A new multivariate meta-analysis model for many variates and few studies

CJ Rose\textsuperscript{1,*}, U Olsen\textsuperscript{2,3}, MF Lindberg\textsuperscript{2,3}, E M-L Denison\textsuperscript{1}, and A Lerdal\textsuperscript{4,5}

\textsuperscript{1}Division for Health Services, Norwegian Institute of Public Health, Oslo, Norway
\textsuperscript{2}Department of Nursing Science, Institute of Health and Society, Faculty of Medicine, University of Oslo, Norway
\textsuperscript{3}Department of Orthopaedic Surgery, Lovisenberg Diocesan Hospital, Oslo, Norway
\textsuperscript{4}Department of Patient Safety and Research, Lovisenberg Diocesan Hospital, Oslo, Norway
\textsuperscript{5}Department of Interdisciplinary Health Sciences, Institute of Health and Society, Faculty of Medicine, University of Oslo, Norway

* Corresponding author

September 28, 2020

Abstract

Studies often estimate associations between an outcome and multiple variates. For example, studies of diagnostic test accuracy estimate sensitivity and specificity, and studies of prognostic factors typically estimate associations for multiple factors. Meta-analysis is a family of methods for synthesizing estimates across multiple studies. Multivariate models exist that account for within-study correlations and between-study heterogeneity. The number of parameters that must be estimated in existing models is quadratic in the number of variates, which means they may not be usable if data are sparse with many variates and few studies. We propose a new model that addresses this problem by approximating a variance-covariance matrix that models within-study correlation and between-study heterogeneity in a low-dimensional space using random projection. The number of parameters that must be estimated in this model is quadratic in the dimensionality of the low-dimensional space, making estimation more tractable. We demonstrate the method using data from an ongoing systematic review on predictors of pain and function after total knee arthroplasty.

Keywords multivariate random-effects meta-analysis; missing data; sparsity; Bayesian statistics; random projection

Git revision 4b45278
Introduction

Meta-analysis is a family of statistical methods used to synthesize estimates of one or more common variates reported by multiple studies [2]. The aim is to obtain a single estimate that statistically characterizes the totality of the available evidence, often including any between-study heterogeneity. For example, a variate of interest might be the prevalence of a particular disease, or a risk ratio comparing a treatment to a comparator. As in the example, the most commonly-used meta-analysis models are univariate, which means that each primary study contributes an estimate of a single variate. Univariate meta-analysis is not necessarily appropriate if there are multiple variates. While it may be tempting to apply univariate meta-analysis to each variate separately, this does not account for possible correlations between variates and does not allow “borrowing of strength” across variates and studies [3]. Univariate meta-analyses applied in the multivariate setting are expected to provide excessively biased and imprecise estimates [10].

The alternative is multivariate meta-analysis, which in principle can model all variates of interest, as well as any within-study correlation and between-study heterogeneity, simultaneously. Perhaps the most well-known application of multivariate meta-analysis within biomedical research is in studying diagnostic test accuracy (DTA), in which the variates sensitivity and specificity are of interest [8]. It is generally recognized that univariate meta-analysis is inappropriate for DTA because changing the threshold that distinguishes positive from negative test results to increase sensitivity will typically decrease specificity. Correlation between sensitivity and specificity is not modeled by univariate meta-analyses. Multivariate meta-analysis is also of use in the study of prognostic factors in which a given outcome may be associated with more than one factor. Network meta-analysis (multiple treatment comparison) can also be posed as multivariate meta-analysis [20]. In addition to the challenges faced in univariate meta-analysis, the multivariate setting poses additional ones, some or all of which may be addressed by available methods:

1. It is rare for authors of primary studies to report an estimate of the full variance-covariance or correlation matrix.

