0.0363 and the 95% credible interval for the value of the mean log odds ratio is -0.0144 to 0.4877. These can be further converted to the odds ratios and associated 95% confidence intervals generated by exponentiating all figures since the log transformation is monotonic.

The second random-effects Bayesian model (Model 2) was also employed to inspect the impact of various study design characteristics that differed between the four studies under examination. Trace plots and posterior density plots for the model parameters were inspected for stability and conformity to the anticipated distributions. In all cases, these characteristics were confirmed.

Estimates of the posterior mean, standard deviation and 95% credible interval for θ, ξ, μ, under each of the three alternatives C1, C2, C3 are given in Table V. These results suggest that the overall odds ratio from the three case control studies is greater than unity and that from the cohort study is less than unity, although both estimates have 95% credible intervals that span unity. Similarly, those studies conducted in the USA have an overall odds ratio that is greater than unity, whereas those conducted in the UK have an overall odds ratio that is less than unity, but again the two credible intervals both include unity. Finally, the overall odds ratio for the three studies that controlled for confounding is greater than unity compared to a reduced odds ratio for the study that did not control for such issues. The overall odds ratios for the three analyses are not substantially different in light of the very wide credible intervals which are a consequence of the disparate study estimates and vague priors.

TABLE I
DESCRIPTION OF STUDIES CONSIDERED FOR THE META-ANALYSIS

Study	First Author	Time frame	Year of publication	Country	Design	Exposition
1	Roberts	5 years	1996	USA	Retrospective case-control	Questionnaire
2	Cooper	2 years	1989	UK	Prospective case-control	Questionnaire
3	Cooper	2 years	1993	UK	Limited prospective Cohort	Questionnaire
4	Snell	5 years	1971	USA	Retrospective case-control	Interview

TABLE II
DESCRIPTION OF STUDIES CONSIDERED FOR THE META-ANALYSIS (CONTINUED)

Study	No. cases	Source of cases	Age of cases	No. controls	Source of cohort	Correction for confounding
1	258	Population	64.8	614	Population	Yes
2	171	Suspicion	55	1992	Hospital	Yes
3	171	Suspicion	55	727	Suspicion	Yes
4	352	Hospital	55.5	670	Hospital	No

TABLE III
STUDY-SPECIFIC ESTIMATES USED IN THE META-ANALYSIS

Study	Odds Ratio	95% CI	Log odds ratio	Precision
1	0.96	0.66-1.41	-0.0408	27.77
2	0.65	0.44-0.96	-0.4308	26.29
3	0.59	0.41-0.85	-0.5276	30.10
4	1.73	1.26-2.36	0.5481	40.63

TABLE IV
SUMMARY STATISTICS FOR THE POSTERIOR MEAN LOG ODDS RATIOS θ AND μ

Log odds ratio	Mean	S.D.	2.5%	97.5%
θ_1	-0.0414	0.1878	-0.4113	0.3287
θ_2	-0.4057	0.1943	-0.7858	-0.0217
θ_3	-0.4998	0.1822	-0.8564	-0.1403
θ_4	0.5237	0.1580	0.2112	0.8326
μ	-0.0363	0.2606	-0.0144	0.4877

IV. DISCUSSION

By allowing for differences in study design, the present analysis supports the findings from the study by Duijts et al. [4], where it was concluded that life events relating to changes in financial status are not statistically significantly related to breast cancer. The disparate nature of the results from these four studies may arise due to differences in study design, location, method of analysis, among other factors. These differences can be acknowledged and explored through the addition of hierarchies to the meta-analysis model, as demonstrated using the model presented. Due to the small number of studies involved, the analyses under this model are intended to be indicative rather than substantive.

Unfortunately, there was insufficient information to further investigate these suggested differences in odds ratios associated with different study design characteristics, or to identify whether there were interactions between these study characteristics. However, the analyses did serve to demonstrate the application of a random-effects Bayesian meta-analysis model by combining results from studies while accommodating partial exchangeability between studies, acknowledging that some studies are more similar due to common designs, partial exchangeability between studies, acknowledging that some studies are more similar due to common designs, locations. The ease in with which additional hierarchical levels between study-specific parameters and the

overall distribution can be incorporated in the Bayesian framework is also demonstrated.

TABLE V
SUMMARY STATISTICS FOR THE POSTERIOR MEAN LOG ODDS RATIO Θ AND μ

Log odds ratio	Mean	S.D.	2.5%	97.5%
C1: Accounting for study design: case control (ξ_1) or cohort (ξ_2)				
θ_1	-0.0362	0.1889	-0.4061	0.3372
θ_2	-0.3936	0.1955	-0.7777	-0.0085
θ_3	-0.5155	0.1846	-0.8787	-0.1516
θ_4	0.5178	0.1588	0.2031	0.8283
ξ_1	0.0140	0.4504	-0.8937	0.9118
ξ_2	-0.3227	0.6616	-1.631	1.0090
μ	-0.0527	0.4981	-1.178	0.9288
C1: Accounting for study location: USA (ξ_1) or UK (ξ_2)				
θ_1	-0.0189	0.1899	-0.3903	0.3580
θ_2	-0.4239	0.1937	-0.8083	-0.0445
θ_3	-0.5156	0.1817	-0.8733	-0.1578
θ_4	0.5251	0.1581	-0.2121	0.8346
ξ_1	0.1816	0.4945	-0.8233	1.1550
ξ_2	-0.3544	0.5023	-1.3260	0.6784
μ	-0.0285	0.4775	-1.0780	0.9382
C1: Accounting for whether the study adjusted for confounding: Yes (ξ_1) or No (ξ_2)				
θ_1	-0.01224	0.1897	-0.3837	0.3636
θ_2	-0.4265	0.1940	-0.8088	-0.0438
θ_3	-0.5156	0.1819	-0.8741	-0.1576
θ_4	0.5181	0.1575	0.2051	0.8267
ξ_1	0.2697	0.4403	-0.6313	1.1330
ξ_2	-0.3445	0.5125	-1.3330	0.7137
μ	-0.0107	0.4721	-1.0210	0.9700

REFERENCES

[1] C. L. Cooper and R. Payne, Personality and stress: individual differences in the stress process. Chichester, UK: Wiley, (1991).
[2] A. Rozanski, J. A. Blumenthal and J. Kaplan, Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. Circulation; 99, 2192-217 (1999).
[3] C. C. Chen, A. S. David, H. Nunnerley, M. Michell, J. L. Dawson, H. Berry, J. Dobbs and T. Fahy, Adverse life events and breast cancer: case-control study. British Medical Journal 311, 1527-30 (1995).
[4] S. F. Duijts, M. P. Zeegers and B. V. Borne, The association between stressful life events and breast cancer risk: A meta-analysis. International Journal of Cancer 107, 1023–1029 (2003).
[5] F. D. Roberts, P. A. Newcomb, A. Trentham-Dietz, B. E. Storer, Self-Reported Stress and Risk of Breast Cancer. Cancer 77, 1089-1093 (1996).
[6] W. H. DuMouchel, Bayesian meta analysis in statistical methodology. Pharmaceutical Sciences, D Berry Ed Marcel Dekker (1990).
[7] D. G. Fryback, N. K. Stout and M. A. Rosenberg, An Elementary Introduction to Bayesian Computing Using WinBUGS. International Journal of Technology Assessment in Health Care, 17, 98-113 (2001).
[8] D. J. Spiegelhalter, A. Thomas and N. Best, "WinBugs Version 1.4 User Manual", MRC Biostatistics Unit, software available at http://www.mrcbsu.cam.ac.uk/bugs/ winbugs/contents.shtml (2000).