2. It cannot be assumed that every primary study provides estimates for all variates of interest.

3. As in univariate meta-analysis, it is typical to observe between-study heterogeneity in the estimates.

A further challenge, addressed herein, is the scenario in which the number of variates (e.g., prognostic factors) is large relative to the number of primary studies. It may not be possible to use existing multivariate meta-analysis models in such circumstances because the number of parameters needed to estimate within-study correlation and between-study heterogeneity is quadratic in the number of variates. Our contribution is to use a low-dimensional variance-covariance matrix that approximates common within-study correlation and between-study heterogeneity. We do this using a dimensionality reduction method called random projection [12, 5]. This allows us to reduce the number of parameters that must be estimated to be quadratic in the dimension of the approximating space. Estimation is thereby more tractable when there are few studies and many variates.

This paper begins with a motivating example from an ongoing systematic review of predictive factors in which existing methods could not be used. We then provide mathematical background on multivariate meta-analytical methods and explain in more detail why estimation is challenging when
there are many variates and few studies. We then introduce our model and demonstrate its use on data from the motivating example. We close with a discussion and suggest avenues for future research in this area.

Motivating example

This work was motivated by an ongoing systematic review of factors that may predict chronic pain and physical function after total knee arthroplasty (TKA) [14]. About 20% of patients who undergo TKA experience post-surgical pain and reduced function [1], and numerous prognostic factors have been studied. Being able to characterize factors predictive of post-surgical pain could lead to better health outcomes and resource use. Following the inclusion criteria specified in our protocol, we extracted 37 estimates of correlation coefficients between 23 predictors and pain from 6 studies that included a total of 5428 patients (approximately 2700 patient-years of follow-up). We had planned to perform multivariate meta-analysis using an extension of the common correlation model of Riley et al. [15], as implemented in the MVMETA add-on command for Stata [18, 19]. However, because the extracted data are sparse, it was not possible to perform the prespecified analysis unless only predictors supported by at least four studies were included. This limited the number of predictors for which multivariate meta-analysis estimates could be obtained to just two (see figure 1). We then attempted to use Lin and Chu’s model [13], which was developed to address such problems, but the available data were too sparse to support this model.

Background

Let $S$ be a set of $m$ studies. The $i$-th study provides $1 \leq t_i \leq p$ point estimates and sampling variances, where $p$ is the total number of unique variates studied. Let the point estimates provided by study $i$ be denoted $y_i \in \mathbb{R}^{t_i}$ and the corresponding diagonal matrix of sampling variances be denoted $D_i \in \mathbb{R}_{>0}^{t_i \times t_i}$. Given the $y_i$ and $D_i$, we wish to estimate the true value of the $p$ variates, $\mu \in \mathbb{R}^p$, accounting for

---

![Figure 1: Multivariate meta-analysis could only be performed for two of the 23 predictors using the model of Riley et al.](image-url)
within-study correlation and between-study heterogeneity. We assume that none of the studies report within-study correlation or variance-covariance matrices.

Riley et al. [15] proposed a bivariate meta-analysis model that assumes a common within-study correlation parameter. A multivariate version of this model has been implemented for Stata by White [18, 19]. Assuming such a model is parameterized in terms of a common within-study correlation matrix and a between-study variance-covariance matrix that models heterogeneity, the parameters to be estimated are the \( p \) elements of \( \mu \), the \( \frac{1}{2}p(p-1) \) elements of the common within-study correlation matrix (which is unitriangular), and the \( \frac{1}{2}p(p+1) \) elements of the upper or lower triangle of the variance-covariance matrix that models heterogeneity. Such a model requires a total of \( p^2 + p \) parameters to be estimated. It may be challenging to fit such a model unless the total number of point estimates provided by the studies \( n = \sum_{i \in S} t_i \geq p^2 + p \). For many research questions, sufficient studies and estimates may not exist, particularly if \( p \) is large. We say the problem is sparse if \( n < p^2 + p \).

Lin and Chu [13] developed on the model of Riley et al. and addressed the sparsity problem by modeling the variance-covariance matrix for each study as a sum of sampling and additional variances, and by assuming a common correlation matrix for all \( p \) variates. Their model requires estimating the \( p \) elements of \( \mu \), the \( p \) additional variances, and the \( \frac{1}{2}p(p-1) \) correlations, for a total of \( \frac{1}{2}p(p+3) \) parameters.

### A low-dimensional model

As in previous work, we make the simplifying assumptions that variates not estimated by particular studies are missing completely at random, and have a common correlation structure across the primary studies [13, 15]. We develop on Lin and Chu’s model by assuming that the within-study correlation structure and between-study heterogeneity are well approximated in a low-dimensional space. This allows us to reduce the number of parameters that must be estimated. Our model is:

\[
y_i \sim \mathcal{N}(X_i \mu, \Phi_i) \text{ for } i \in S
\]

where

\[
\Phi_i = D_i + X_i R^\top \Sigma R X_i^\top
\]

\(X_i\) is a \( t_i \times p\) indicator matrix. Its \((j, k)\)-th element is unity if the \( j \)-th estimate reported by the \( i \)-th study corresponds to the \( k \)-th of the \( p \) variates and is zero otherwise. \( R \in \mathbb{R}^{q \times p} \) is a matrix that maps between the full \( p \)-dimensional space and a \( q \)-dimensional space (where \( q < p \)) in which the full within- and between-study variance-covariance matrix is approximated by the symmetric positive-definite matrix \( \Sigma \in \mathbb{R}^{q \times q} \). Table 1 summarizes the number of parameters that must be estimated by

| Model                | Number of Parameters |
|----------------------|----------------------|
| Riley et al.         | \( p^2 + p \)        |
| Lin and Chu          | \( \frac{1}{2}p(p+3) \) |
| Our Model            | \( p + \frac{1}{2}q(q-1) \) |

Table 1: The number of parameters that must be estimated for each model; \( p \) is the total number of variates across all studies, and \( q \) is the dimensionality of the space in which within-study correlation and between-study heterogeneity are modelled.
the models of Riley et al., Lin and Chu, and our model. In brief, the number of parameters that must be estimated for our model scales as $O(q^2)$ rather than $O(p^2)$, as for those of Riley et al. and Lin and Chu. Figure 2 plots number of model parameters as a function of number of variates for Riley’s and Lin and Chu’s models, and our model with $q = 4$ and $q = 8$.

We use a random projection matrix $R$ for $R$, but any suitable dimensionality-reduction method could be used instead. There is a large literature on the theory and applications of random projection, and [5] provides a good introduction. The following is a brief and informal treatment of the relevant concepts. Further background on the approach is provided in the Discussion section.

Recall that an orthogonal linear transformation matrix $T \in \mathbb{R}^{p \times p}$ maps between orthonormal bases and preserves the magnitudes and angles between vectors. Such transforms are of interest in multivariate statistics, in particular in principal component analysis (PCA), which has application in dimensionality reduction. Briefly, PCA can be posed as follows: given a variance-covariance matrix $\Lambda$, find matrix $T$ such that $T \Lambda T^\top$ is a diagonal matrix of variances that preserves the total variance of $\Lambda$. This can be achieved via eigendecomposition of $\Lambda$, giving a matrix of eigenvectors ($T$) and their associated eigenvalues, the latter of which provide information about the proportion of the total variance explained in the direction of each of the eigenvectors. Dimensionality reduction can be achieved via the matrix $T' \in \mathbb{R}^{q \times p}$, with $q < p$. $T'$ is formed by dropping those eigenvectors from $T$ that have the smallest associated eigenvalues. Hence, a dimensionality-reducing transform can be defined that preserves at least a given proportion of the total variance. In short, the original matrix $\Lambda$ can be approximated in a $q$-dimensional space by $T' \Lambda T'^\top$.

PCA is applicable when the variance-covariance matrix is known (or can be estimated). When it is not known, other methods must be used. Random projection is a method of establishing an approximately orthogonal linear transform, $R$, that defines a basis in a low-dimensional space. Interestingly, the elements of $R \in \mathbb{R}^{q \times p}$ are sampled randomly from a particular distribution. A variance-covariance matrix

Figure 2: Number of model parameters as a function of number of variates for Riley’s and Lin and Chu’s models, and our model with $q = 4$ and $q = 8$. Curves for our model are shown for $q < p$. 
matrix \( \Lambda \) can be approximated in a \( q \)-dimensional space by \( R\Lambda R^\top \). There are three key differences between PCA and random projection that are relevant to our model:

1. Unlike PCA, a random projection matrix can be constructed without knowing anything about the variance-covariance matrix, except for its dimension.

2. Unlike PCA, \( R\Lambda R^\top \) is not necessarily diagonal. It is therefore necessary to estimate all elements of the low-dimensional matrix \( \Sigma \), which is then transformed to the \( p \)-dimensional space by \( R^\top \Sigma R \).

3. The eigenvectors and eigenvalues obtained in PCA are often of interest to the analyst in their own right. For example, it may be useful to know that 95% of total variance can be explained by three principal components, and how they relate to the original variates. In random projection, however, provided the transform defines a sufficiently useful basis, it is of little interest to the analyst because it is essentially arbitrary.

The number of parameters to be estimated could be reduced further by modeling \( \mu \) in the same \( q \)-dimensional space via \( \beta = R\mu \in \mathbb{R}^q \). While modeling within- and between-study variances and covariances in a low-dimensional space might be expected to lead to poorer quantification of precision, modeling \( \mu \) in this way might be expected to lead to bias, which is arguably more serious.

**Application to knee pain data**

We used our model to analyze the TKA knee pain data introduced in the motivating example above. As for that analysis, we extracted correlation coefficients or imputed them from extracted estimates of regression coefficients, risk ratios, and odds ratios. We converted correlation coefficients using Fisher’s \( z \)-transform (hyperbolic arctangent function) prior to meta-analysis. Further details are given in our protocol [14].

We implemented our model within the Bayesian framework using Stan version 2.24.1 [4], although frequentist implementations would also be possible. We used the priors \( \mu \sim \mathcal{N}(0,10^3) \) and \( \Sigma \sim \mathcal{W}^{-1}(I,q+1) \), where \( \mathcal{W}^{-1} \) is the inverse Wishart distribution and \( I \) is the identity matrix. We chose to model the within- and between-study variance-covariance matrix in \( q = 4 \) dimensions. Estimation was performed using four Hamiltonian Monte Carlo chains, the No-U-Turn Sampler (NUTS) [7], and default numbers of warmup and sampling draws (1000 each per chain). We only accepted posterior samples if the potential scale reduction factor was less than 1.01. Meta-analytical estimates were transformed back to the correlation coefficient scale using the hyperbolic tangent function.

Figure 3 shows estimates of correlation coefficients between each predictor and pain measured six months post-TKA. We present posterior means and equal-tailed 95% credible intervals. Note that because the systematic review is ongoing, we have disguised the names of the predictors. Note also that the predictor identifiers used in figures 1 and 3 do not correspond in this version of the manuscript. We compared the estimates from our model with those from White’s implementation of the model of Riley et al., and to univariate meta-analysis. There is good agreement between the three approaches with respect to the point estimates. Figure 4 shows the posterior mean variance-covariance matrix \( R^\top \Sigma R \) (computed element-wise) estimated for the TKA knee pain data.
Figure 3: Forest plot showing posterior means and 95% credible intervals for the 23 predictors included in our ongoing systematic review on pain and function after total knee arthroplasty. Note that the predictor identifiers do not correspond to those in figure 1.
Discussion

We have described the problem of multivariate meta-analysis when data are sparse and proposed a tractable model that approximates within-study correlations and between-study heterogeneity in a $q$-dimensional space, where $q$ is smaller than $p$, the total number of variates of interest. The main advantage of this model is that it can be used when data are too sparse for methods proposed by Riley et al. and Lin and Chu.

Our implementation uses random projection, which is typically used in problems in which both $p$ and $q$ are of much higher dimension than is commonplace in multivariate meta-analysis. Analogously to the preservation of magnitudes and angles in orthogonal linear transforms, Johnson and Lindenstrauss showed that it is possible to embed high-dimensional data into much lower-dimensional spaces while preserving the relative distances between points [9]. Indyk et al. subsequently proposed using random matrices [12]. A wide range of distributions have been shown to yield good results with high probability, for example almost all zero mean, unit variance distributions can be used [5]. Lu and Lió studied random projection in low dimensions [10] and showed, perhaps unsurprisingly, that the distortion introduced when $q$ is small can be negligible if the intrinsic dimensionality of the original space is low. Random projection and related methods have subsequently become an important tool in high-dimensional statistics and machine learning, and have been applied to a range of problems, including regression [17], mixture modeling [5], text analysis [3], and medical imaging [11].

The main disadvantages of our model are the requirement to choose $q$, potentially increased bias and poorer coverage induced by dimensionality reduction, and the model’s inability to disentangle the within- and between-study variances and covariances, which may be of interest in their own right. However, these disadvantages may be acceptable when $n < \frac{1}{2}p(p + 3)$ and multivariate meta-analysis is preferred over $p$ univariate meta-analyses. Our model should probably not be used when data are not sparse (i.e., where models that make fewer assumptions can be used instead). We suggest that authors wishing to use our model report the possible limitations introduced by the low-dimensional approximation, and compare the estimates provided by our model to those from
a conventional multivariate meta-analysis that includes as many variates as possible, as well as to univariate meta-analyses. Inconsistencies between models should be reported and their implications explained in a way that can be understood by non-quantitative decision-makers.

Future work could address estimating the free parameter $q$ (e.g., by placing a prior over the dimensionality of the model or otherwise integrating over $q$); characterizing the statistical properties of the method and comparing them to other approaches (e.g., Lin and Chu’s); using alternative dimensionality-reduction methods; modeling the distinction between studies that adjusted estimates for other variates, versus studies that did not; and modeling the scenario in which variates are not missing completely at random. These are challenging problems in the face of sparse data.

References

[1] Beswick AD, Wylde V, Gooberman-Hill R, et al. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. BMJ Open 2012;2:e000435

[2] Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2011). Introduction to meta-analysis. John Wiley & Sons.

[3] Bingham, E. and Mannila, H. (2001). Random projection in dimensionality reduction: applications to image and text data. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 245-250.

[4] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M. et al. (2017). Stan: A probabilistic programming language. Journal of Statistical Software, 76(1).

[5] Dasgupta, S. (2000). Experiments with random projection. In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, 143-151.

[6] Deeks, J. J. (2001). Systematic reviews of evaluations of diagnostic and screening tests. BMJ, 323(7305), 157-162.

[7] Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15(1), 1593-1623.

[8] Irwig, L., Macaskill, P., Glasziou, P., and Fahey, M. (1995). Meta-analytic methods for diagnostic test accuracy. Journal of Clinical Epidemiology, 48(1), 119-130.

[9] Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. Contemporary Mathematics, 26(189-206), 1.

[10] Lu, Y. E., Lió, P., and Hand, S. (2008, October). On low dimensional random projections and similarity search. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, 749-758.

[11] Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. (2008). Compressed sensing MRI. IEEE Signal Processing Magazine, 25(2), 72-82.
Author contributions
CJR developed the model, performed the analyses, and wrote the manuscript. UO, MFL, EMLD, AA, and AL planned the systematic review, collected data, and contributed to the manuscript.

Funding
This work is supported by funding from the Norwegian Research Council of Norway (grant #287816) and the South-Eastern Regional Health Authority (grant #2018060).

Conflicts of interest
Within the previous five years, CJR was employed by OncoImmunity AS. He has patents and patent applications with no relevance to this study. The other authors do not report any potential conflicts of interest.